DNA evidence resides at the center of modern criminal justice (1) and is used to help apprehend, convict, and exonerate suspects (2). An ideal DNA system would provide identification information with thoroughness, accuracy, and objectivity. These desirable features are already found in the data generation process, in which a largely automated DNA laboratory transforms biological specimens into quantitative computer signals. However, the subsequent data interpretation process may not possess these ideal attributes.

Rule-based human or computer (3) review can work well with pristine DNA data, such as reference samples. Casework evidence is usually not pristine, however. DNA evidence extracted under real-world conditions is often mixed (having multiple contributors), damaged (by heat or bacteria), or low level (thus hard to discern with certainty).

The DNA data pattern that results from such evidence may suggest multiple genotype possibilities, thereby reducing identification information. Human review of ambiguous DNA can be a time-consuming process that considers only some genotype possibilities, and does not fully elicit all the information that the data contain (4). Moreover, human comparison of DNA evidence and suspect genotypes may not be entirely objective (5–8).

Computer interpretation of DNA evidence overcomes these issues. Specifically, quantitative computer modeling is:

- **Thorough**, with parallel computers considering virtually every genotype possibility;
- **Accurate**, able to employ mathematical models that more fully preserve the identification information residing in the DNA data; and
- **Objective**, interpreting evidence without knowledge of any suspect’s genotype.

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

Computers can thoroughly examine every genotype possibility through statistical Markov chain Monte Carlo (MCMC) computing (9,10) on probability models (11,12). Probabilistic genotypes have been recognized by regulatory bodies such as SWGDAM (13) and standards organizations such as ANSI/NIST (14) as a valid approach to DNA interpretation and reporting. Genotype probability has a long scientific tradition in genetics (15–17) and human identification (18–20), with statistical computer solutions for DNA mixtures (21,22). Human review simplifies DNA data and considers only some genotype possibilities (22). The well-established laws of probability (23,24) and scientific inference (25,26) permit a more complete interpretation.

This study examined Cybergenetics TrueAllele® Casework, a probabilistic genotyping computer system that interprets DNA evidence using a statistical model. We analyzed 368 items of anonymized, adjudicated evidence from the New York State Police (NYSP) Forensic Investigation Center (FIC). We studied 81 mixture items, focusing on contributors with mixture weights between 5% and 95%; these contributors yielded 87 inferred genotypes that matched a known suspect. We measured the sensitivity, specificity, and reproducibility of these genotypes using

1Cybergenetics Corp, 160 North Craig Street, Suite 210, Pittsburgh, PA 15213.
2Northeast Regional Forensic Institute, 1400 Washington Avenue, Albany, NY 12222.
3New York State Police Forensic Investigation Center, 1220 Washington Avenue, Albany, NY 12226.
4Presented at the 62nd Annual Scientific Meeting of the American Academy of Forensic Sciences, February 22–27, 2010, in Seattle, WA.

Received 18 Jan. 2012; and in revised form 5 Aug. 2012; accepted 16 Sept. 2012.
match information. We compared computer results with those from human review.

This study was submitted to the DNA Subcommittee of the New York State Commission on Forensic Science as part of the technology approval process mandated for NYS public DNA laboratories. The DNA Subcommittee made a unanimous binding recommendation for TrueAllele approval on May 20, 2011, which was ratified by the Commission on June 27, 2011.

In this paper, we first discuss computer methods for interpreting quantitative DNA mixture data and information-based validation metrics. We then describe the case materials used in the study and how we classified items. Finally, we present validation results that quantify the sensitivity, specificity, and reproducibility of computer interpretation on this mixture data set.

Methods and Materials

Interpreting Uncertain DNA Evidence

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

The DNA identification task can thus be understood as a two-step process (27):

- Objectively inferring genotypes from evidence data, accounting for allele pair uncertainty using probability, and
- Subsequently matching genotypes, comparing evidence with a suspect relative to a population, to express the strength of association using probability.

The match strength is reported as a single number, the likelihood ratio (LR), which quantifies the gain in identification information produced by having examined the DNA evidence (28–30).

The TrueAllele Casework system is a MATLAB® (The MathWorks, Natick, MA) computer implementation of this two-step DNA identification inference approach. The computer objectively infers genotypes from DNA data through statistical modeling (31,32), without reference to any suspect. To preserve the identification information present in the data, the system represents genotype uncertainty using probability (33). These probabilistic genotypes are stored on a PostgreSQL relational database (34) for persistent and secure storage (35). Subsequent comparison with suspects provides investigative and evidentiary identification information. A user visually asks interpretation questions of DNA case data, reviews the computer’s answers, and generates a report on the match statistics found.

Identification Information

This study uses identification information (36) to quantify the sensitivity, specificity, and reproducibility of DNA interpretation methods and for making comparisons between methods (37). The LR logarithm (powers of ten or “order of magnitude”) provides an additive measure of match information (38,39) called the “weight of evidence.”

With a strong DNA match between evidence and suspect, relative to a population, the LR is typically more than a million (log(LR) > 6). The LR can be in the quintillions (log(LR) > 18) when using 13 or more short tandem repeat (STR) loci on unambiguous DNA data. A mismatch will have a small LR under 1, often less than a trillionth (log(LR) < -12). The LR thus quantifies the extent of match, based on DNA data (40,41). A numerical presentation of scientific support is more precise than qualitative words like “inclusion” or “exclusion” (42), which are binary decisions perhaps best left to a trier of fact.

The weight of DNA evidence is used as an information metric throughout this study (Results). For method sensitivity, log(LR) measures the identification information inferred from DNA data. Interpretation specificity is measured by a negative log(LR) value between evidence and a nonmatching individual. We quantify reproducibility through information variation within a case. The productivity of DNA processing can be assessed by how often an informative sample produces a reportable match statistic.

Genotype Probability

Uncertainty in DNA evidence translates into genotype uncertainty. All mixture interpretation methods report out a list of allele pair possibilities and assign a probability to each pair (43). Prior to examining DNA data, our belief is that an evidence genotype has an allele pair probability distribution based on allele frequency in a human population. Observing data at a genetic locus can change this belief, because some contributor genotype possibilities become more likely than others, based on how well they explain the data. Combining likelihood and prior will form a new genotype, whose posterior probability distribution (44) typically eliminates hundreds of allele pairs and concentrates on just a few probable ones.

DNA mixture interpretation methods differ by the likelihood function they employ to explain the data. The most informative methods have a likelihood function that uses all the quantitative peak height data and other available information (45); these include statistical computer modeling systems (21,46) such as TrueAllele (47). Human interpretation often simplifies the observed peak heights into qualitative all-or-none allele events to form a list of allele pairs (27). Such threshold-based binary likelihood functions can lose identification information, as they do not use all of the available data (45). Combined likelihood ratio (CLR) qualitative interpretation considers known (e.g., victim) contributor genotypes, whereas combined probability of inclusion (CPI) does not (22).

Mixture Weight Probability

Individuals contribute their DNA to a mixture item in a certain proportion, or “mixture weight.” Analysts can infer this mixture weight (mean and variance) for the DNA template from quantitative peak height data and assumed contributor genotypes. This human approximation uses only those genetic loci where every allele peak can be unambiguously assigned to some contributor. Computers can instead use all of the loci to infer the mixture weight of the DNA template (32), achieving greater precision.

Mixture weight is an auxiliary variable that can help assess the complexity of a mixture item (Results). A small mixture weight can produce a lower quantity of contributor DNA, which may reduce match information through stochastic effects. In a 50:50 mixture interpreted without assuming a victim reference, more allele pair combinations are possible, and this genotype uncertainty can decrease DNA match information.
Evidence Items

We reanalyzed 41 cases (39 adjudicated and two proficiency tests) that had been previously analyzed by competent DNA analysts at the NYSP FIC. These 41 cases covered many interpretation situations commonly encountered in forensic casework. The cases contained 32 sexual assaults (some having a victim reference) and homicides involving up to 30 evidence items with multiple victims (Table 1A).

The 368 study items were derived from 206 distinct biological source samples. These samples included diverse evidence sources, such as vaginal swabs, anal swabs, oral swabs, penile swabs, dried secretions, blood stains, semen stains, weapons, cigarette butts, condoms, human hair, bite marks, and fingernail scrapings (Table 1B).

The NYSP laboratory processed the items using Applied Biosystems (AB; Foster City, CA) ProfilerPlus<sup>®</sup> and Cofiler<sup>®</sup> STR kits and generated electronic data .fsa files on an AB 3100-series genetic analyzer. Cybergenetics uploaded these files to the TrueAllele system for probabilistic genotyping. The data and results were then available for a retrospective comparison with the corresponding case reports.

The TrueAllele-inferred mixture weights of these evidence items were broadly distributed (Fig. 1). Our goal was to study actual mixtures, rather than background contamination or single-source samples. We therefore focused on the 87 inferred evidence genotypes from the 81 matching items that had a mixture weight between 5% and 95%. Different mixture interpretation methods may define a match in different ways (e.g., by item, contributor, or genotype). In this multimethod study, a match is always between a genotype of an evidence contributor and a known reference genotype.

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

<table>
<thead>
<tr>
<th>(A) Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexual assault</td>
<td>32</td>
</tr>
<tr>
<td>Homicide</td>
<td>4</td>
</tr>
<tr>
<td>Assault</td>
<td>2</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B) Source</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaginal swab</td>
<td>17</td>
</tr>
<tr>
<td>Anal swab</td>
<td>7</td>
</tr>
<tr>
<td>Penile swab</td>
<td>1</td>
</tr>
<tr>
<td>Semen stain</td>
<td>9</td>
</tr>
<tr>
<td>Clothing item</td>
<td>10</td>
</tr>
<tr>
<td>Bedding item</td>
<td>3</td>
</tr>
<tr>
<td>Weapon</td>
<td>11</td>
</tr>
<tr>
<td>Cigarette butt</td>
<td>2</td>
</tr>
<tr>
<td>Condom</td>
<td>1</td>
</tr>
<tr>
<td>Dried secretion</td>
<td>6</td>
</tr>
<tr>
<td>Hair</td>
<td>3</td>
</tr>
<tr>
<td>Bite mark</td>
<td>2</td>
</tr>
<tr>
<td>Fingernail</td>
<td>9</td>
</tr>
<tr>
<td>Blood stain</td>
<td>69</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>206</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(C) Category</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>39</td>
</tr>
<tr>
<td>Intermediate</td>
<td>18</td>
</tr>
<tr>
<td>Complex</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>81</td>
</tr>
</tbody>
</table>

Prior to computer analysis, mixture evidence items were classified as simple, intermediate, or complex and assessed by an experienced DNA analyst. About half of the items were in the simple category, with the remainder divided between the intermediate and complex categories (Table 1C). The number and nature of the contributors to the items were tabulated, as shown (Table 2). Other mixture classifications (48,49) have been tailored to manual mixture review protocols having limited sensitivity and are not relevant to the study of data information content or statistical computer interpretation.

Most of the simple mixture case items had two contributors (Table 2, simple) and came from sexual assault differential extractions. Usually one contributor was known, and the task was to infer the unknown second contributor. The DNA quantity was adequate, and the data showed clear major and minor contributors.

The intermediate mixture items were more challenging for people to interpret. Some were derived from low-template DNA sources, while others contained contributors in approximately equal 50:50 mixture weights. They had two or three contributors, with multiple unknown contributors appearing in most of the items (Table 2, intermediate).

Complex mixture items had two or three contributors, usually with multiple unknown contributors (Table 2, complex). The STR data often showed peak height imbalance or equally weighted contributors. There were low-template DNA items, and several cases had multiple suspect or victim references.

Results

The study examined 4958 alleles in 202 single-source profiles across the 41 cases. The genotypes inferred by the TrueAllele statistical computing system were in complete concordance with human review of the same data. With single-source concordance established, the remainder of the paper focuses on validating DNA mixture interpretation.

Opinions vary about what genotype should be inferred from DNA mixture data (45). Regardless of interpretation method, the...
log(LR\textsubscript{method}) of any inferred genotype, relative to a known suspect and reference population, provides a measure of identification information in a single number. Human (e.g., CPI, CLR) and computer (e.g., TrueAllele) mixture interpretation methods thus share a common information vocabulary, so we can quantitatively assess and compare DNA interpretation methods through their match statistics.

Sensitivity

Identification information measures sensitivity, quantifying how well an interpretation method preserves the log(LR) weight of DNA evidence. We examined TrueAllele sensitivity on a set of 87 matches, comparing inferred genotype information from simple, intermediate, and complex cases (Fig. 2). These histograms show that the computer preserved considerable identification information. Simple cases (blue) had an average log(LR) of 14.92, intermediate cases (green) averaged 12.05, and the complex case (orange) average was 10.47 (Table 3, $\theta = 1\%$). These log(LR) averages correspond to match statistics at the quadrillion, trillion, and billion levels, respectively.

With current manual mixture interpretation methods, not all items of evidence will yield a reportable match statistic. Prior to SWGDAM’s 2010 mixture interpretation guidelines, a DNA analyst might have reported an item using qualitative “consistent with,” “cannot be excluded,” or “insufficient” language, without attaching a match number. Now, under current SWGDAM guidelines, a laboratory’s newly elevated peak threshold may result in no statistic (50). The TrueAllele computer, however, has no choice but to infer a genotype that can be used later on in a statistical match comparison. The computer is programmed to report a LR value at every locus in a genotype comparison, unlike human review (5). If the data are uninformative, that fact is reflected in a diffuse genotype distribution and a low LR score.

The weight of DNA evidence in this study ranged from a log(LR) of 2.02 (hundred) to 21.35 (sextillion), with a median value of 14.50 (quadrillion). On mixtures found in forensic practice,

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Number of Total Number Knowns in Knowns Unknowns Unknowns DNA Item Number of</th>
<th>Number of</th>
<th>Number of</th>
<th>Number Reported by</th>
<th>Number Reported by</th>
<th>Fraction Reported by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>2 1 2</td>
<td>32</td>
<td>17</td>
<td>0.531</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>2 1 1</td>
<td>8</td>
<td>2</td>
<td>0.250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>2 1 1</td>
<td>20</td>
<td>4</td>
<td>0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 0 2</td>
<td>7</td>
<td>2</td>
<td>0.286</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 1 2</td>
<td>5</td>
<td>2</td>
<td>0.400</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0 3</td>
<td>1</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0 3</td>
<td>15</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0 3</td>
<td>5</td>
<td>2</td>
<td>0.400</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0 3</td>
<td>27</td>
<td>4</td>
<td>0.148</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0 3</td>
<td>87</td>
<td>25</td>
<td>0.288</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LR, likelihood ratio.

**TABLE 2**—Item complexity. The rows are organized into three groups, based on our simple, intermediate, and complex mixture classification. In the left half of the table, the number of contributors is shown, arranged by total, known, and unknown contributor columns. The right half shows the number of inferred genotypes, giving the number having a TrueAllele match statistic, how many of those were reported by the human review laboratory with an associated match statistic, and the corresponding fraction of human to computer reporting (the computer always generated a match statistic when human review produced a statistic).

![FIG. 2—Genotype information sensitivity. The distribution of genotype match information (as inferred by TrueAllele) is shown as a log(LR) histogram of counts for each complexity category, with theta at 1%. The simple items (blue) are distributed more to the right than the intermediate items (green). The leftmost distribution is for the complex items (orange), which tend to be less informative.](image)
TrueAllele can produce match statistics comparable to single-source DNA random match probabilities. This occurs when there is clear genotype separation of the component contributors.

Specificity

When two genotypes are dissimilar, their match information can be negative. The greater the dissimilarity, the greater the magnitude of the negative log(LR) value. Equivalently, when a person did not contribute their DNA to the biological evidence, the support in the STR data for their having contributed is generally far less than the alternative view that some else contributed, that is,

\[
\text{LR} = \frac{\Pr \{ \text{data|person contributed DNA} \}}{\Pr \{ \text{data|someone else contributed} \}} < 1
\]

so that log(LR) is negative. The degree of negativity can be used to quantify how specifically an interpretation method discriminates between nonmatching genotypes.

We compared each of the 87 matched mixture evidence genotypes with the (≈87) reference genotypes from the other 40 cases. Each of these 7298 comparisons should generate a mismatch between the unrelated genotypes from different cases and hence a negative log(LR) value. A genotype inference method having good specificity should exhibit mismatch information values that are negative in the same way that true matches are positive.

The (empirical) distribution functions of information are shown in Fig. 3 for the simple, intermediate, and complex items on a log(LR) scale with theta at 1%. The average weight of evidence was −22.65 for the simple genotypes, with a maximum of −6.55 (Table 4). Restated linearly, the average statistic of a mismatching simple item was under one in sextillion and never exceeding one in a million. The information distribution of intermediate items moved to the right, having a mean of −20.46 and a maximum of −4.70. Complex items shifted further, averaging −19.56, but not exceeding −3.67. While less definite genotypes gave less definitive mismatch statistics, the log(LR) was always less than zero and was well separated from the positive values of matching genotypes.

Reproducibility

We measured interpretation reproducibility by examining the information variation of duplicate computer runs on the same evidence item. These log(LR) variations can be combined across many case items in the same category to compute a within-group standard deviation (4). The computer lends itself to reproducible solutions—it is far easier to rerun a program on the same data than have people arduously rework the same problem manually.

We had previously observed that one unknown mixture problems are more reproducible than two unknown problems (4). Across the spectrum of the 87 inferred matching genotypes from mixture items in this study, we again see that more informative results tend be more reproducible (Fig. 4). In this figure, the two replicate TrueAllele runs (blue, green) are shown sorted by descending information, with a median log(LR) of 14.50. As identification information decreases (from left to right), we see increasing divergence between an item’s two inferred log(LR) values (red line).

This information divergence is a natural consequence of statistical sampling and is independent of any particular mathematical model or data interpretation method. Lower DNA quantity increases a data peak’s coefficient of variation, which in turn diffuses genotype probability. Because less probable allele pairs are sampled by the computer less frequently, they exhibit greater match statistic fluctuation. TrueAllele’s information variation is relatively small (Table 3).

<table>
<thead>
<tr>
<th>Mixture Category</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>3380</td>
<td>−22.65</td>
<td>3.300</td>
<td>−6.55</td>
</tr>
<tr>
<td>Intermediate</td>
<td>1664</td>
<td>−20.46</td>
<td>4.111</td>
<td>−4.70</td>
</tr>
<tr>
<td>Complex</td>
<td>2254</td>
<td>−19.56</td>
<td>4.964</td>
<td>−3.67</td>
</tr>
</tbody>
</table>

FIG. 3—Genotype information specificity. The distribution of genotype mismatch comparison information (as inferred by TrueAllele) is shown as a log(LR) histogram of counts for each complexity category, with theta at 1%. The simple items are most specific (blue), distributed more to the left than the intermediate items (green). The rightmost distribution is for the complex items (orange), which are the least specific.

FIG. 4—Replicate genotype information. The inferred genotypes are sorted by descending match information, with theta at 1%. For each matched genotype, the first (blue bar) and second (green bar) independent TrueAllele computer runs’ match statistics are shown on a log(LR) scale. The information difference between the two genotype replicates is shown (red line).
Categorizing items by their complexity confirms that more informative data produce (on average) more reproducible interpretation (Table 3, $\theta = 1\%$). The simple items had a within-group log(LR) standard deviation of 0.249, while complex ones more-than-doubled the spread yet again to 0.519. The observed with-in group standard deviations (Table 3) were all less than the match statistic variation due to population sampling, typically one power of ten with standard population databases having over a hundred individuals (51). TrueAllele mixture interpretation was thus reproducible in all the data situations examined.

**Case Reporting**

A DNA match statistic should be reproducible. When reporting a case, Cybergenetics will process an item in (at least) two independent TrueAllele computer runs. When duplicate runs yield match statistics that differ by more than two log units, one of the Markov chain solutions may have failed to properly sample from the posterior distribution.

**Human Comparison**

Human review may not always produce a numerical match value. For the 40 computer-matched genotypes inferred from simple items, human review in this study yielded a statistic (e.g., CPI, CLR) 42.5% of the time (Table 2, fraction reported by laboratory with DNA statistic). The human yield was less on the 20 intermediate items, with people generating a statistic for just 20% of the inferred genotypes. On the 27 computer-inferred genotypes from complex items, only 14.8% resulted in a statistic upon human review. Overall, the laboratory's human review reported a numerical DNA match statistic for 25 of the 87 computer genotype matches (28.8%), or 25 statistics produced for the 81 evidence items (30.9%).

We can represent the relative information yield for these mixtures in a chart that shows the amount of preserved identification information, along with the mixture interpretation method (Fig. 5). The TrueAllele weight of evidence (blue) is shown in the background and sorted in descending log(LR) order for the 87 computer-inferred genotype matches. For the 24 recorded human interpretation numerical scores, a LR equivalent match result is shown, with bars for the 4 single-source conditional match probability (gray), 8 mixture CLR (green), and 12 mixture CPI (orange) results. Overall, we see less human-derived match information, both in the number of reported match statistics and in the statistic’s value.

The human inferred log(CLR) statistic of 8.20 (Fig. 5, third green bar from left) was appreciably lower than its CLR neighbors, which had statistics of 15.03 and 13.52 (the corresponding computer log(LR) values were all around 18). This match statistic reduction occurred because the human interpretation reported on just 8 of the 13 STR loci in this case. At the other five loci, the unknown 60% major contributor shared its alleles with the victim. Many laboratories do not report victim genotypes, which can reduce the match score. The computer is not restricted in
this way and, instead, infers and reports contributor genotypes at every locus.

**Productivity**

TrueAllele interpretation produced a match statistic for every item. With human review, however, three mixture items were examined for each match statistic that was produced (Table 2).

**Mixture Weight**

For comparison purposes, we manually computed the mixture weights on Excel spreadsheets for 45 two person mixtures, considering only those loci where the contributor allele sets did not overlap. At each locus, the contributor peak height sum was divided by the allele peak height total to estimate mixture weight. Examining the differences between TrueAllele-computed mixture weight (using all loci in a probability model computation) and these less exact human spreadsheet estimates, we found good agreement between the computer and human solutions (Fig. 6).

For each of the three classifications, we paired each DNA match’s log(LR) ($\theta = 1\%$) together with its mixture weight. These scatter plots show how greater item complexity is associated with reduced match information and somewhat lower mixture weight (Fig. 7). On average (Table 3), the simple mixture contributor genotypes (Fig. 7A) had more information (log(LR) of 14.92) and a higher mixture weight (54.9%). Intermediate genotypes (Fig. 7B) showed less identification information (12.05) and somewhat lower mixture weights (48.9%). The complex items (Fig. 7C) had the least information (10.47) and about the same mixture weights (50.1%).

There is an apparent simple category outlier in the bottom left corner of the scatter plot (Fig. 7A), having a mixture weight of 7% and a log(LR) of 2.02 (see also Fig. 2, leftmost blue bar). The item had two contributors, each with its own matched genotype. The item classification was based on the (easier to interpret) major contributor genotype, rather than on the minor contributor genotype involved in the match shown. Human review examines unseparated item data peaks, whereas a probabilistic computer infers and compares a mathematically separated contributor genotype with a reference genotype.

**Conclusions**

In today’s forensic laboratory, a greater range of biological evidence undergoes DNA analysis (e.g., property crimes, touch DNA), with ever more challenging items submitted (e.g., low-template mixtures). Many laboratories have responded by introducing automation into their workflow. This greater analytical capacity, in turn, has created bottlenecks in data interpretation and technical review. To overcome these bottlenecks (53), we tested how Cybergenetics TrueAllele Casework statistical modeling system could provide automated DNA data interpretation.

The study reanalyzed 41 previously reviewed cases spanning diverse interpretation situations commonly encountered in forensic casework. There were 368 evidence items across many categories of evidence, including 81 mixture items that yielded 87 genotype matches. We found genotype concordance in all the items, verifying the computer’s ability to accurately separate genotypes across a wide range of mixture weights and complexities.

A thorough mixture interpretation considers all genotype possibilities, whether or not they are immediately evident from the
peak height data. Thorough statistical review also determines allele peak uncertainty directly from the observed data (54), obviating the need for a less precise threshold approach to addressing stochastic variation. Such thoroughness is needed for robust probability inference (44). Computers can achieve this thorough examination through statistical sampling (9,10), whereas a person might take many years to explore one mixture problem’s vast variable space.

The interpretation of STR DNA data by TrueAllele probability modeling provides objectivity, eliminating examiner bias in forensic DNA casework (8). The evidence genotypes are inferred from the data without any knowledge of a suspect genotype (4). Using this system, a genetic testing laboratory could design their workflow to ensure that analysts complete their evidence examination and record inferred genotypes, before they see a suspect’s genotype (7).

The TrueAllele system achieves greater genotype accuracy in probabilistic inference from mixture data than does current manual review (13,22). The system preserves more of the identification information present in the evidence by making better use of the data (45). When human review does report a match statistic, the computer’s positive log(LR) can provide greater sensitivity. When the data do not support a match, the computer quantifies its specificity through a negative log(LR). With inconclusive human review on informative data, the computer will typically produce a match statistic (55,56).

The study showed that TrueAllele’s statistical processing was reproducible across independent replicated computer runs (4). This reproducibility, coupled with greater accuracy, establishes the validated system’s reliability for forensic casework and experimental science (57). A reproducible DNA interpretation method permits standardization of mixture reporting in the laboratory. Current human review approaches can lead to very different match statistics on the same DNA data (45,58,59). Trained TrueAllele analysts, however, can produce statistically similar results when interpreting the same data.

In summary, we have shown that TrueAllele Casework is a reproducible DNA interpretation method. It provides a means of minimizing examiner bias in forensic DNA mixture interpretation. The system preserves more of the identification information present in the evidence by making better use of the data. A computerized system can provide greater sensitivity and specificity than human review.

Acknowledgments

New York State Police FIC forensic scientist Elizabeth Staude collected the case data, receiving considerable assistance from scientists Russell Gettig, Missy Lee, and Shannon Morris. NYSP counsel Steve Hogan provided legal guidance that enabled us to conduct this study. Cybergenetics staff Cara Lincoln and Jessica Smith performed most of the TrueAllele processing, with additional support from Meredith Clarke and Matthew Legler. NYS DNA Subcommittee members gave valuable feedback in the course of their review of preliminary manuscript drafts. Two anonymous reviewers made suggestions that greatly improved the clarity and concision of the paper.

References