CONTINUING LEGAL EDUCATION

Fall 2013

November 7, 2013

INTERPRETATION OF DNA MIXTURES AND STATISTICS

HEATHER MILLER COYLE, PH.D. AND MICHAEL ALPERSTEIN, ESQ.

SPONSORED BY:
APPELLATE DIVISION, FIRST AND SECOND JUDICIAL DEPARTMENTS
CO-SPONSORED BY:
THE ASSIGNED COUNSEL PLAN OF THE CITY OF NEW YORK

Alphabet Soup Analogy (also known as All-pairs Trawls and the Birthday Problem)

Concept: Instead of trawling DNA databases for partial matches, FST use with mixtures trawls for all possible combinations of DNA profiles that could be included to generate a match rate and validation studies have yielded the false inclusion rate of non-contributors. Reference: Kaye, D. 2009. Trawling DNA Databases for Partial Matches: What is the FBI Afraid of? Cornell Journal of Law and Public Policy 19(1): 145-171.

Point: Arizona DNA Database: when searched for 9 locus matches out of core 13 CODIS loci, partial matches in non-degraded samples with a DNA database the size of 65, 493 yielded 122 pairs of individuals that matched at 9 loci and 20 pairs that matched at 10 loci.

Point: Likelihood ratios increase the chance of coincidental matching due to lack of an inconclusive category option in DNA reporting. The chance of coincidental matching also increases as information level decreases due to DNA degradation. Reference: Tedeschi, S., Coyle, H. 2013. Likelihood Ratios: Uses and Issues. Northeastern Association of Forensic Scientists Annual Meeting, Cromwell, CT (posted on www.identacode.org).

Concept: Individualization without uniqueness in DNA mixtures. Reference: Kaye, D. 2010. Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies. Brooklyn Law Review. 75(4): 1163-1185.

Two competing theories (Likelihood Ratio analogy): if there is a DNA match, then what are the odds it is the defendant versus anyone else?

Is it more like, what are the odds you will suffocate when all the oxygen molecules suddenly shift to the other side of the room (which seems unlikely) or more like Alphabet Soup?

With Alphabet Soup (DNA mixtures), the best answer is "could be included or cannot be excluded" as well as the total of all the other people calculated as coincidental matches from the trawl of the reference database plus anyone who also was undetected by DNA for scientific accuracy. DNA cannot establish which combination of alleles is most likely in a mixture it can only establish those alleles that were detected. If those alleles can be attributed to someone else as a known person (victim, for example), then our confidence level regarding the remaining alleles increases since the options then become smaller for all the remaining possible pairwise combinations. If not, then it is anyone's guess.

The Birthday Problem is similar. Most people would say there is a 1 in 365 chance of someone having the same birthday since there are 365 days in the year but with actual careful study, the real probability estimate is 1 per 23 individuals would by coincidence share the same birthday.

Likelihood ratios:
For some mixtures wherein an individual contributor's DNA profile cannot be determined, a known person's DNA profile can still be compared to the mixture. The comparison DNA profile can be from a known person, or from a single source or deduced profile from within a case. For these comparisons, a statistical value known as a likelihood ratio (LR) may be calculated. The LR value provides a statistical measurement of the strength of support for one scenario over another, i.e., one scenario being that the known person contributed to the mixture versus the scenario that an unknown, unrelated person contributed instead.

Limited, moderate, strong or very strong support: These terms describe the strength or weakness of different ranges of a likelihood ratio (as shown in the table below). Examples of factors that affect the LR value include the amount of DNA tested, the type of mixture (for example, the number of contributors), instances when one or more of the individual's DNA alleles are not seen in the mixture, the presence of rare alleles in the mixture, and the presence of extra DNA alleles in the mixture.

Reported value	Qualitative interpretation
1	No conclusions
1 to 10	Limited support
10 to 100	Moderate support
100 to 1000	Strong support
Greater than 1000	Very strong support

Note, if the LR value is less than one, this means that the mixture is better explained if an unknown, unrelated person contributed to the mixture rather than the known person. This situation is reported as 1/LR and the qualitative terms from the table above are applied.

Sample OCME – NYC Report Conclusion for Likelihood Ratios

(1) Is the defendant included?
(2) What is the weight of the LR value?
(3) On this scale, the categories are very broad
(4) 0-1 is still a no conclusion category (not shown here)
(5) Limited support – is formerly known as "cannot be excluded" which is also "inconclusive"
(6) 10-20 is also very, very weak association with typically estimates of approximately 100 or more individuals that could also contribute to the mixture if of 3 –persons and an empirical reference database of 500,000 to compare to
(7) False inclusion rates exist – and some LR values fall into higher categories (see Table 4A and Table 4B, Likelihood Ratio Statistics for Analysis of Single Source, Mixed and Degraded Evidence Samples – Executive Summary OCME-NYC. These are explained by source attribution errors and by coincidental matching between samples.

Table 4A. Frequency of observed LRs among non-contributors to two-person samples

	Observed frequency (1 in x):	
LR greater than:	Deducible	Non-Deducible
0.001	270	900
0.01	580	2,000
0.10	1,500	4,800
1	5,000	10,000
10	16,000	19,000
100	55,000	176,000
1,000	> 166,000	> 176,000
10,000	> 166,000	> 176,000

Table 4B. Frequency of observed LRs among non-contributors to three-person samples

	Observed frequency (1 in x):	
LR greater than:	Deducible	Non-Deducible
0.001	80	300
0.01	170	610
0	410	1,300
1	1,200	3,100
10	4,400	7,600
100	13,000	40,000
1,000	31,000	60,000
10,000	> 93,000	121,000

These are estimated false inclusion rates for coincidental matching when a mixture is interrogated against a 500,000 individual database; these are the number of individuals that are estimated to match by coincidence due to allele sharing. Therefore, it is proven possible to have a high LR value and still be a noncontributor due to coincidental matching. Source: Table 4A and Table 4B, Likelihood Ratio Statistics for Analysis of Single Source, Mixed and Degraded Evidence Samples – Executive Summary OCME-NYC.

J Forensic Sci, Nov. 2005, Vol. 50, No. 6
Paper ID JFS2004475
Available online at: www.astm.org

David R. Paoletti,[1] *M.S.; Travis E. Doom,*[1,2] *Ph.D.; Carissa M. Krane,*[3] *Ph.D.; Michael L. Raymer,*[1,2] *Ph.D.; and Dan E. Krane,*[4] *Ph.D.*

Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures

ABSTRACT: Samples containing DNA from two or more individuals can be difficult to interpret. Even ascertaining the number of contributors can be challenging and associated uncertainties can have dramatic effects on the interpretation of testing results. Using an FBI genotypes dataset, containing complete genotype information from the 13 Combined DNA Index System (CODIS) loci for 959 individuals, all possible mixtures of three individuals were exhaustively and empirically computed. Allele sharing between pairs of individuals in the original dataset, a randomized dataset and datasets of generated cousins and siblings was evaluated as were the number of loci that were necessary to reliably deduce the number of contributors present in simulated mixtures of four or less contributors. The relatively small number of alleles detectable at most CODIS loci and the fact that some alleles are likely to be shared between individuals within a population can make the maximum number of different alleles observed at any tested loci an unreliable indicator of the maximum number of contributors to a mixed DNA sample. This analysis does not use other data available from the electropherograms (such as peak height or peak area) to estimate the number of contributors to each mixture. As a result, the study represents a worst case analysis of mixture characterization. Within this dataset, approximately 3% of three-person mixtures would be mischaracterized as two-person mixtures and more than 70% of four-person mixtures would be mischaracterized as two- or three-person mixtures using only the maximum number of alleles observed at any tested locus.

KEYWORDS: forensic science, DNA typing, DNA mixtures, short tandem repeats, Combined DNA Index System, allele sharing, bioinformatics

PCR-based amplification of STR loci has become the method of choice for the purpose of human identification in forensic investigations (1,2). While alternatives exist (3,4), most DNA-typing laboratories use commercially available kits to amplify and label STR alleles associated with evidence and reference samples that are then size fractionated with capillary electrophoresis systems such as the ABI 310 or 3100 Genetic Analyzers (5,6). Software such as GeneScan® and Genotyper® are then used to determine the presence or absence of STR alleles associated with a sample.

Interpreting evidence samples containing mixed DNA profiles is more complicated than the analysis of single source samples. Programs exist that can aid analysts attempting to "deconvolve" the contributors of mixed samples on the basis of associations between peak heights or areas (7,8) and, in some instances, using the genotype information from individuals presumed to have been contributors (7). Analysis of a multiple contributor sample is particularly challenging when potential contributors have several alleles in common (such as is often the case with close relatives), when stochastic variations in peak heights occur, or when technical artifacts such as stutter, allelic dropout, and degradation/inhibition occur.

With only rare exceptions, an individual should possess exactly two actual alleles for every locus. These alleles may differ from each other (heterozygous) or be effectively indistinguishable (homozygous). For example, at one STR locus an individual may be found to have alleles 11 and 12, or 12 and 13 (Fig. 1*A*, 1*B*). When more than two actual alleles are observed in the testing results from any single locus, it can be reasonably assumed that the presence of DNA from more than one contributor is the most likely explanation. The absence of a fifth or sixth actual allele is often interpreted as support of there being only two contributors to mixtures (Fig. 1*C*) even though it is formally possible for the number of contributors to be greater than two. However, if three actual alleles are observed, the sample may arise from a mixture of two individuals, a mixture of three individuals with overlapping alleles, or even a mixture of four or more individuals. Likewise, observing five or more actual alleles at one locus is an indication of three or more contributors. However, it becomes increasingly difficult to determine the exact number of contributors as the number of observed alleles increases. Although previous research (9) has analyzed how often a two-person mixture will present 1, 2, 3, or 4 alleles at six individual loci, no published studies address the issues of how often a three-person mixture will present no more than four different alleles at any tested locus, or how often a four-person mixture will present no more than six different alleles at any tested locus.

Exhaustive analysis of 959 complete 13-locus STR genotypes from a population dataset, as well as of randomized sets of comparable genotypes in this study could assist DNA analysts by formally addressing the relative statistical confidences of declarations regarding the number of contributors to a DNA mixture on the basis of alleles observed at typed STR loci. Peak heights and areas sometimes provide additional data that is useful for the purpose of mixture deconvolution but this information is not utilized in the analysis presented here. Instead, this study examines the interpretation of STR data in cases where this information is unreliable (i.e. when degradation has occurred and/or stutter complicates interpretation), unavailable (i.e., only a laboratory's summary report is provided for review) or uninformative (i.e., the relative contributions by two or more contributors are similar).

[1] Computer Science Department, Wright State University, Dayton, OH 45435.
[2] Forensic Bioinformatics, Inc., 2850 Presidential Drive, Suite 150, Fairborn, OH 45324.
[3] Biology Department, University of Dayton, Dayton, OH 45969.
[4] Department of Biological Sciences, Wright State University, Dayton, OH 45435.

Received 6 Nov. 2004; and in revised form 28 April 2005; accepted 9 June 2005; published 14 Sept. 2005.

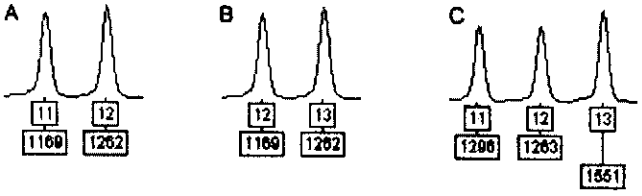


FIG. 1—*Sample electropherograms from possible single-source and mixed samples. A single STR locus with alleles 11 and 12 (A); the same STR locus but now with alleles 12 and 13 (B); and the same STR locus but now with three distinct alleles: 11, 12, and 13 (C). Allele number designations for each peak appear immediately below it with corresponding peak height information (in relative fluorescence units) immediately below.*

Materials and Methods

In order to conduct this study, a dataset containing complete STR-DNA profiles of several hundred individuals was needed. Such a dataset from the FBI, used for the determination of allele frequencies, has already been analyzed for Hardy-Weinberg equilibrium (10), and is publicly available (11). This FBI dataset consists of complete typing information for the 13 commonly used CODIS STR loci for 959 individuals from six different racial groups: Bahamian (153 individuals), Jamaican (157 individuals), Southwest Hispanic (202 individuals), Trinidadian (76 individuals), US African American (177 individuals), and US Caucasian (194 individuals). The original dataset (11) contains typing information for a larger number of individuals, but any with incomplete information, i.e., allele "0," were discarded for this study.

Published analyses of the FBI dataset make no specific mention of the extent of the effort made to assure that close relatives were not included in the population sampling (10). Thus, to guarantee that that there are absolutely no relation-based linkages among individuals for some aspects of this study, a dataset of "randomized individuals" was also generated. In this randomization, the actual alleles observed in the FBI dataset were distributed randomly to produce a new set of genotypes equal in number to the original dataset. Allele frequencies in this randomized dataset are the same as in the original dataset but individuals are unequivocally unrelated by descent (alleles are not the same because they have been faithfully passed from a common ancestor). Instead, any allele sharing can arise only through identity by state (alleles are the same because there is a finite number of different alleles that can be detected). Each locus was considered independently during the production of randomized genotypes. For each locus, the alleles of all individuals in the original dataset (without respect to racial classification) were randomly redistributed among the same number of synthetic individuals; an example of one possible redistribution amongst three individuals is shown in Table 1. This redistribution occurs without replacement, thus each locus in a randomized dataset has the same allele frequencies as the corresponding locus in the original dataset. Source code for all of the analyses can be found at (12).

All individuals in the original dataset are assumed to have two and only two alleles per locus (rare conditions resulting in unusual allele counts such as null alleles, triploidy or chimerism are beyond the scope of this study). Similarly, all simulated mixtures of genotypes are considered to be free of any typing errors that might further complicate the interpretation.

TABLE 1—*Example of alleles being redistributed amongst three individuals.*

	vWA	
Individual	Original	Redistributed
A	18, 19	15, 17
B	17, 18	18, 18
C	14, 15	14, 19

Shared Allele Counts

Homozygotes were deemed to share two alleles with other homozygotes with the same genotype (e.g. an individual who was 12, 12 was determined to share two alleles with another 12, 12 individual but none with a 10, 10 individual). Homozygotes could share either one or no alleles with heterozygotes (e.g., a 12, 12 homozygote would share one allele with an 11, 12 individual and none with a 10, 11 individual). Average shared allele counts were the average pair wise total (with a maximum of 26 arising from two alleles across 13 loci) number of shared alleles observed between all possible pairs of individuals in a dataset.

Shared Allele Counts with Related Individuals

The greater the number of shared alleles between pairs or clusters of individuals within a population, the greater the chance that maximum numbers of alleles observed per locus may suggest an incorrect minimum number of contributors. In order to determine upper bounds on the number of shared alleles observed between pairs of individuals, we consider a "worst case" situation. What would be observed if there were relatives of every individual in the dataset? To answer this question, virtual families of individuals were created using genotypes from the randomized dataset.

Each virtual family consists of two "cousins" (C1–C2), their four "parents" (P1–P4, of which P2 and P3 are siblings), and their six "grandparents" (G1–G6) (Fig. 2). The grandparents of each family are drawn from datasets of randomized individuals to preclude complications from the possible presence of related individuals already being present in the FBI dataset. Thus, each set of six randomly selected virtual individuals produces one family.

In this simulation, each of two parents contributes one of its alleles (chosen randomly) for each locus in the production of their virtual offspring. For each dataset of 959 individuals, 159 sets of six individuals were chosen to be "grandparents" to produce 159 3-generation families, each containing a cousin and sibling pair. From these synthetic families two datasets containing related individuals were created: one populated with pairs of siblings and one populated with pairs of cousins. In order to maintain similarity in the scope of the study, roughly the same number of two-person combinations in the virtual family datasets were considered as in the original dataset (459,361). This is done by choosing the grandparents randomly over the course of 2,889 runs (459,361 two-person combinations/159 virtual families).

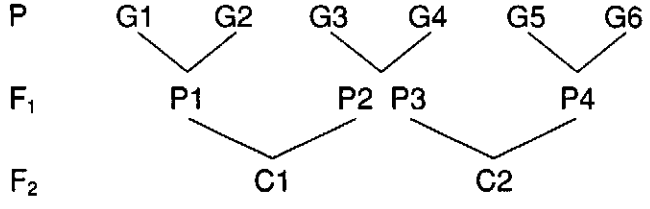


FIG. 2—*Creation of virtual families from the dataset. The individuals in the "P" line (G1–G6) are 6 profiles chosen randomly from the dataset of synthetic individuals, representing grandparents. They produce offspring shown in line F1 (parents P1–P4) with siblings P2/P3 as shown. Line F2 shows the grandchildren (C1, C2) of the original profiles, who are first cousins.*

Three-person Mixture Analysis

The number of three-person combinations of a set of n individuals is determined as:

$$N = \frac{n!}{(n-3)!3!}$$

where n is the number of individuals, and N the number of combinations. For instance, for 4 individuals (A, B, C, and D) there are $4!/(1! \cdot 3!) = 24/6 = 4$ different three-person combinations (ABC, ABD, ACD, and BCD). Given 959 individuals, each of the 146,536,159 different possible three-person 13 locus genotype combinations were considered and the number of different alleles represented at each of the 13 loci was determined for each mixture. The same process was also performed using the genotypes within each of five randomized datasets, and the results averaged.

If no more than two different alleles were observed at all of the 13 loci considered, then the three-person mixture was considered to be potentially mischaracterized as the profile of a single individual. Likewise, if no more than four alleles were observed over all loci, then the three-person mixture has the potential to be mischaracterized as a mixture of only two individuals. Potential differences in peak heights (i.e. due to additivity associated with shared alleles) were not considered in this study.

We have observed that the standard operating procedures of forensic DNA testing laboratories sometimes allow analysts to discard information from loci that they determine to be anomalous based on their training and experience. One factor that could conceivably cause a locus to appear anomalous would be the observation of five or six alleles (suggesting a minimum of three contributors to a mixture), while the other 12 loci possess only four or fewer different actual alleles (consistent with a mixture of two contributors). To assess the ramifications of invoking analyst discretion to discard such a locus, we analyzed the number of three-person genotype mixtures where discarding a single locus with the highest number of different observed alleles produces results consistent with mischaracterization of the mixture as a single source sample or as a two-person mixture.

Four-person Mixture Analysis

Computing all possible four-person mixtures of a set of 959 individuals is impractical (there are 35,022,142,001 such mixtures). Consequently, analyses of four-person mixtures was restricted to a subset of the FBI dataset (specifically, the 194 Caucasians) resulting in 57,211,376 different four-person mixtures. For this analysis, we assume that mixed genotypes that allow the observation of 7 or 8 different alleles at even one locus will be correctly identified as a four-person mixture. In the same way, mixed genotypes where the locus or loci with the greatest number of different alleles observed have either 5 or 6 alleles will be considered to be mistakenly characterized as a three-person mixture. Mixed genotypes where the locus or loci with the greatest number of different alleles have either 3 or 4 alleles observed will be considered to be mistakenly characterized as a two-person mixture.

The number of loci that need to be considered for at least 95% of the simulated four-person mixtures to be correctly characterized as having originated from at least four contributors (e.g., had at least one locus with 7 or 8 alleles) was empirically determined. Since 13 loci proves to be insufficient to reach 95% confidence, new virtual loci were introduced by randomly selecting one of the original 13 loci and creating a simulated locus with equivalent discriminating power by randomly redistributing alleles. This process was repeated in five parallel simulations until a level of 95% correct characterization was independently achieved in each simulation.

Results

Shared Allele Counts

The 959 individuals of the FBI dataset can also be combined in $n = 459{,}361$ different pairings. The distribution of the number of alleles shared between pairs of unrelated individuals shows expected similarity ($p = 1.00$ by a two-tailed t-test) between the original dataset ($\bar{x} = 8.59$, $\sigma = 2.16$) and the randomized dataset ($\bar{x} = 8.59$, $\sigma = 2.15$). No pairs of individuals were found to share more than 25 of 26 alleles in the original dataset, or 20 of 26 in the randomized datasets. The distribution of the number of alleles shared between virtual cousins ($\bar{x} = 10.95$, $\sigma = 2.27$) and virtual siblings ($\bar{x} = 16.94$, $\sigma = 2.30$) was significantly different ($p = 0.00$) as were the distributions for shared alleles between virtual siblings and unrelated individuals ($p = 0.00$) but not for virtual cousins and unrelated individuals ($p = 0.29$). The distributions for the number of shared alleles in pairings of randomized individuals, virtual siblings, and virtual cousins are roughly Gaussian (Fig. 3) as was the distribution of pair wise allele sharing in the original dataset.

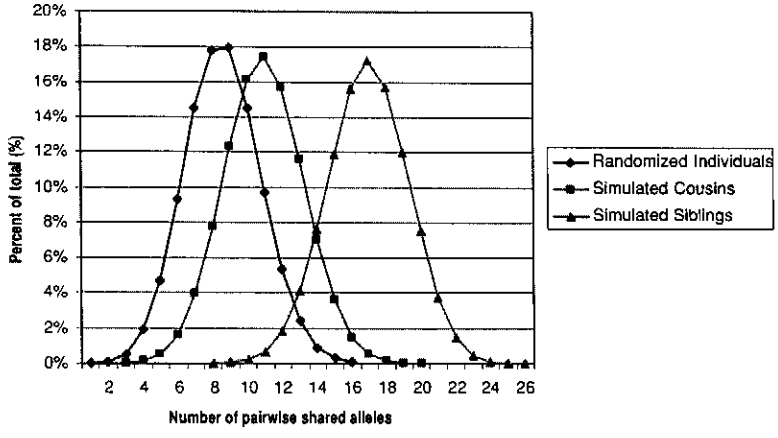


FIG. 3—*The distributions for the number of shared alleles amongst all possible pairs of synthetic individuals, pairs of synthetic siblings, and pairs of synthetic cousins. A total of 459,361 pairs each of randomized individuals, simulated cousins, and simulated siblings were considered.*

Shared Allele Counts with Related Individuals

The original dataset's mischaracterization rate of 3.39% (appearing to be a two-person mixture rather than a three-person mixture on the basis of maximum allele count per locus) is more than two standard errors of the mean above the average observed in the five datasets of randomized individuals. The higher mischaracterization level in the original dataset is attributable at least in part to a greater number of pair wise shared allele counts at or above 19 out of a possible 26 in the original dataset relative to each of the five randomized datasets (3 vs. an average of 1.4). Although neither of these numbers are statistically significant, given the averages and standard deviations, these unusual results are most easily explained by hypothesizing that one or more pairs of individuals in the original dataset may be related. This possibility was the motivation for producing and analyzing the random datasets against the original.

Analyses of virtual families suggest that such high levels of allele sharing are likely to be due to identity by descent, not by state (Fig. 3). It is worth noting that while it is uncommon for virtual siblings in this simulation to have indistinguishable genotypes (matching at all 26 alleles) an average of 3.0 such perfectly matching sibling pairs were generated in the five repetitions of this simulation (459,361 sibling pairs each).

Three-person Mixture Analysis

All possible different three-person mixtures of the 959 individuals in the FBI dataset were considered. Of those 146,536,159 three-person mixtures, 4,967,112 (3.39%) had contributors that possessed overlapping alleles such that none of the 13 loci exhibited more than four alleles (Table 2). Consequently, each of these 4,967,112 mixtures could be mischaracterized as a mixture of two individuals using information from maximum allele count alone. A smaller fraction ($\bar{x} = 3.18\%$, $\sigma = 0.2131\%$, and standard error of the mean = 0.0953%) of similarly mischaracterized mixed profiles was found with the five randomized datasets. None of these three-person mixtures can be misinterpreted as a single contributor, as nowhere are there only one or two alleles observed across all loci.

Many of the simulated three-person mixtures were found to possess just one locus that contained more than four alleles. When these single loci were not considered (e.g., because they were "inconsistent with" or "anomalous relative to" the majority of loci) the mischaracterizations increased dramatically (Table 3) both in the original and five randomized datasets. In this case, 26,791,938 (18.28%) of the three-person mixtures from the original dataset could be mischaracterized as a mixture of just two individuals. An average of 25,591,392.6 or 17.46% ($\sigma = 0.5444\%$) were similarly mischaracterized in the five randomized datasets. As in the previous analysis, none of these three-person mixtures can be misinterpreted as a single contributor.

When this locus with the largest number of observed alleles is discarded, the category can change at most once. For example, if a mixture has only one locus having 6 alleles, one locus having 5 alleles, and the remaining loci with 4 or fewer alleles, it would be counted as changing from a maximum of 6 observed alleles to a maximum of 5 observed alleles (Table 4). Further, if the two highest maximum observed alleles counts are the same, no change occurred. Other useful information, such as per-locus information similar to Table 3, is available on-line (12).

TABLE 2—*Count and percent of three-person mixtures in which a particular number of unique alleles was the maximum observed across all loci, both for the original and randomized individuals**.

Unique Alleles	Count	Percent (%)
2	0	0.00%
3	78	0.00%
4	4,967,034	3.39%
5	93,037,010	63.49%
6	48,532,037	33.12%

A—Original dataset.

Unique Alleles	Count	Percent (%)
2	0.0	0.00%
3	115.8	0.00%
4	4,653,064.2	3.18%
5	92,019,609.6	62.80%
6	49,863,369.4	34.03%

B–Average over five randomized datasets.

* The unique allele column reports the maximum number of different alleles that were observed across all loci for these cenceptual three-person mixtures.

TABLE 3—*Count and percent of three-person mixtures in which a particular number of unique alleles was the second highest observed across all loci, both for the original and randomized individuals, after removal of the locus with the maximum number of unique alleles**.

Unique Alleles	Count	Percent
2	0	0.00%
3	3,398	0.00%
4	26,788,540	18.28%
5	112,469,398	76.75%
6	7,274,823	4.96%

A—Original dataset.

Unique Alleles	Count	Percent
2	0.0	0.00%
3	3,872.0	0.00%
4	25,587,520.6	17.46%
5	113,412,323.0	77.40%
6	7,532,443.4	5.14%

B—Average over five randomized datasets.

* The unique allele column reports the maximum number of different alleles that were observed across all loci for these conceptual three-person mixtures but only after a single locus with the highest count has been discarded.

Four-person Mixture Analysis

A large majority (43,667,840 or 76.34%) of the 57,211,376 possible four-person conceptual mixtures of the 194 Caucasians in the FBI dataset can be mischaracterized as two- or three-person mixtures when maximum allele count observed across all 13 loci was used as the only basis for characterization (Table 5). As expected, the mischaracterization rate decreases as the number of loci considered is increased (Fig. 4). The simulation was run five times, each run halting independently when the mischaracterization rate dropped below 5%. The fewest number of additional loci required to fall below a 5% level of mischaracterization in these simulations was 171 while the most was 177. The addition of information from an average of 27 simulated loci resulted in at least half of the four-person mixtures having at least one locus with more than six different alleles (and thus correctly classified). As with the three-person mixtures, per-locus data similar to Table 5A is available on-line (12).

Discussion

The extent of allele sharing observed between pairs of individuals is clearly influenced by the degree to which the individuals being

TABLE 4—*Count and percent of three-person mixtures in which a particular number of unique alleles was the second highest observed across all loci, both for the original and randomized individuals, after removal of the locus with the maximum number of unique alleles*.*

Unique Alleles			
From	To	Count	Percent of Total
6	5	37,751,585	25.76%
6	4	3,505,340	2.39%
6	3	289	0.00%
5	4	18,317,945	12.50%
5	3	1,252	0.00%
4	3	1,779	0.00%

A—Original dataset.

Unique Alleles			
From	To	Count	Percent of Total
6	5	38,850,621.8	26.51%
6	4	3,479,995.2	2.37%
6	3	309.0	0.00%
5	4	17,456,523.4	11.91%
5	3	1,385.0	0.00%
4	3	2,062.2	0.00%

B—Average over five randomized datasets.

* The unique allele column reports the maximum number of different alleles that were observed across all loci for these conceptual three-person mixtures before and after a single locus with the highest count has been discarded. The middle four rows represent those cases where a change of interpretation has occurred, from three contributors to two.

TABLE 5—*Count and percent of four-person mixtures (from the FBI Caucasian dataset) in which a particular number of unique alleles was the highest observed across all loci, both for the original 13 loci and after the addition of 182 new loci derived from the original 13 loci*.*

Unique Alleles	Count	Percent
4	13,480	0.02%
5	8,596,320	15.03%
6	35,068,040	61.30%
7	12,637,101	22.09%
8	896,435	1.57%

A—Original 13 loci.

Unique Alleles	Count	Percent
4	0	0.00%
5	0	0.00%
6	2,542,148	4.44%
7	45,542,753	79.60%
8	9,126,475	15.95%

B—After the addition of 182 new loci.

* The unique allele column reports the maximum number of different alleles that were observed across all loci for these conceptual four-person mixtures. Three or less alleles are never observed, and thus omitted.

compared are related to each other. Simulations confirm that first-degree relatives (siblings) are more likely to have alleles in common than second degree relatives (cousins) or unrelated individuals (13) (Fig. 3). The larger amount of pair wise allele sharing observed between individuals in the original FBI dataset relative to the five datasets of randomized individuals suggests that the FBI dataset may contain some pairs of closely related individuals. The extent of allele sharing between siblings in large-scale simulations also suggests that perfect 13 locus matches (26 out of 26 possible alleles) occur at a frequency (an average of 3.0 per 459,361). This frequency suggests that some are likely to exist in large populations such as the general population of the United States and even eventually in DNA profile datasets that contain large numbers of close relatives.

A large number (4,967,112) of the possible three-person combinations of the 959 actual individuals included in the FBI population dataset have sufficient allelic overlap between individuals to allow none of 13 STR CODIS loci to exhibit more than four alleles in a mixed sample (Table 2). This observation has important implications for the interpretation of forensic DNA testing results given that mixtures where both the number of contributors and the genotypes

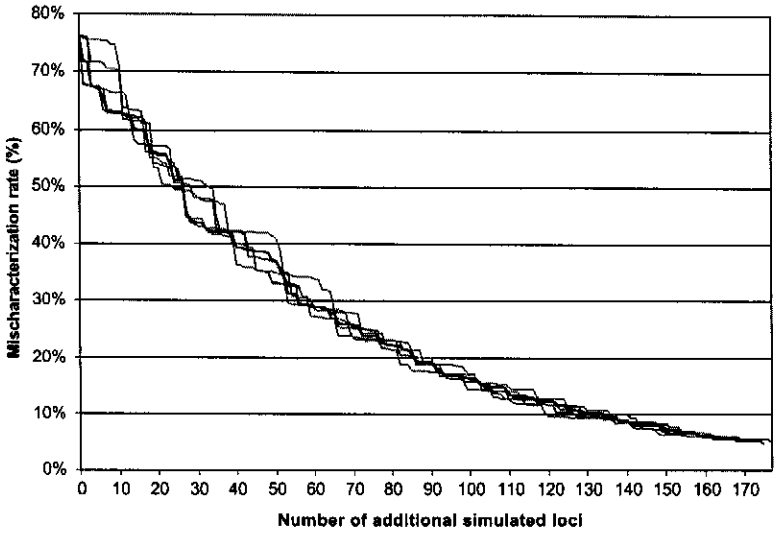


FIG. 4—*The percent of four-person mixtures that could be mischaracterized as arising from just two or three contributors solely on the basis of the maximum number of different alleles observed at any locus considered. Mischaracterization occurred at a rate of 76.34% when only the thirteen CODIS loci in the original FBI dataset were considered. A black line shows the average obtained from five simulations, each of which is shown individually in gray. Each simulation was halted when the mischaracterization rate dropped below 5%. A total of 57,211,376 conceptual four-person mixtures were made for each data point in each simulation as new loci were added.*

of one or more likely contributors are often disputed. Inclusions in such mixtures should always be accompanied by statistics that convey the strength of such a finding since allele sharing and large numbers of observed alleles both diminish the possibility of exclusion. Simulations with equivalent datasets containing randomized individuals yield similar results suggesting that the mischaracterization level is not attributable solely to artifacts or population substructure within the original FBI dataset (data not shown).

The mischaracterization of conceptual three-person mixtures increased more than five-fold when a single locus with the largest number of different alleles was eliminated from consideration. Reasoning along the lines of "a single locus with more than four alleles is likely to be attributable to technical artifacts when all other tested loci are consistent with there being only two contributors" is intuitively appealing. However, the observed dramatic increase in mischaracterization rate in both the original and five randomized datasets suggests that such rationalization is not well-founded.

Much higher rates (76.34%) of potential underestimates of the number of contributors to mixed samples were observed when four-person mixtures were considered. Difficulties in inferring the correct number of contributors to both three- and four-person mixtures are ultimately due to overlapping alleles between individuals. Allele sharing between two individuals is attributable to only two conditions: 1) identity by descent or, 2) identity by state. Allele sharing due to identity by descent can make mixtures involving related individuals particularly problematic. Allele sharing due to identity by state is a potential problem in all mixtures and arises from two related characteristics of the commonly employed STR CODIS loci. First, many of these loci possess a small number of detectable alleles (e.g., in this dataset, there are only six observed alleles for the TPOX locus and only seven for the D13, D5, D3, and TH01 loci, therefore mixtures that display more than six alleles at these loci must be rare). Second, some alleles at some loci are relatively common and therefore likely to overlap between contributors to a mixture. The key factor is that the addition of more individuals (and thus more alleles) into the mixture causes the mixture to become more likely to hide any indications of subsequent individuals, as the relative proportion of present versus absent alleles at each locus increases with each new contributor.

Conclusions

Maximum allele count by itself is not very reliable in terms of predicting the number of contributors to mixed forensic DNA samples, particularly when: the number of loci considered is small, the number of contributors may be large (4 or more), and/or a single point of seemingly inconsistent/anomalous information can be disregarded. However, maximum allele count still results in mistaken inference of the number of contributors at a rate of over 3% even in the best of circumstances (e.g., the 13 STR CODIS loci are considered, there are only three contributors, and no seemingly inconsistent/anomalous information is disregarded). In light of these observations, the practice of many testing laboratories to simply report that a sample arises from "two or more individuals" when more than two alleles are observed at one or more loci during the course of testing is both reasonable and appropriate.

Acknowledgments

We gratefully acknowledge both Jason Gilder and Carolyn Rowland for their assistance and critical reviews of this manuscript. The comments of two anonymous reviewers were also extremely helpful and very much appreciated.

References

1. Edwards A, Hammond HA, Lin J, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992;12:241–53. [PubMed]
2. Frégeau CJ, Fourney RM. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. BioTechniques 1993;15:100–19. [PubMed]
3. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Sprecher CJ. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47(4):773–85. [PubMed]
4. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 2001;46(3):647–60. [PubMed]
5. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B. Validation of STR typing by capillary electrophoresis. J Forensic Sci 2001;46(3):661–76. [PubMed]
6. Frégeau CJ, Bowen KL, Fourney RM. Validations of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers. J Forensic Sci 1999;44(1):133–66. [PubMed]
7. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci 2001;46(6): 1372–7. [PubMed]
8. Wang T, Xue N, Rader M, Birdwell JD. Least square deconvolution (LSD) of STR/DNA mixtures. 7th CODIS User's Conference. Available at http://www.lit.net/presentations2/wang_lest_square%20STRDNA_2001_3.pdf. 2001.
9. Gill P, Sparkes R, Klimpton C. Development of guidelines to designate alleles using an STR multiplex system. Forensic Sci Int 1997;89(3):185–97. [PubMed]
10. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. 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.
11. http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnaloci.txt.
12. birg.cs.wright.edu/~dpaolett/JFS05-materials.zip.
13. Presciuttini S, Ciampini F, Alu M, Cerri N, Dobosz M, Domenici R, Peloso G, Pelotti S, Piccinini A, Ponzano E, Ricci U, Tagliabracci A, Baley-Wilson JE, De Stefano F, Pascali V. Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I. AmpFlSTR® Profiler Plus database. Forensic Sci Int 2003;131(2–3):85–9.

Additional information and reprint requests:
Dan E. Krane, Ph.D.
Biological Sciences
3640 Colonel Glenn Highway
Dayton, OH 45435-0001
E-mail: Dan.Krane@wright.edu

Statistical Analysis of STR Data

By Kimberly A. Huston
e-mail genetics@promega.com

Statistical analysis is used to interpret DNA results for genetic identity. In order to determine the significance of a match, it is necessary to support DNA typing results with statistical analysis. These analyses assign a value to the results obtained and enable easier resolution of forensic or paternity cases.

Polymorphic loci contain different sequences at the same locus within and between individuals. These highly variable loci are used in DNA analysis because of their ability to differentiate individuals. Population databases are used to determine the frequency of each allele for a given locus. These databases are generally defined by racial group and geographical region because alleles may have different frequencies in different populations.

Q: What are the Laws of Probability?

A: The First Law of Probability is denoted by the equation:

$$0 \leq Pr(A/E) \leq 1$$

$$Pr(A/A) = 1$$

Zero is less than or equal to the probability that A is true, given that E is known, which is less than or equal to one. If we know that A is true, then it has a probability of one.

The Second Law of Probability is denoted by the equation:

$$Pr(A \text{ or } B/E) = Pr(A/E) + Pr(B/E)$$

so: $Pr(\bar{A}/E) = 1 - Pr(A/E)$

If A and B are mutually exclusive and E is known, the probability that A or B is true, given E, equals the probability of A, given E, plus the probability of B, given E. Thus, it follows that the probability of A not happening, knowing E, is equal to one minus the probability of A, knowing E.

The Third Law of Probability is denoted by the equation:

$$Pr(A \text{ and } B/E) = Pr(A/B, E) \times Pr(B/E)$$

The probability of A and B, given that we know E, equals the probability of A occurring, knowing B and E, multiplied by the probability of B occurring, knowing E.

From all of the above laws, the Law of Total Probability follows:

$$Pr(A) = Pr(A/B)Pr(B) + Pr(A/C)\ Pr(C)$$

If B and C are two mutually exclusive and exhaustive events, the probability of A is equal to the [probability of A, given B, multiplied by the probability of B] plus [the probability of A given C, multiplied by the probability of C.]

Q: What do exhaustive and exclusive mean?

A: Exhaustive events include all possible outcomes. Exclusive events require that there is no overlap between the outcomes of events.

Q: What does independent mean?

A: Any two events that have no influence on what happens to each other are independent or unassociated. Therefore, for independent events, the probability of both events happening is the product of the probability for each event.

Q: What is meant by being in Hardy-Weinberg Equilibrium?

A: For a population to be in Hardy-Weinberg Equilibrium (HWE), the alleles must be randomly inherited.

Q: What is Bayes' Theorem?

A: Bayes' Theorem is a useful tool for presenting DNA data in a logical manner. The odds form of Bayes' Theorem states:

$$Pr(E/A) = \frac{Pr\ E \text{ and } A}{Pr(A)} = \frac{Pr(A/E)\ Pr(E)}{Pr(A)}$$

and

$$Pr(E/A) = \frac{Pr\ \bar{E} \text{ and } A}{Pr(A)} = \frac{Pr(A/\bar{E})\ Pr(\bar{E})}{Pr(A)}$$

Taking the ratio of the above equations:

$$\frac{Pr(E/A)}{Pr(\bar{E}/A)} = \frac{Pr(A/E)}{Pr(A/\bar{E})} \times \frac{Pr(E)}{Pr(\bar{E})}$$

From this, we derive that the posterior odds are equal to the likelihood ratio multiplied by the prior odds.

Prior odds $\left(\frac{Pr(E)}{Pr(\bar{E})}\right)$ are odds that are assigned based on initial information.

The likelihood ratio $\left(\frac{Pr(A/E)}{Pr(A/\bar{E})}\right)$ is the ratio of two conditional probabilities. By multiplying the prior odds by the likelihood ratio the posterior odds $\left(\frac{Pr(E/A)}{Pr(\bar{E}/A)}\right)$ are obtained.

Q: What is conditional probability?

A: All probabilities are conditional based on what we know to be true.

Q: What is the difference between heterozygotes and homozygotes?

A: Heterozygotes are individuals who have two different alleles at the same locus. Homozygotes are individuals who have two identical alleles at a given locus. Heterozygosity is also called the frequency of heterozygotes and is represented by *h* in the following equation.

$$h = \frac{n_h}{n}$$

Where n_h is the number of individual observations with two alleles and n is the total number of individuals.

Since one is either a homozygote or a heterozygote, the frequency of heterozygotes (h) plus the frequency of homozygotes (H) is equal to one.

$$h + H = 1$$

Q: What is matching probability?

A: Matching probability, also known as probability of match (pM), is the number of individuals that may be surveyed before finding the same DNA pattern in a randomly selected individual. This is represented as:

$$pM = \sum_{i=a}^{n} \sum_{j\geq 1}^{n} P_{ij}^2$$

Where i and j represent the frequencies of all possible alleles a through n, P_{ij} represents the frequencies of all possible genotypes.

The combined matching probability for more than one locus is the product of the individual matching probability at each locus, assuming that they are not linked.

Q: What does the paternity index represent?

A: The paternity index reflects how many more times likely it is that the person being tested is the biological father, rather than a randomly selected individual. The typical paternity index is assigned to a locus rather than an individual case. Generally, a $PI_{typical}$ of less than one is indicative of non-relatedness. The $PI_{typical}$ is represented by the following equation:

$$PI_{typical} = \frac{1}{2H}$$

The $PI_{typical}$ of several loci is the product of the individual $PI_{typicals}$.

Q: What is the power of discrimination?

A: The power of discrimination is one minus pM. The combined power of discrimination for multiple loci may be calculated by the following equation:

$$P_{d\ combined} = 1 - \prod_{i=1}^{n} (1 - P_{di})$$

Q: What is the Wahlund principle?

A: The Wahlund principle is seen within subpopulations where little gene exchange occurs. In these subpopulations, the allele frequencies differ from those predicted by the Hardy-Weinberg principle. This is seen as an increase in homozygotes and a deficiency in the number of heterozygotes in comparison with the expectations of HWE.

Q: What is the power of exclusion?

A: The power of exclusion, PE, is defined as the fraction of individuals having a DNA profile that is different from that of a randomly selected individual in a typical paternity case. The value for each individual case will vary. The average for a given locus is represented by the following equation.

$$PE = h^2(1-2hH^2)$$

The $PE_{typical}$ for several loci is represented in the following equation:

$$PE_{typical} = \prod_{i=1}^{n} 1 - \pi\ (1-PE_i)$$

Q: What is θ?

A: θ, the co-ancestry coefficient, is the probability that two alleles in the population have descended from the same allele and are identical by descent. This is a measure of the coancestry of populations diverging due to genetic drift. The larger θ, the longer it has been since the populations diverged.

Q: Where can I get more information on statistical calculations?

A: For CODIS calculations, contact the Federal Bureau of Investigation. They can provide Popstats 5.1 software, which will assist in calculations. For advice on paternity case calculations, contact the American Association for Blood Banking at 8101 Glenbrook Rd., Bethesda, MD 20814-2749 or on the Internet at www.aabb.org.

REFERENCES

Brenner, C. and Morris, J. (1990) In. *Proceedings of the International Symposium on Human Identification, 1989*, Promega Corporation.

Evett, I.W. and Weir, B. (1997) *Interpreting DNA Evidence*, Sinauer (in press).

Science and Justice 51 (2011) 204–208

Contents lists available at SciVerse ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Subjectivity and bias in forensic DNA mixture interpretation ☆

Itiel E. Dror [a,b,*], Greg Hampikian [c]

[a] Institute of Cognitive Neuroscience, University College London (UCL), London, UK
[b] Cognitive Consultants International (CCI), London, UK
[c] Departments of Biology and Criminal Justice, Boise State University, USA

ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form 30 July 2011
Accepted 4 August 2011

Keywords:
Human cognition
Bias
Forensic decision making
Contextual influences
DNA interpretation

ABSTRACT

The objectivity of forensic science decision making has received increased attention and scrutiny. However, there are only a few published studies experimentally addressing the potential for contextual bias. Because of the esteem of DNA evidence, it is important to study and assess the impact of subjectivity and bias on DNA mixture interpretation. The study reported here presents empirical data suggesting that DNA mixture interpretation is subjective. When 17 North American expert DNA examiners were asked for their interpretation of data from an adjudicated criminal case in that jurisdiction, they produced inconsistent interpretations. Furthermore, the majority of 'context free' experts disagreed with the laboratory's pre-trial conclusions, suggesting that the extraneous context of the criminal case may have influenced the interpretation of the DNA evidence, thereby showing a biasing effect of contextual information in DNA mixture interpretation.

Seeking and interpreting information in a biased way so that it fits existing beliefs, expectation, hope, or motivation is a result of how we reason and is widespread [1]. The potential for such biases in forensic science disciplines has been suggested before [2,3], and has now been highlighted by the National Academy of Science (NAS) report on *Strengthening Forensic Science in the United States: A Path Forward* [4]. It directly discusses "the potential for bias and error in human observers" (p. 8), and states that "the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, [or] the threat of bias .... [is] significant" (p. 9). Indeed, empirical research supports the effects of bias in some forensic disciplines; for example, in fingerprinting, the same forensic experts may arrive at different conclusions when identical evidence is presented within different extraneous contexts (e.g., whether the detective believes the suspect is guilty, or the suspect confessed) [5–8].

However, in contrast to other forensic disciplines, DNA is regarded as the gold standard of forensic science [9]. DNA has been held as objective and immune to subjectivity and bias; "In the past several years, it has become commonplace in the courts, in the media, and in much of the technical literature, to contrast the scientific and objective evidence supplied by DNA profiling, with the experiential or subjective opinions given by traditional forensic experts" [9] (p. 97). Indeed, even the NAS distinguishes between "forensic science disciplines [that] are laboratory based (e.g., nuclear and mitochondrial DNA analysis, toxicology and drug analysis)" [4] (p. 38), and other forensic disciplines that are "based on expert interpretation of observed patterns (e.g., fingerprints, writing samples, toolmarks, bite marks, and specimens such as hair)" [4] (p. 38).

If correct, then DNA analyses should be consistent and not affected by domain irrelevant contextual circumstances. It seems, however, that at least in complex situations (such as with DNA mixtures) DNA does require and rely on human examiners making a variety of subjective judgements that are susceptible to bias. Indeed, in contrast to the view that DNA is objective, some have proposed that DNA analysis interpretations may be subjective and may even be influenced by a variety of factors [10,11].

However, such claims – both for the subjectivity or for the objectivity – of DNA analysis have rarely been examined and tested through empirical research. To investigate the subjectivity and biasability of mixture DNA analysis we observed and compared the conclusions on identical DNA evidence that was presented within and between different extraneous contextual information. To properly investigate this issue, it was critical to: 1. conduct the study with qualified DNA expert analysts who conduct real casework in accredited laboratories, and 2. that the examiners genuinely believed the contextual information, as contrived context within an experimental setup does not have the effect or impact as that of genuinely believed real context [8].

To achieve these goals we used mixture DNA analysis from a real adjudicated criminal case, using records obtained through the Georgia Freedom of Information Act. The case we chose provided us with analysis within extraneous context. We then took the same DNA evidence and presented it to 17 independent North American DNA expert analysts, but without the potentially biasing contextual case

☆ *One sentence summary:* DNA mixture interpretation is subjective and may be susceptible to bias by extraneous context, as evidenced by conflicting conclusions concerning the inclusion or exclusion of suspects.

* Corresponding author.
E-mail address: i.dror@ucl.ac.uk (I.E. Dror).
URL: http://www.cci-hq.com (I.E. Dror).

1355-0306/$ – see front matter © 2011 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.scijus.2011.08.004

information. First we compared the consistency in interpretation and conclusion within those 17 examiners to assess subjectivity in DNA analysis. Then we compared between them and those who examined the DNA mixture within the extraneous context of the criminal case to assess biasability in DNA analysis. The DNA evidence related to a gang rape case in which one of the assailants testified against the other suspects in return for a lesser sentence as part of his cooperation in a plea bargain deal. However, those identified through the plea bargain denied any involvement in the rape.

The mixture DNA from the sexual assault was examined by experts in the real criminal case, and their analysis and conclusions were that the suspects that were identified by the cooperative assailant could not be excluded from being contributors to the mixture. The establishment of this corroborating fact was essential to the prosecution of the suspects who claimed innocence. Under the law of that state where this act occurred, the testimony of the admitted rapist would not be admitted without corroborating evidence. Therefore the DNA conclusions were critical to prosecution. If the suspects were excluded by DNA, or even if the DNA was "inconclusive", the incriminating testimony of the admitted rapist would most likely not be allowed. As potentially biasing as this domain irrelevant context was, if DNA was totally objective it should not have affected their analysis.

In this study we took the original materials used by the DNA examiners that concluded that the suspect cannot be excluded, and presented them to 17 other DNA examiners, 'context free.' These 17 DNA examiners were all expert DNA analysts who were working casework in an accredited governmental laboratory in North America. Fourteen were female and three were male; their mean age was 40.7 (SD = 5.86), and their mean years of experience conducting DNA analysis was 8.9 (SD = 3.96). Two examiners had a BSc, 12 had a MSc (either in biology or forensic science), and 2 had PhDs (one participant did not provide information on their level of education).

We asked the 17 independent DNA examiners to examine the DNA mixture along with DNA profiles of the victim and three suspects (Table 1) (one of the suspects, suspect 3, was the point of interest, as he was determined as 'cannot be excluded' by the DNA examiners who examined his DNA within the potentially biasing context). The evidence presented to them was comprised from the electropherograms (Figs. 1 and 2) available to the original examiners, and included the Vaginal Sperm Fraction (Profiler+) and Vaginal Sperm fraction (CoFiler). They were also provided with the *relevant* contextual information that was provided to the original examiners, such as the concentration of DNA in the sperm fraction extract, the DNA amplification conditions, and capillary electrophoresis injection times. Each of the 17 DNA examiners independently examined the evidence, and gave one of three conclusions for each of the suspects: 'cannot be excluded', 'excluded', or 'inconclusive.'

In regard to suspect three, the results obtained from the 17 independent DNA examiners varied. One examiner concluded that the suspect 'cannot be excluded', 4 examiners concluded 'inconclusive', and 12 examiners concluded 'exclude.' The results are revealing in two respects: First, the fact that the 17 DNA examiners were not consistent in their conclusions, by itself, suggests that there is an element of subjectivity in DNA interpretation. If it was totally objective, then all the examiners would have reached the same conclusion, especially since they all work in the same laboratory and follow the same interpretation guidelines. The observed inconsistencies within the 17 examiners who conducted their analysis on the identical evidence, 'context free,' demonstrated subjectivity in DNA mixture analysis, which may reflect individual differences (e.g., training, experience, personality, and motivation). It is interesting that even using the 'gold standard' [9] DNA, different examiners reach conflicting conclusions based on identical evidentiary data.

Table 1
Suspect 3 portion of the allele chart.

Locus	S3
D3	14, 17
vWA	17, 18
FGA	22, 24
D8	14, 15
D21	28, 28
D18	13, 18
D5	12, 13
D13	10, 14
D7	9, 10
D3	No data
D16	9, 13
THO1	7, 8
TPOX	9, 9
CSF1PO	11, 11

Second, comparing the data between examiners, those from the context free condition to those who were exposed to the extraneous context condition, it is possible that the domain irrelevant information may have biased their interpretation. The DNA analysts who concluded that the suspect cannot be excluded within the biasing context of the criminal case, are in sharp contrast to the vast majority of examiners who examined the same evidence without this biasing context. Only 1 (out of 17) gave the same conclusion as the original analysts, 16 other examiners reached a different and conflicting conclusion (either 'exclude', 12 examiners, or 'inconclusive', 4 examiners). Thus, the extraneous context appears to have influenced the interpretation of the DNA mixture, however, it is always hard to draw scientific conclusions when dealing with methodologies involving real casework.

It must be emphasized, however, that these effects were observed for a DNA mixture analysis. Previous research in forensic identification suggests that contextual influences are most powerful when the evidence is ambiguous, complex, and a 'hard call' [8]. When the data is clear and decisions are simple, then the power of context is diminished. Gill has been quoted to say that "If you show 10 colleagues a mixture, you will probably end up with 10 different answers" [12]. The difficulties and challenges presented by complex DNA mixture have been the focus of several discussions [13–21], and are an important component of 'expert systems' and statistical computing that try to more objectively deconvolute and interpret DNA mixtures [22,23].

The study reported here, the first experimental study exploring DNA interpretation, demonstrates that DNA mixture interpretation has subjective elements and may be susceptible to bias and other contextual influences. Minimizing such potential effects is important, and may include specific training on bias issues, as well as procedures and best practices especially designed to limit contextual influences (such as sequential unmasking [24]).

This study also demonstrates that all types of DNA analysis should not be lumped together as the "gold standard." It is true, that in contrast with many areas of forensic science [25], identity testing using DNA has progressed to the point of general acceptance when complete profiles are obtained from a single DNA contributor [26]. Consistent with this level of acceptance in the scientific community, the courts in the United States and elsewhere equate identity with DNA profiles that include complete allelic data from 13 or more of the standard short tandem repeat loci (STRs). However, in cases where low numbers of template molecules are amplified [27], or where complex mixtures are analyzed, subjective conclusions are made by analysts. This is evidenced by our experiment and the case we discuss, however, one cannot estimate its magnitude and impact without more empirical studies.

The great degree of variability in laboratory methods regarding DNA mixtures has been the subject of concern in the DNA community, and the Scientific Working Group on DNA Analysis Methods (SWGDAM). It is also important to note that while some laboratories in North America still report qualitative results such as "cannot exclude" without quantitative measure, the 2010 SWGDAM guidelines state that "The laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis." [28]

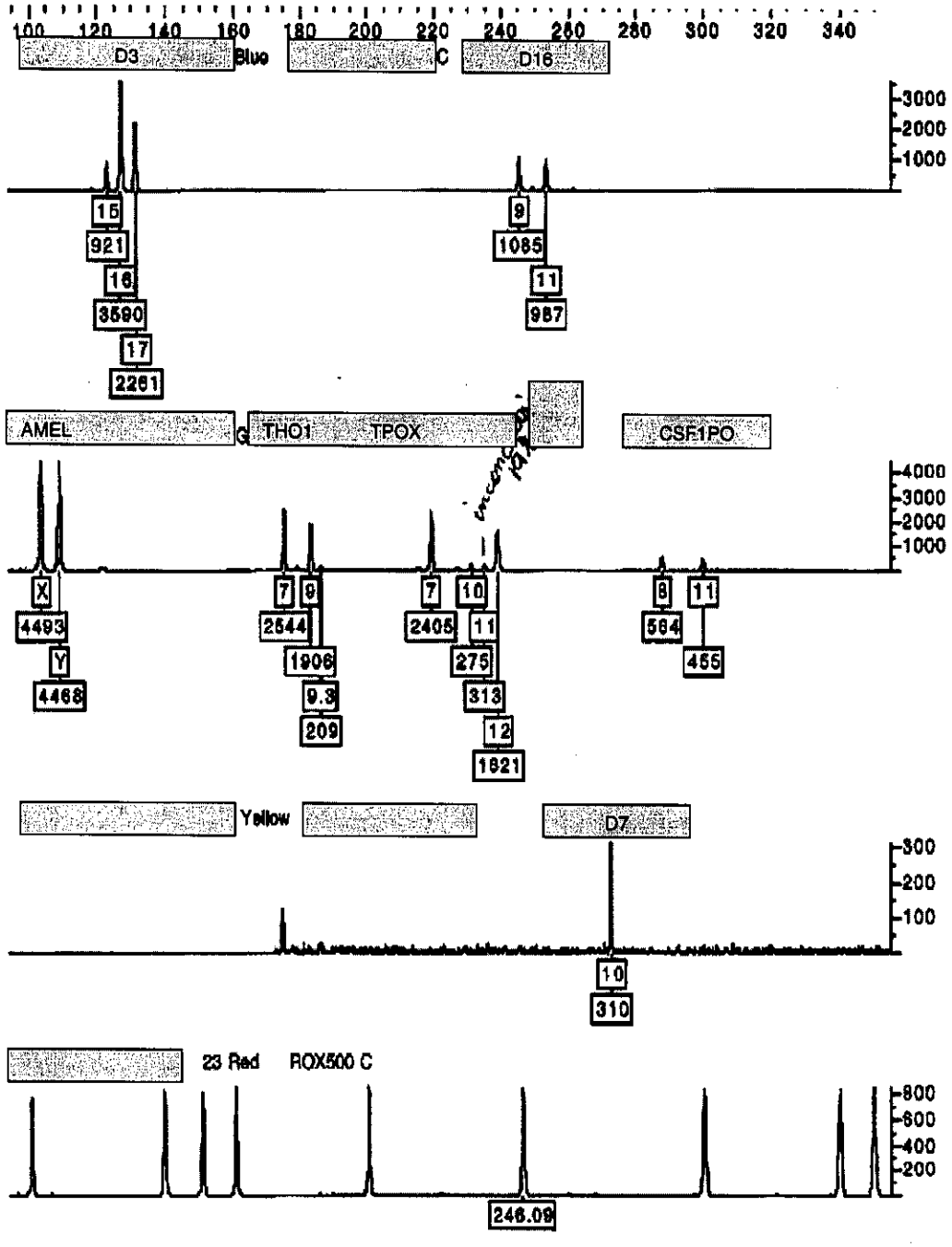


Fig. 1. Sperm fraction electropherogram from victim's vaginal swab, after amplification with CoFiler (ABI). This electropherogram was given to analysts for interpretation. Genetic loci are indicated in boxes above alleles.

These guidelines however are not binding, and are not required for The American Society of Crime Laboratory Directors Laboratory Accreditation Board (*ASCLD/LAB*) accreditation. Outside of North America, the *International Society for Forensic Genetics (ISFG)* DNA commission recommendations on the interpretation of mixtures strongly supports the use of likelihood ratios [16], and this approach is beginning to gain ground in North America.

It is also important to note that while this is the first published empirical study of potential DNA bias, Butler of the NIST laboratories has conducted extensive studies of mixture analysis over several years, wherein he supplies a large number of volunteer laboratories identical DNA mixture data and asks for their analysis. The results of these excellent studies have been presented at conferences and are available at the NIST webpages [29], but have never been published in a peer-reviewed journal.

An interesting and perhaps the most critical point for this paper is that Butler's research findings show that inclusion statistics for the same profiles (using the same data) varied over 10 logs, that is from 1 in 434,600 to 1.18×10^{15}, using the exact same electropherograms [29]. Therefore, although the use of statistics is paramount, it does not resolve the issue of subjectivity and potential bias, the topic of this study.

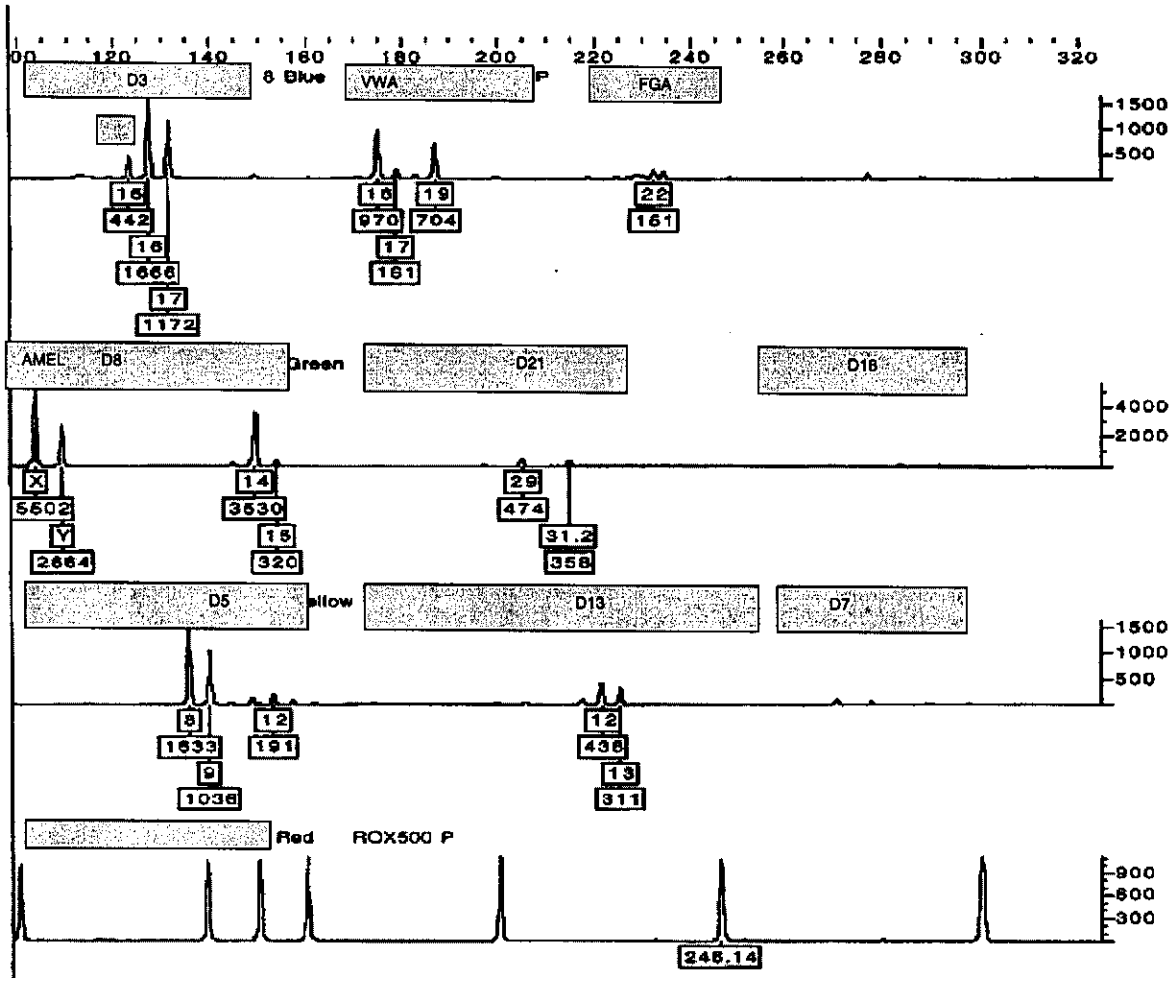


Fig. 2. Sperm fraction electropherogram from victim's vaginal swab, after amplification with Profiler Plus (ABI). This electropherogram was given to analysts for interpretation. Genetic loci are indicated in boxes above allele.

The work presented here is a step in addressing the subjectivity and potential for bias in DNA mixture interpretation. It is clear that additional and follow-up studies are called for. However, acknowledging the role of the human examiner, understanding the role (and weaknesses) of human cognition in making forensic comparisons (including DNA mixtures), is an important step in correctly conceptualizing forensic science and finding ways for improvements [30].

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.scijus.2011.08.004. Additional information can be found at: www.cci-hq.com.

References

[1] R.S. Nickerson, Confirmation bias: a ubiquitous phenomenon in many guises, Review of General Psychology 2 (2) (1998) 175–220.
[2] D.M. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion, California Law Review 90 (1) (2002) 1–56.
[3] I.E. Dror, A. Péron, S. Hind, D. Charlton, When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints, Applied Cognitive Psychology 19 (6) (2005) 799–809.
[4] NAS, Strengthening forensic science in the United States: a path forward, National Academy of Sciences, Washington D.C, 2009.
[5] I.E. Dror, D. Charlton, Why experts make errors, Journal of Forensic Identification 56 (4) (2006) 600–616.
[6] I.E. Dror, D. Charlton, A.E. Péron, Contextual information renders experts vulnerable to make erroneous identifications, Forensic Science International 156 (1) (2006) 74–78.
[7] I.E. Dror, R. Rosenthal, Meta-analytically quantifying the reliability and biasability of fingerprint experts' decision making, Journal of Forensic Sciences 53 (4) (2008) 900–903.
[8] I.E. Dror, S. Cole, The vision in 'blind' justice: expert perception, judgment and visual cognition in forensic pattern recognition, Psychonomic Bulletin & Review 17 (2) (2010) 161–167.
[9] M. Lynch, God's signature: DNA profiling, the new gold standard in forensic science, Endeavour 27 (2) (2003) 93–97.
[10] W.C. Thompson, Subjective interpretation, laboratory error and the value of DNA evidence: three case studies, Genetica 96 (1995) 153–168.
[11] W.C. Thompson, Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation, Law, Probability and Risk 8 (3) (2009) 257–276.
[12] M. Dolan, J. Felch, The peril of DNA: it's not perfect: DNA: genes as evidence, LA Times, December 26 (2008) 2008.
[13] B. Budowle, A.J. Onorato, T.F. Callaghan, A.D. Manna, A.M. Gross, R.A. Guerrierl, J.C. Luttman, D.L. McClure, Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed dNA profiles in forensic casework, Journal of Forensic Sciences 54 (4) (2009) 810–821.
[14] P.M. Schneider, R. Fimmers, W. Keil, G. Molsberger, D. Patzelt, W. Pflug, T. Rothamel, H. Schmitter, H. Schneider, B. Brinkmann, The German Stain Commission: recommendations for the interpretation of mixed stains, International Journal of Legal Medicine 123 (1) (2009) 1–5.
[15] J.S. Buckleton, J.M. Curran, A discussion of the merits of random man not excluded and likelihood ratios, Forensic Science International: Genetics 2 (2008) 343–348.
[16] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures, Forensic Science International 160 (2006) 90–101.
[17] I.W. Evett, P. Gill, J.A. Lambert, Taking account of peak areas when interpreting mixed DNA profiles, Journal of Forensic Sciences 43 (1) (1998) 62–69.

[18] T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Science International 91 (1) (1998) 55–70.
[19] P. Gill, R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker, J. Buckleton, Interpreting simple STR mixtures using allele peak areas, Forensic Science International 91 (1998) 41–53.
[20] B.S. Weir, C.M. Triggs, L. Starling, K.A.J. Stowell, J. Buckleton, Interpreting DNA mixtures, Journal of Forensic Sciences 42 (1997) 213–222.
[21] I.W. Evett, C. Buffery, G. Willot, D.A. Stoney, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, Journal of Forensic Sciences 31 (1991) 41–47.
[22] R.G. Cowell, S.L. Lauritzen, J. Mortera, Probabilistic expert systems for handling artifacts in complex DNA mixtures, Forensic Science International: Genetics 5 (3) (2011) 202–209.
[23] Perlin, M.W, Legler, M.M., Spencer, C.E., Smith, J.L. Allan, W.P., Belrose, J.L., & Duceman, B. W. (in press). Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. (Also presented at the 2010 American Academy for Forensic Science 62nd Annual Scientific Meeting "Casework validation of Genetic Calculator mixture interpretation").
[24] D.E. Krane, S. Ford, J. Gilder, K. Inman, A. Jamieson, R. Koppl, I. Kornfield, D.M. Risinger, N. Rudin, M.S. Taylor, W.C. Thompson, Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation, Journal of Forensic Sciences 53 (4) (2008) 1006–1007.
[25] M.J. Saks, J.J. Koehler, Identification science the coming paradigm shift in forensic, Science 309 (2005) 892–895.
[26] J.M. Butler, Forensic DNA Typing, 2nd Edition, 2005 Elsevier Academic Press, 2005.
[27] N. Gilbert, DNA's identity crisis, Nature 464 (2010) 347–348.
[28] SWGDAM, 4.1 Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories(available at, http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines.
[29] J. Butler, American Academy of Forensic Sciences annual meeting(available at, http://www.cstl.nist.gov/strbase/pub_pres/AAFS2006_mixtures.pdf 2006.
[30] I. E. Dror, Cognitive bias in forensic science, in McGraw-Hill 2012 Yearbook of Science & Technology, McGraw-Hill publishing, in press.

INTERPRETATION OF DNA DATA & MIXTURES

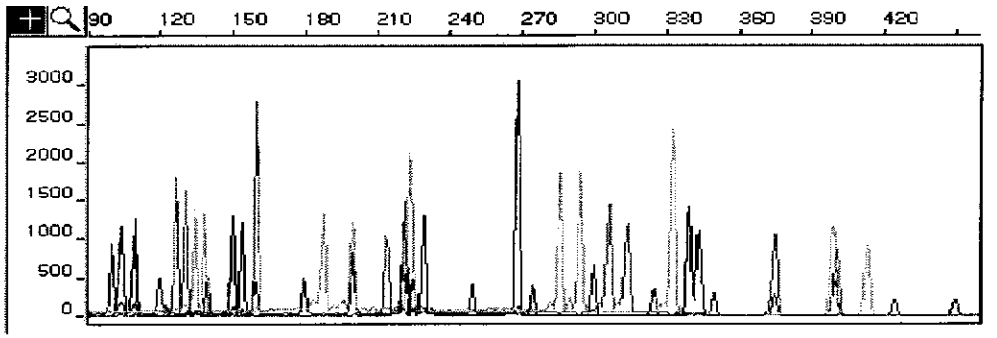
IDENTACODE CONSULTING LLC

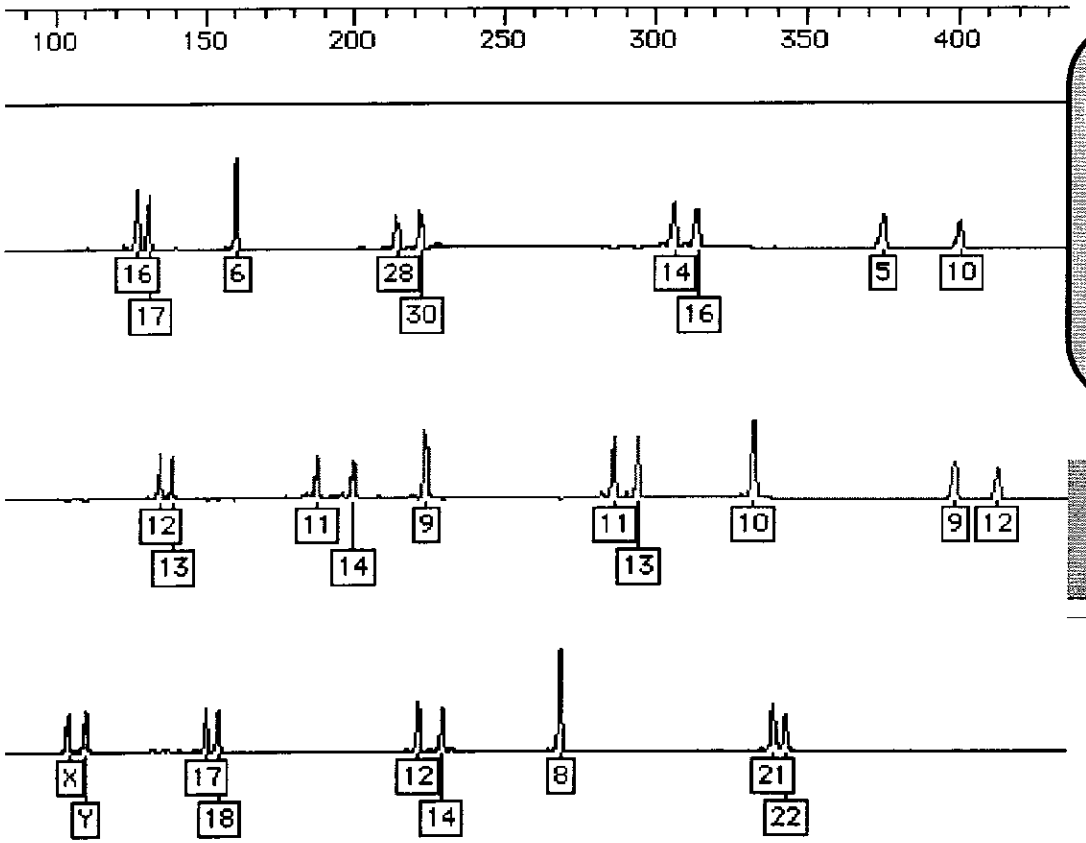
BASIC FORENSIC STATISTICS

- Hardy Weinberg Equilibrium Formula (HWE)
- Random Match Probability (RMP)
- Combined Probability of Inclusion (CPI)
- Combined Probability of Exclusion (CPE)
- Likelihood Ratios (LR)
- DAB Guidelines recommend statistics to give "meaning" or weight to a match of the DNA profile in the evidence when compared to any given individual as a potential contributor – which are most valid?

PROBABILITY THEORY

- These are estimates based on mathematical calculations
- Some require no prior assumptions (RMP, single source)
- Others require assumptions that if incorrect can introduce an error rate (RMP, LR; mixtures)
- Error rates will lead to false exclusion or false inclusion, the latter being most concerning
- Goal: scientific accuracy in classification by DNA





How do we go from raw collected data to processed DNA profiles?

ONCE A DNA PROFILE IS GENERATED, HOW DOES ONE INTERPRET THE DATA?

- DNA data analysis is based on alleles detected
- DNA data can establish if an individual is often likely to have ever been at a scene if sufficient DNA data is present
- DNA data analysis can sometimes tell you exact circumstances based on probative value and how tightly linked the DNA profile is to victim or scene
- DNA data can be an aid to crime scene reconstruction, however, in positioning individuals in certain scene locations

STEPS IN ANALYSIS:

- Evidentiary DNA profiles are analyzed first independently of the known reference sample (this is supposed to prevent bias)
- Was a profile obtained? If not, write the report as no results obtained
- Was a full DNA profile (all the loci obtained)?
- Was a partial DNA profile obtained? (only some of the alleles are present)
- Is the profile a single source? (no more than 2 alleles per locus at all the loci)
- Is the profile a mixture? (evidence at more than 1 locus of 3 or more alleles)

WHY IS DNA ANALYSIS NOT SIMPLE?

KNOWN ARTIFACTS EXIST AND NEED TO BE ASSESSED

SOME ESTABLISHED PCR ARTIFACTS:

- Stutter
- Spectral pull up
- Non-template addition (minus A)
- Amplification product balance
- Stochastic effects (low template quantity)
- Allele drop out (stochastic effect)
- Allele drop in (stochastic effect)
- Key: most occur in a certain location or have a characteristic appearance

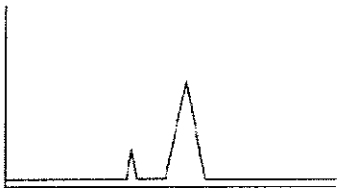
WHAT IS A PEAK?

- A peak is defined by a visible thin apex a few pixels wide that has consistent sharp shape and resolution
- Not every fluorescence above baseline is consistent with being a peak
- Lint, urea crystals, air bubbles etc. can create changes in baseline and must be reviewed by the analyst

What is a Peak?

The location of the allele(s) from the minor contributor in a mixture may be an issue in the interpretation of the data

Is the peak a minor allele or stutter in the stutter position?

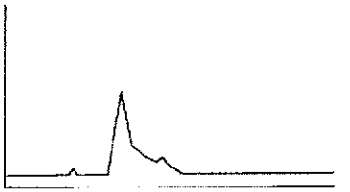


20%

If the height or a stutter peak is 20% of the major allele peak, and observed for stutter for this allele is usually 6.5%, this demonstrates that the stutter position contains a true allele.

The empirically derived stutter percentage data is used in the interpretation of this scenario

Is the peak in the shoulder a minor component to a mixture or machine noise?



Interpretation is based upon how many RFUs the peak is above the shoulder.

If the peak is an allele it's height must be greater than 50 RFU above the shoulder.

raised baseline, shoulder

STUTTER CALCULATIONS AND FILTERS:

- Formula for recognizing stutter is n - 4 where n is the size (bp) of the main peak
- Software filters out allele labeling for all peaks at the n – 4 position that are typically 10% of the value of the main peak (peak height ratio calculation) (10% rule)

Stutter

Stutter – a reproducible characteristic of Taq activity

Stutter is the result of strand slippage during the DNA elongation step of PCR

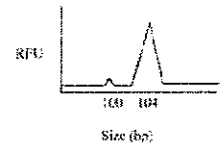
Mechanism:

Taq Polymerase may fall off DNA template during DNA elongation

Extended strand breaks apart

Reanneals out of register

The most common result is a secondary peak of fluorescence at N-4 base pairs.



Stutter is observed at all STR loci used in forensic testing

The percent stutter is empirically derived for each locus by testing known samples

-manufacturer
-by user

General Rules of Stutter

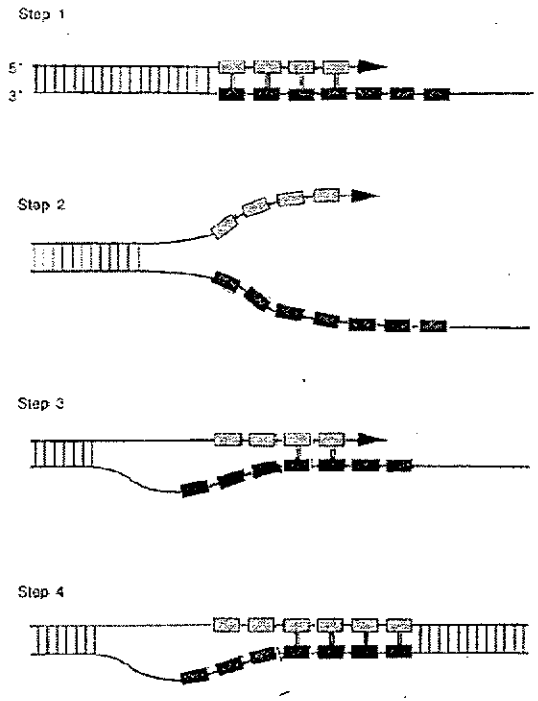
The larger the amplification product the greater the percentage stutter observed

-locus to locus: generally more stutter is observed in the larger loci

-allele to allele: within a locus, the larger alleles will generally have observed stutter

Proposed Mechanism for the Generation of Stutter

Slipped Strand Mispairing Model



Step 1: Taq Polymerase extends through four repeats

Step 2: If the enzyme falls off the extending strand, the DNA molecule can "breath" apart

Step 3: The template and extending strand reanneal with the template strand looped out and in the wrong register

Step 4: Taq Polymerase completes the newly extended strand, which only has six repeats while the template strand has seven

Stutter

Stutter – a reproducible characteristic of Taq activity

Stutter is the result of strand slippage during the DNA elongation step of PCR

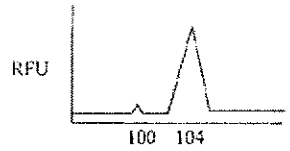
Mechanism:

Taq Polymerase may fall off DNA template during DNA elongation

Extended strand breathes apart

Reanneals out of register

The most common result is a secondary peak of fluorescence at N-4 base pairs.



Stutter is observed at all STR loci used in forensic testing

The percent stutter is imperically derived for each locus by testing known samples

-manufacturer
-by user

General Rules of Stutter

The larger the amplification product, the greater the percentage stutter observed

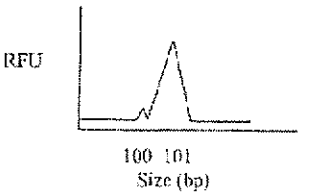
-locus to locus: generally, more stutter is observed in the larger loci

-allele to allele: within a locus, the larger alleles will generally have higher observed stutter

Minus A or non template addition is due to overload of DNA template such that the taq polymerase enzyme cannot add the adenine to the end of the DNA fragment (n – 1)

Minus A (-A)

Taq Polymerase tends to add a single non template A to the end of amplification products



Amplification systems are optimized for the addition of the A

Typically do not see any –A (or a very small amount)

When is –A observed?

-When larger quantities of DNA are amplified.

If the sample is degraded, the effective number of template molecules for the smaller loci will be greater than the larger loci

-When inhibitors of Taq Polymerase are present.

How can –A be minimized? "It's not nice to fight mother polymerase."

Heat soak sample at 60 °C for 90 minutes to facilitate complete A addition.

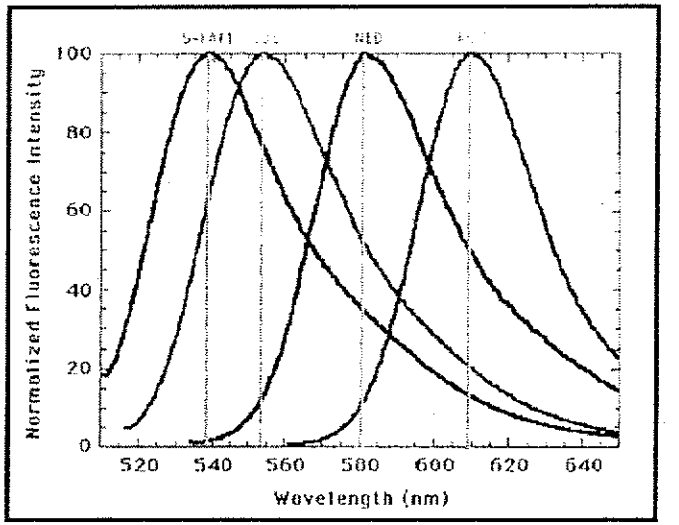
True (variant) alleles predominately contain half-repeats, not n+1 or n-1 base pair.

SPECTRAL PULL UP:

- Collection software has a matrix (virtual filter) that instructs the CCD camera in the DNA sequencer as to what true red, green, yellow etc. represents and gates the collection channel for that color
- Spectral pull up is when sample fluorescence is too high and bleeds through into a different color channel
- This can look like a peak but is not a real peak and data must be reviewed in all 4 color channels*

Spectral Overlap

The light emitted from the fluorescent dyes overlaps into other colors. This overlap needs to be corrected.



GeneScan software uses virtual filters to "look" at specific wavelengths of light for each dye

Gray lines represent the wavelength range used to detect fluorescence from each dye

Wavelength of emitted light from the dyes overlaps (spectral overlap)

Need to correct for overlap

A Matrix File is generated for each different run condition. This file corrects for spectral overlap.

Pull-Up

Incomplete subtraction of spectral overlap results in:

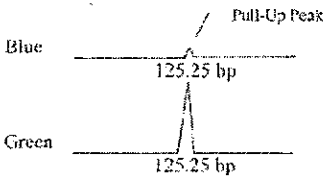
Peak Pull-Up:

A peak (in color A) in the same position as the actual peak (in color B)

-The actual peak is much higher than the pull-up peak

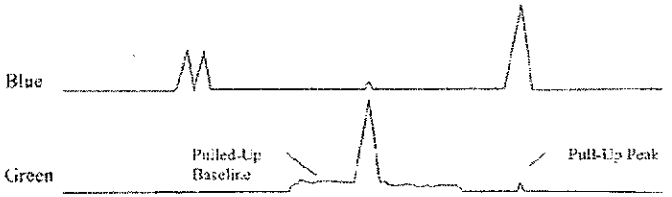
-The greatest level of pull-up is observed in colors with the closest wavelengths

Blue to Green and Green to Blue
Green to Yellow and Yellow to Green
Yellow to Red and Red to Yellow



Baseline Pull-Up:

A region in an electropherogram where the baseline level of fluorescence rises above normal. Baseline pull-up may or may not be accompanied by peak pull-up.



Pull-Up can be corrected by the application of a different Matrix File

- Does not compromise result interpretation.

-The use of different Matrix Files cannot cause the addition or loss of real data peaks, only the reduction or elimination of the pull-up peaks or baseline.

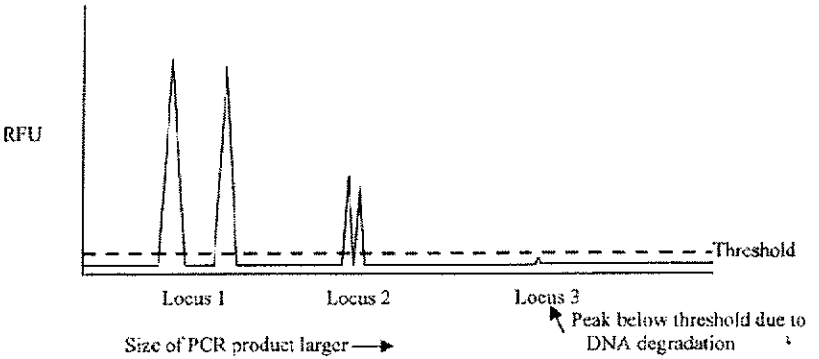
THRESHOLD VALUES:

- Peaks above threshold are valid to report
- Peaks below threshold must be noted in some way to alert the reviewer that another person may be present, however, the amount is so small that the laboratory has no confidence in labeling the allele calls* (potentially exculpatory)
- Individual labs set their own thresholds*

Environmental Insults

Conditions that case material are exposed to can and often do have deleterious effects upon DNA. These the result of these alterations are observed in the analysis of STR data

Degraded Template (DNA)



Result of Degraded Template

-Effective reduction in the number of templates for the larger loci resulting in less amplification product

Potential loss of the larger loci

THRESHOLD VALUES

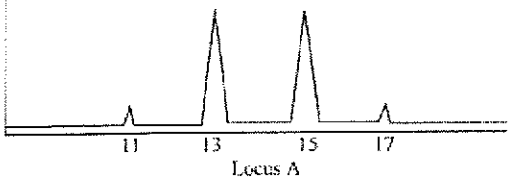
- Differs from laboratory to laboratory
- The minimum threshold used nationwide is 50RFU
- The average baseline for the capillary electrophoresis detection instrumentation is 30RFU
- Some laboratories use a minimum threshold and in addition have a "grey zone" where alleles are called or interpreted with caution (50RFU – 200RFU)
- All laboratories are required to report data if potentially exculpatory
- It is worth it to have a broad review of electronic data (on occasion akin to an audit) or if there is some reason to suspect a second individual is present but not reported (common source of error)
- Many contamination events are discovered by examination of the negative controls or key probative samples in a case

IDENTIFYING SINGLE SOURCE VS. MIXTURES IN DATA

Mixtures

Mixtures are an expected aspect of routine forensic casework

Example 1

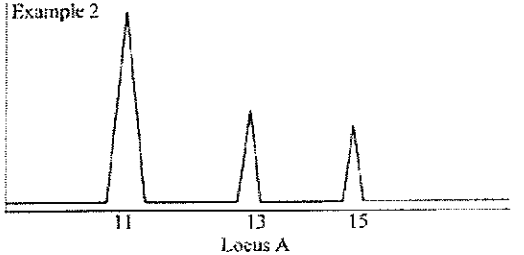


Detection of four alleles – the sample must be a mixture
2 or more contributors

This example can see major and minor contributors
Major 13, 15
Minor 11, 17

Mixture – more than 2 alleles detected at any given locus at more than one locus in a profile

Example 2



How to interpret this result ?

It is a mixture – 3 alleles
What are the major and minor contributors?

11, 12 and 11, 13 individuals
11, 11 and 13, 15
or more than two contributors

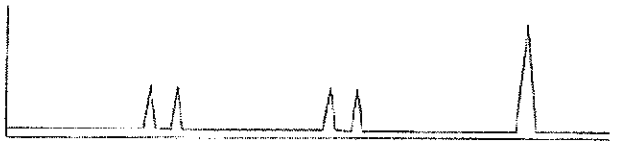
Many laboratories use the heterozygous allele balance and stutter percentage information to separate the genotypes for the contributors to a mixture. The net effect is to report the expected frequency for the knowns, rather than a mixture calculation.

Amplification Product Balance

Desire balance in amplification efficiency between:

Loci
Alleles

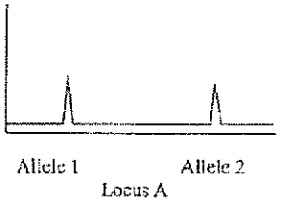
Locus to Locus:



Goals: Heterozygous loci should be approximately equal height: A and B

Homozygous loci should be approximately twice as high as heterozygous loci: B and C

Allele to Allele:



Goal: The smallest allele in a locus should be approximately the same height as the largest

Amount of amplification product is the result of primer binding efficiency

Need primer pairs for the different loci in a multiplex to have similar efficiencies

Kits are designed for allele or peak balance within a certain percent (25-30% on average).

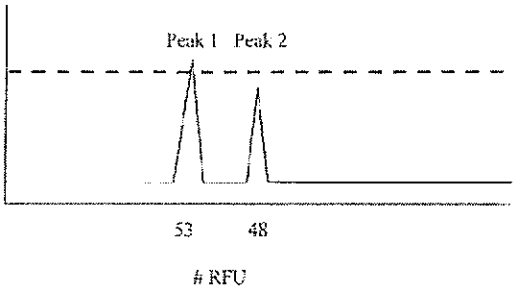
When peak heights show 60% or more imbalance in height, some laboratories will use this to assign values to a major contributor versus a minor contributor (this source attribution can lead to a high percentage of error (10-30% on average based on several studies).

If fragments are assigned incorrectly then an artificial DNA profile is generated.

STOCHASTIC EFFECTS:

- These are fairly straightforward for single source samples
- Stochastic effects are very difficult with complex mixtures*

Stochastic Effects



When amplifying small amounts of DNA, one can observe 1 peak above the threshold and a second below.

However, the peak below threshold cannot be called. Calling just the one peak above 50 RFU would misrepresent the data. The peak at 48 RFU is significantly above background (noise).

Reporting:

If the observed peaks match the defendant:

Call the locus inconclusive

If the observed peaks do not match the defendant:

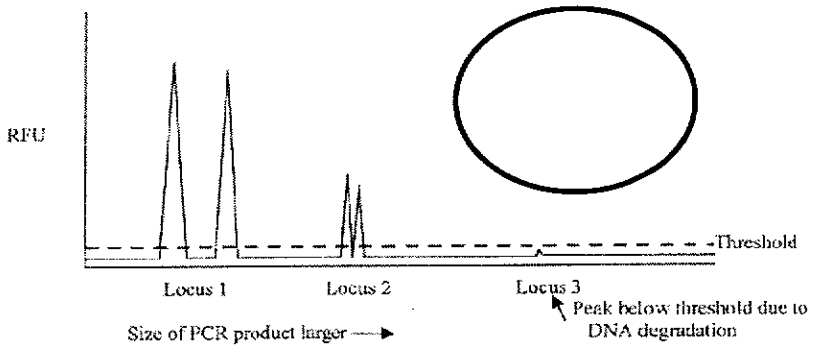
Report that the defendant is excluded

DNA Degradation

Environmental Insults

Conditions that case material are exposed to can and often do have deleterious effects upon DNA. These the result of these alterations are observed in the analysis of STR data

Degraded Template (DNA)



Result of Degraded Template

-Effective reduction in the number of templates for the larger loci resulting in less amplification product

Potential loss of the larger loci

If software uses peak height to establish number of contributors, what happens when peaks are not consistent with amount of DNA due to breakage?

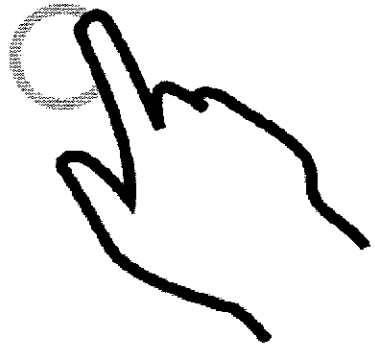
This is why it is best to say >2 contributors, scientists really can't with certainty ever tell you more than that. Studies have shown that with controlled mixtures, 3, 4, 5, 6 sources and more all begin to look the same.

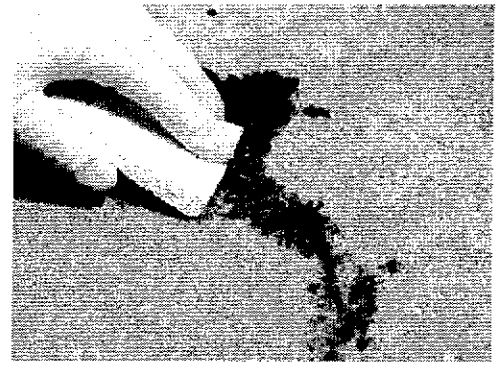
TYPES OF DNA IN CASEWORK

- Cases with no detectable DNA
- Cases with partial DNA profiles only
- Cases with full single source DNA profiles only*
- Cases with two person mixtures (deconvolute or not)
- Cases with three person mixtures (deconvolute or not)
- Cases with more than three people in the mixture

* Most discriminating data by statistics; loss of information (partial profiles) or gain of information (complex mixtures) reduces the statistical probability of uniqueness

Gloves



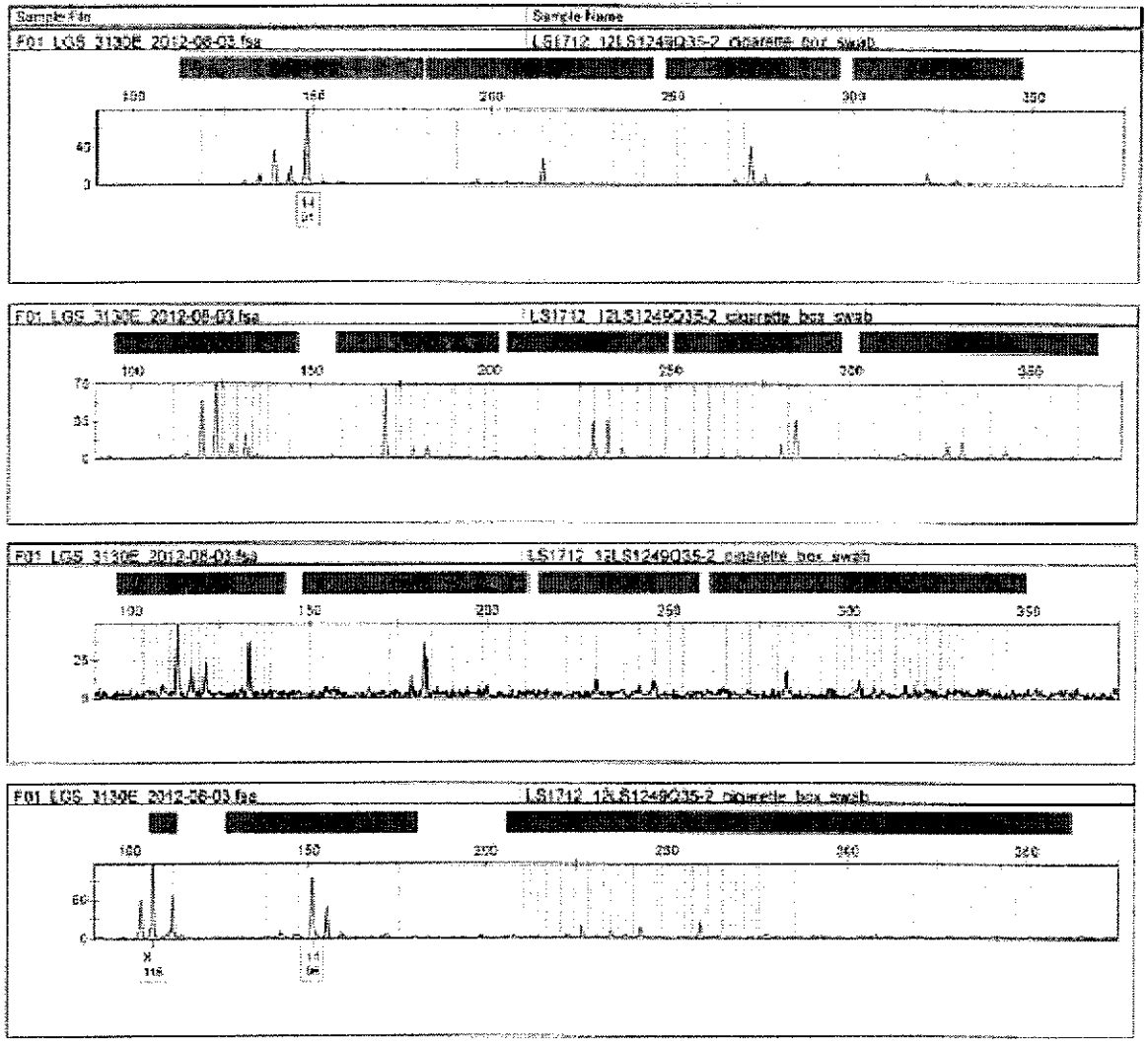


Cleaned scene

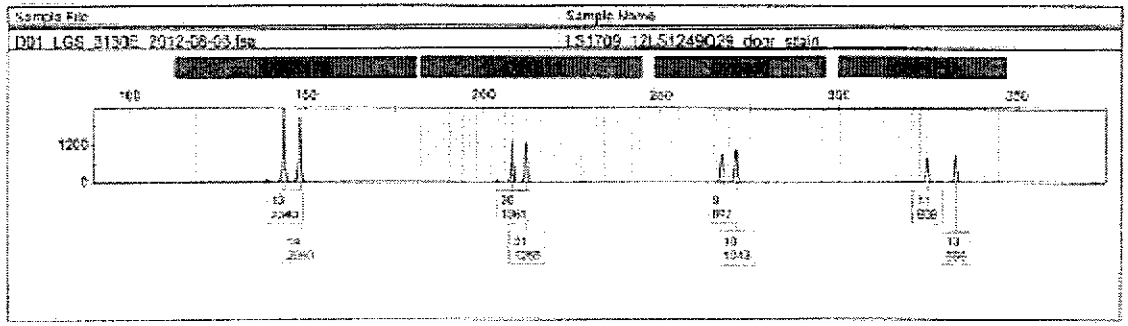
Less than 30 seconds contact time
or non-shedder status

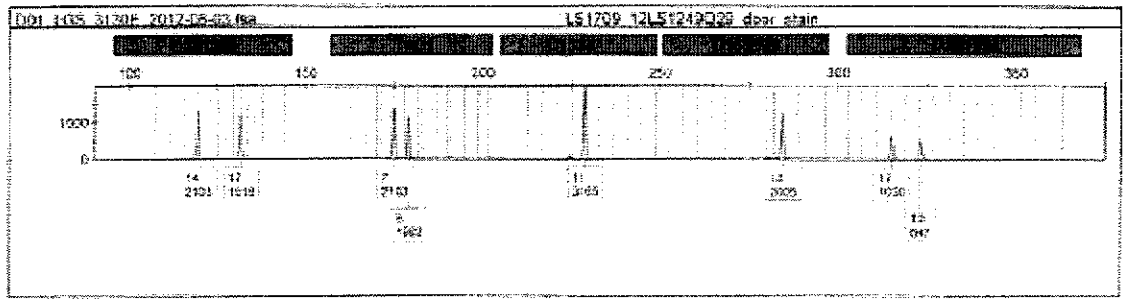
IS IT POSSIBLE TO HAVE TOUCHED AN OBJECT AND LEAVE NO DNA?

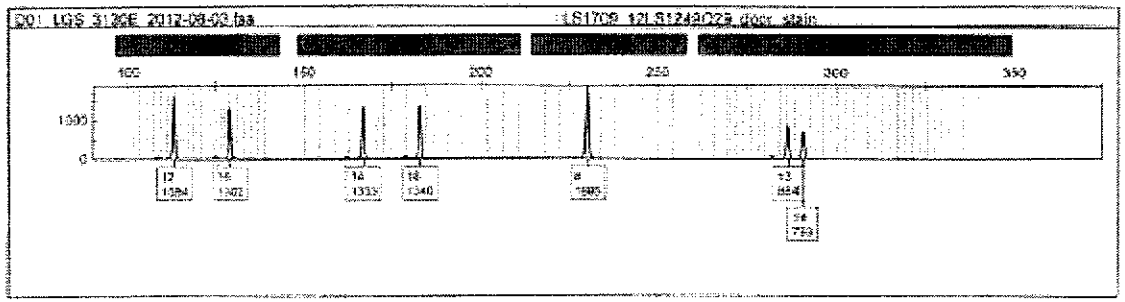
OF COURSE, LACK OF EVIDENCE DOES NOT MEAN AN INDIVIDUAL WAS THERE BUT THERE IS NO PROOF EITHER

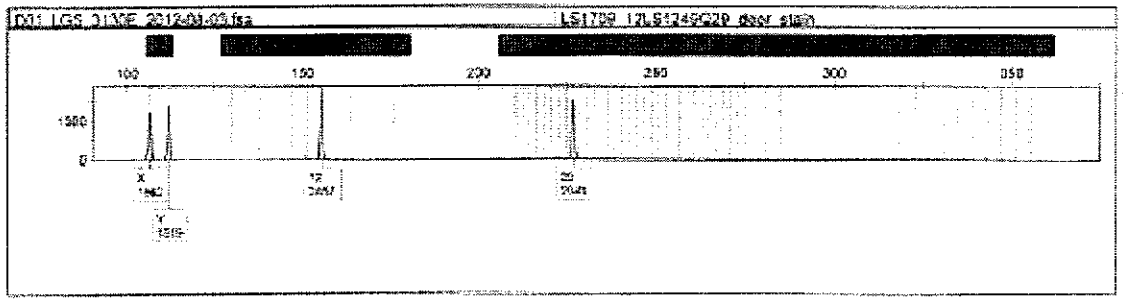


low level mixture, not very informative; inconclusive or call alleles but would have a high statistical match probability due to coincidence

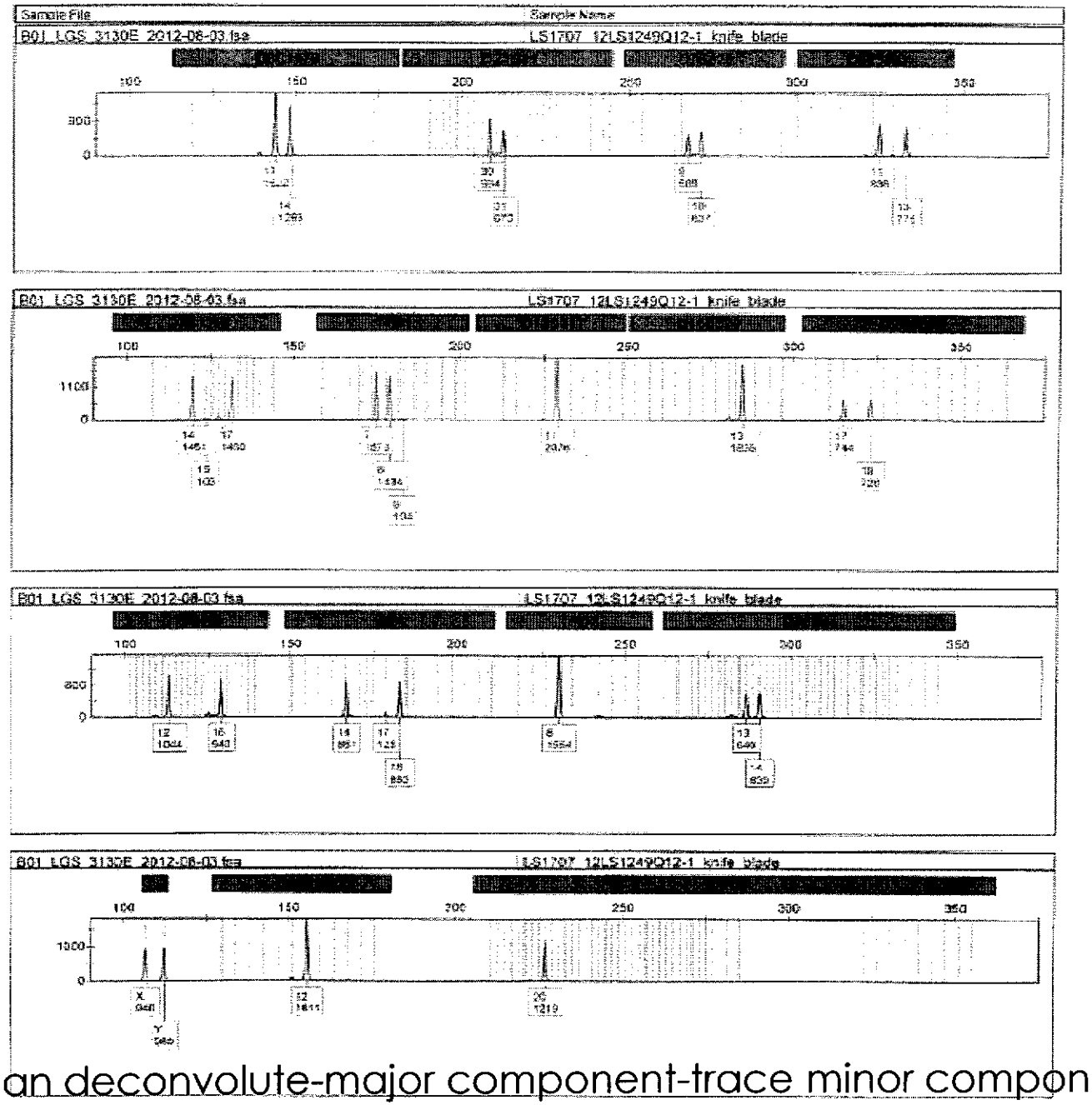








clean single source DNA profile



Can deconvolute-major component-trace minor component

- Peaks per locus
- Peak height ratios
- Peaks in baseline
- Degraded sample
- with drop out
- 2, 3, 4 or more?

difficult to deconvolute

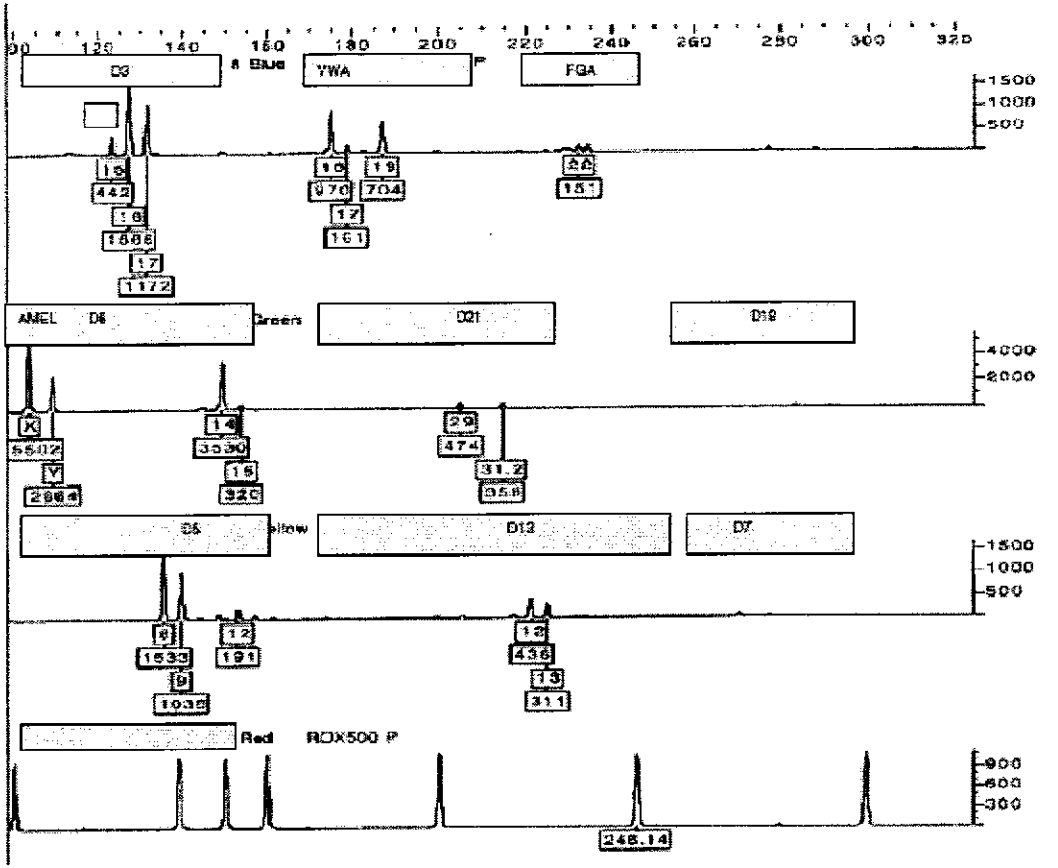
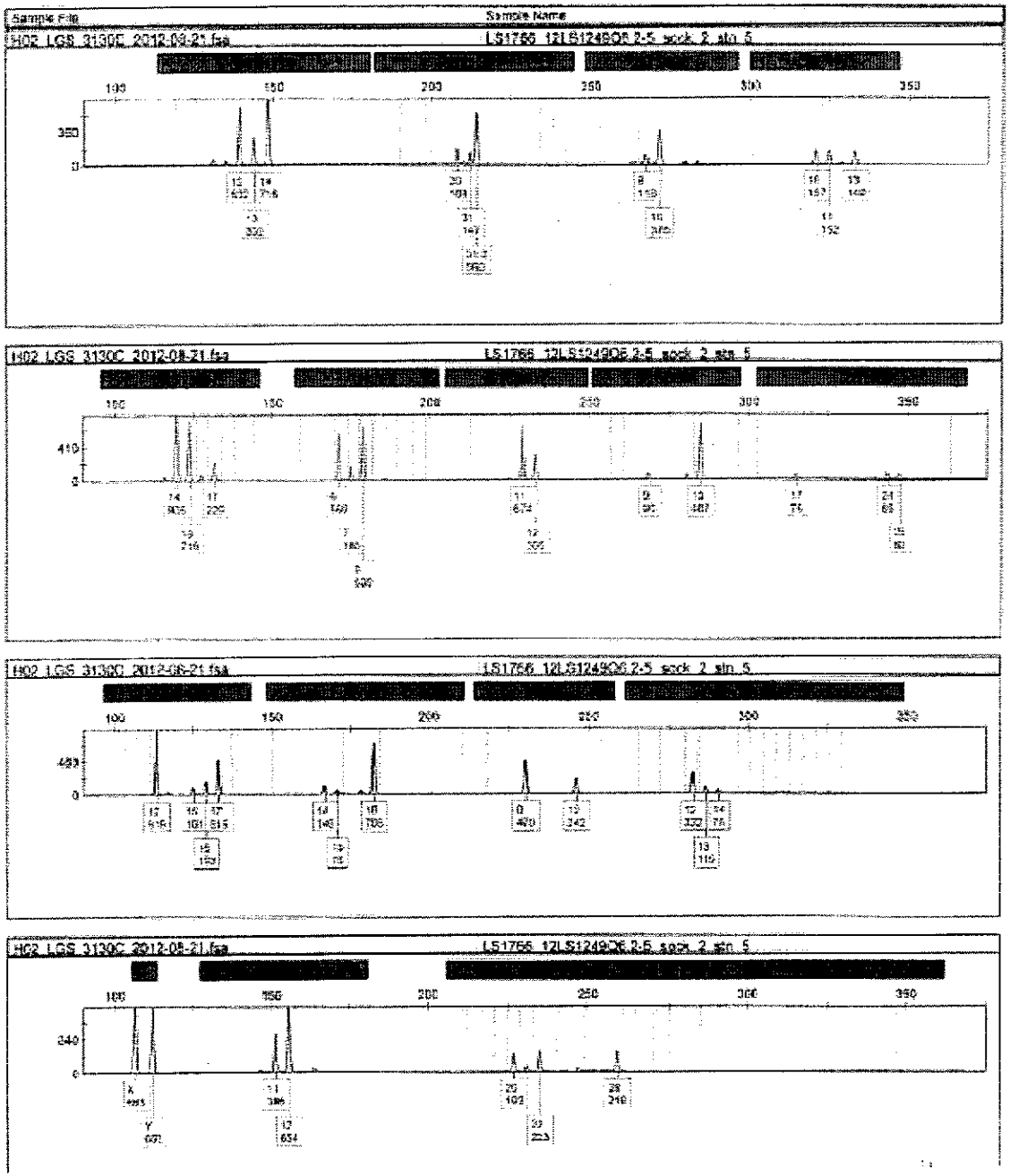


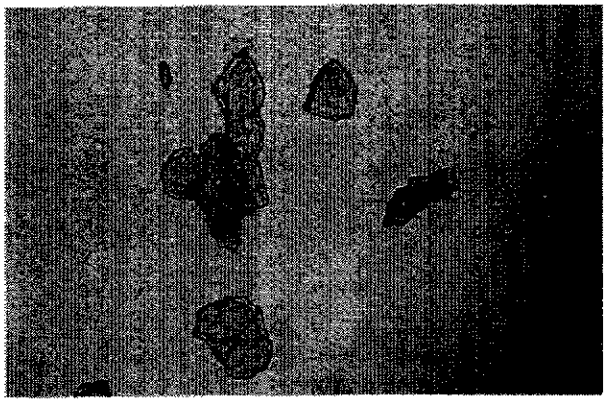
Fig. 2. Sperm fraction electropherogram from victim's vaginal swab, after amplification with Profiler Plus (ABI). This electropherogram was given to analysts for interpretation. Genetic loci are indicated in boxes above alleles.

HOW MANY PEOPLE COULD BE IN A COMPOSITE DNA PROFILE?

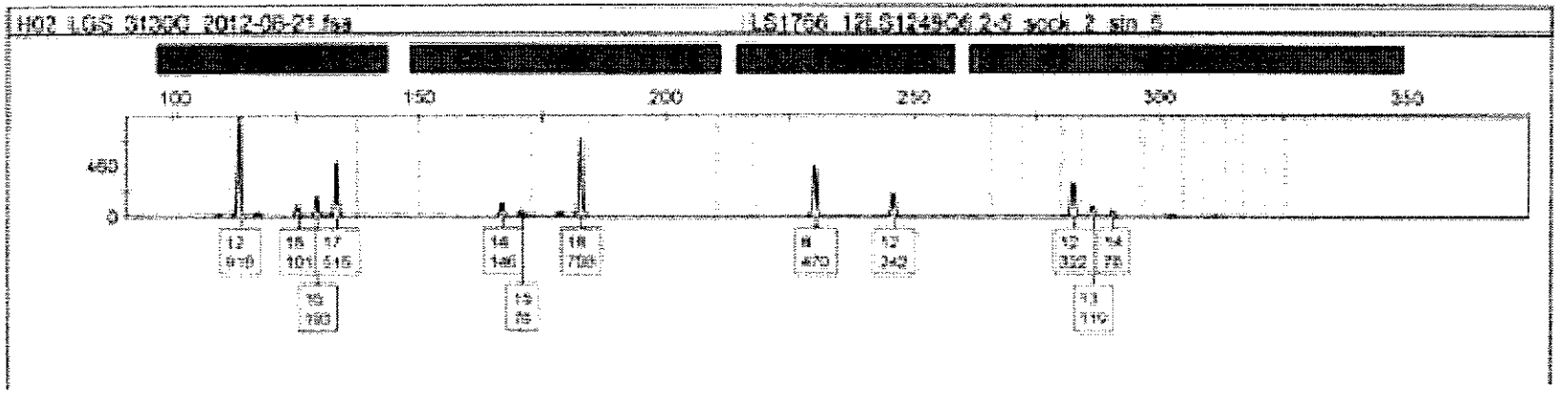
ASSUMPTIONS MADE BASED ON NUMBER OF PEAKS DETECTED



complex mixture; deconvolute or not?

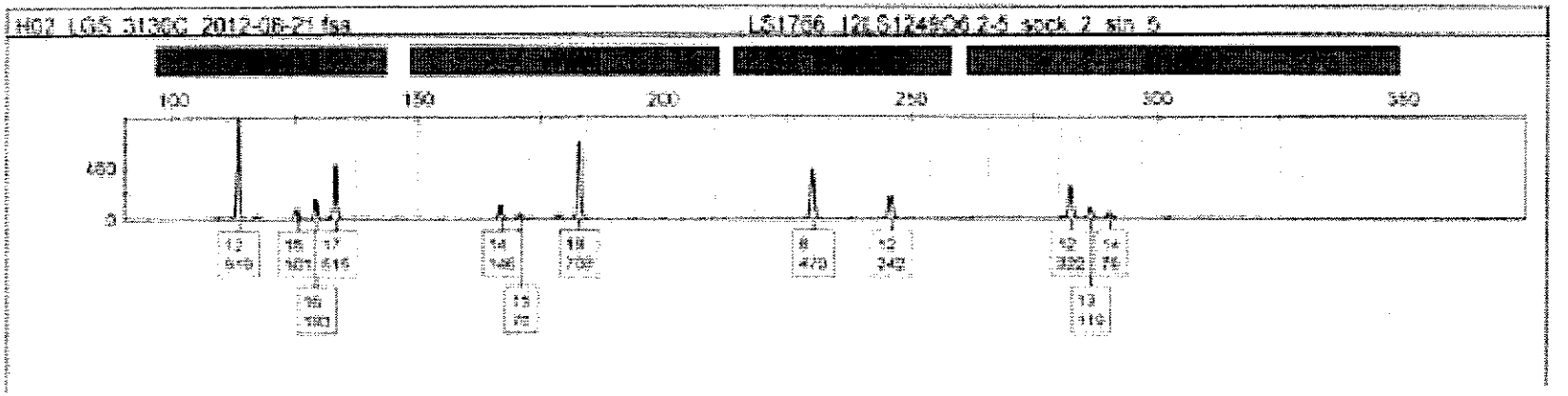


coincidental matching – like the birthday rule – some individuals by coincidence can have the same allele values; genetic relatives and complex mixtures have the highest rate for coincidental matching



IS IT POSSIBLE TO BE INCLUDED IN A MIXTURE ARTIFICIALLY?

SURPRISINGLY, THE ANSWER IS YES

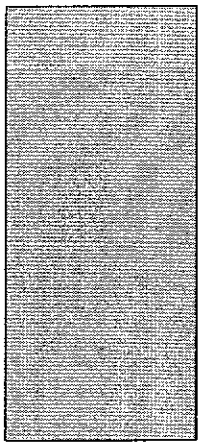


Individual #1: 12, 17; 18, 18; 8, 12; 12, 12
Individual #2: 15, 16; 14, 15; 8, 12; 13, 14
Individual #3: 15, 17; 18, 18; 8, 8; 12, 14

- coincidental match rate due to source attribution error can be as high as 1/900-1200 in some mixture studies – this is for non-contributors; i.e. individuals who were never actually present at the scene. Error rate is important when giving this data to a jury for establishing weight of the evidence.

FORENSIC STATISTICS

HEATHER MILLER COYLE, PHD



BASIC FORENSIC STATISTICS

- Hardy Weinberg Equilibrium Formula (HWE)
- Random Match Probability (RMP)
- Combined Probability of Inclusion (CPI)
- Combined Probability of Exclusion (CPE)
- Likelihood Ratios (LR)
- DAB Guidelines recommend statistics to give "meaning" or weight to a match of the DNA profile in the evidence when compared to any given individual as a potential contributor – which are most valid?

PROBABILITY THEORY

- These are estimates based on mathematical calculations
- Some require no prior assumptions (RMP)
- Others require assumptions that if incorrect can introduce an error rate (RMP, LR)
- Error rates will lead to false exclusion or false inclusion, the latter being most concerning
- Goal: scientific accuracy in classification by DNA

ASSUMPTIONS INHERENT IN STATISTICAL CALCULATIONS

- Independence of loci (on separate chromsomes)
- No significant population substructure
- All contributors of same race (admixture)
- All unrelated (some level of relationship in databases does exist)
- No allele dropout
- No intensity (peak height) differences
- Defined hypothesis (accurate assessment of no. of contributors)

- If the original premise is not true-then how valid is the calculation?

COMPARISON OF CPI VS. LR

- *Combined Probability of Inclusion (CPI)* -All inclusive approach where no assumptions are made about the number or identity of an individual(s) contributing to the DNA profile in the evidence. [*who knows how many people in the net?*]
- No assumptions are made regarding genotype assignment of major versus minor contributors. (qualitative statement with statistic)
- Alternatively, you can attempt to assign data to contributors based on peak height values in some circumstances. This choice is left to the laboratory to decide. ASCLD enforces lab policy: "say what you do; do what you say". Choice of how this is accomplished is left up to individual Lab Directors but encouraged to follow TWGDAM guidelines.

- *Likelihood Ratio (LR)* –Several assumptions must be made regarding contributors. OCME uses a 2 or 3 person assumption in their formula, but there could be many, many more individuals contributing.[*3 people are in the net, maybe more, maybe a lot more]*
- Two versions are presented; *Hp* = includes the defendant + 2 unknown unrelated individuals, *Hd* = 3 unknown unrelated individuals

COMMONLY USED FORMULAS

Random Match Probability: formula: (RMP) = 1/profile frequency, also called the frequentist approach (gives an estimated frequency)

Combined Probability of Exclusion (CPE): estimates the proportion of a population that would be excluded from an observed mixture (conservative, makes no assumptions re: number or identity of contributors)
Combined Probability of Inclusion (CPI): CPI = 1 - CPE

Likelihood Ratio (LR):, two propositions- one of which is the alternative of the other:
Hp = DNA profile came from defendant
Hd=DNA profile came from some other person
formula: LR = Pr(E/Hp)/Pr(E/Hd)

Bayesian Approach (BA), prior odds x likelihood ratio = posterior odds:
formula: [PrHp/PrHd] x Pr(E/Hp)/Pr(E/Hd)

WHAT'S THE IMPACT OF USING CPI VS. LR?

Combined probability of inclusion/exclusion calculations allow for inconclusive results; phrases such as:

- could be included
- cannot be excluded
- included but low or high estimate of rarity 1/ "x" no. of individuals in population of any given ethnicity (we presume no knowledge of ethnicity only that DNA was detected)
- Inconclusive (can't be determined)
- Un-interpretable (extremely complex mixtures of 4+ mixtures) – differs in forensic community vs. OCME protocol for how many individuals they consider interpretable

WHAT'S THE IMPACT OF USING CPI VS. LR?

- Likelihood ratios (LR) are not used as frequently in DNA statistical calculations (a few laboratories across the country) due to the inability to leave a DNA sample in the inconclusive category.
- Effectively, this decreases scientific accuracy and increases the error rate for false inclusions and false exclusions.

LR INTERPRETATION

NYC OCME

May 24, 2012

Option	Numerator (Prosecutor's Hypothesis)	Denominator (Defense Hypothesis)
1	Comparison	Unknown
2	Comparison + Known	Known + Unknown
3	Comparison 1 + Comparison 2	2 Unknowns
4	Comparison + Unknown	2 Unknowns
5	Comparison + 2 Unknowns	3 Unknowns
6	Comparison + Known + Unknown	2 Unknowns + Known
7	Comparison + 2 Knowns	Unknown + 2 Knowns

Table 1A. Numerator and denominator options available in LR software. "Known" refers to an elimination profile from an individual who is assumed to be a contributor to the evidence sample. "Comparison" refers to the comparison profile of interest (often the suspect). "Unknown" refers to a randomly selected individual from a population of individuals that are unrelated to the Known, Comparison, or to one another.

COMMON ERRORS ON REVIEW

ALLELE SHARING RESULTS IN COINCIDENTAL MATCHING

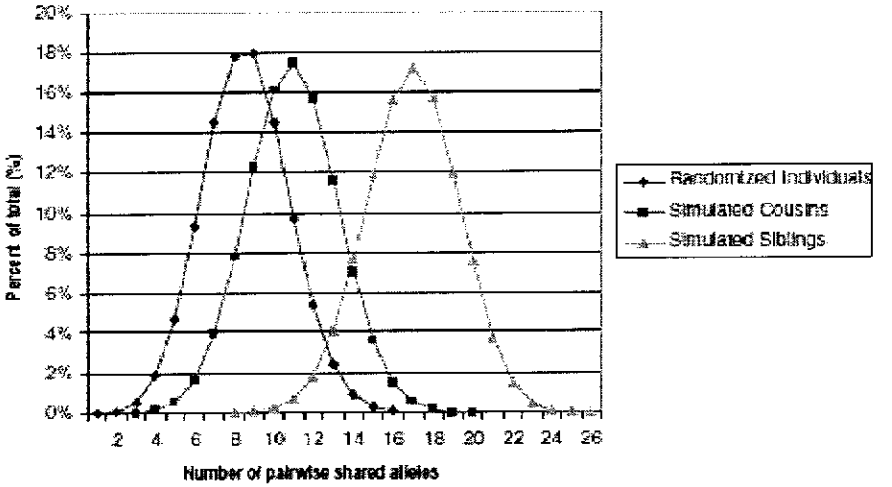


FIG. 3—*The distributions for the number of shared alleles amongst all possible pairs of synthetic individuals, pairs of synthetic siblings, and pairs of synthetic cousins. A total of 459,361 pairs each of randomized individuals, simulated cousins, and simulated siblings were considered.*

Reference: Paoletti et al. 2005. JFS. 50:6, 1-6

SOURCE ATTRIBUTION ERROR

- Coincidental matching

- Study of laboratories (anonymous):
- major contributor in mixture misclassified 10-13% of the time
- minor contributor misclassified 13-33% of the time
- some laboratories choose not to deconvolute mixtures for this reason

Ladd et al. 2001. CMJ 42:3, 244-246.

ALLELE SHARING

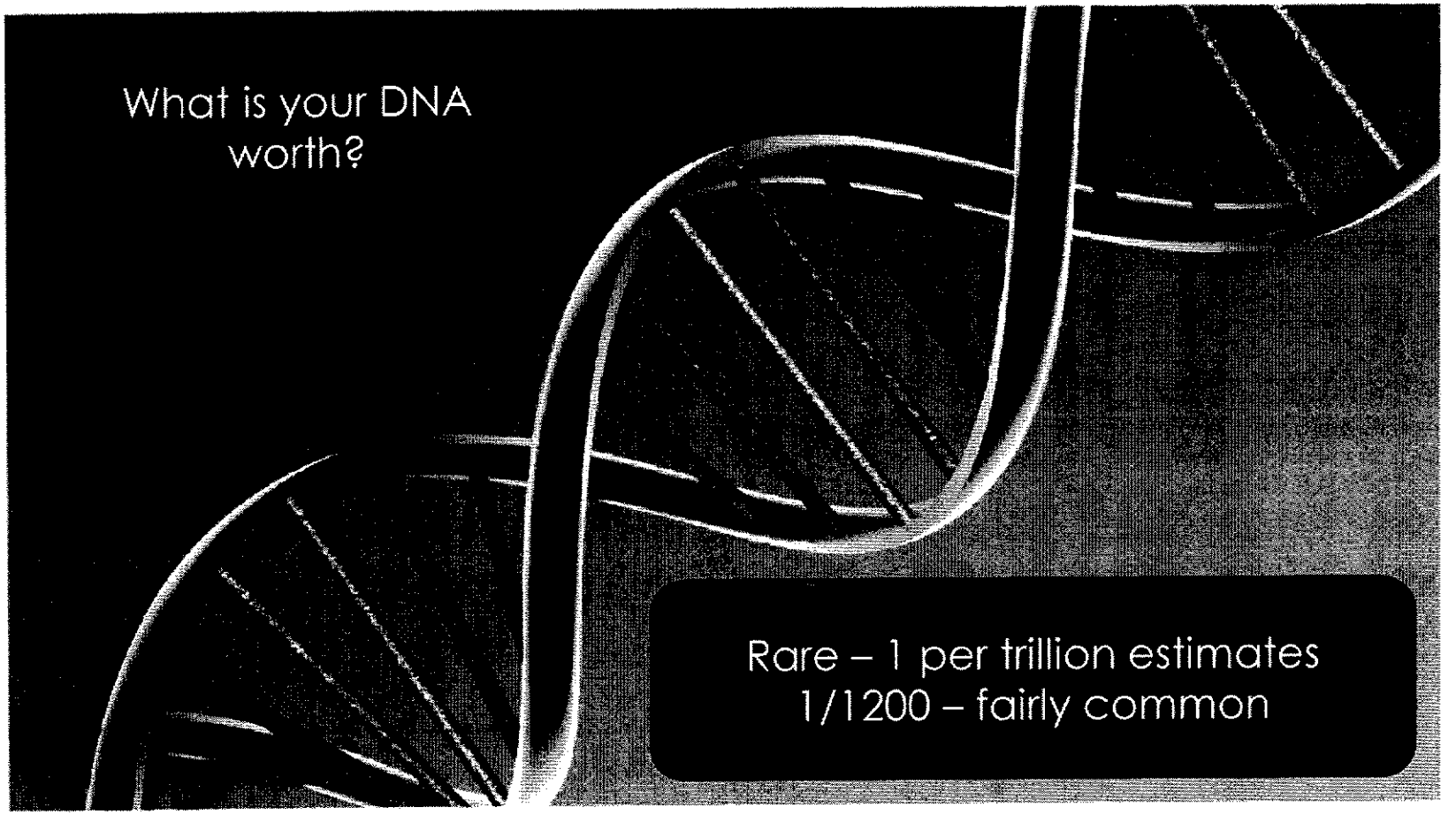
- Coincidental matching

163 other individuals could be included in a 3 person mixture when searched against the reference DNA database; estimates approximate coincidental match rate is 1 per 3000 individuals

Reference: trial testimony – FST software
People State of New York v. Devon Thomas, 2013

INCORRECT SAMPLING IN DATABASE

- Did you know there is variation both within and between ethnic groups across the US?
- Original articles of sampling in different states led to the FBI database.
- Sampling or calculations using different population databases can give very different estimates of RMP or LR so realize that these statistical calculations are very rough approximations.
- Statistical calculations that are on the margin of include or exclude or weakly associated in LR calculations can easily be affected by ethnicity or database sampling issues.



IF DNA IS A LARGE PART OF CASE OR CRITICAL PROBATIVE EVIDENCE..

FALSE INCLUSION RATES NEED TO BE CAREFULLY EVALUATED

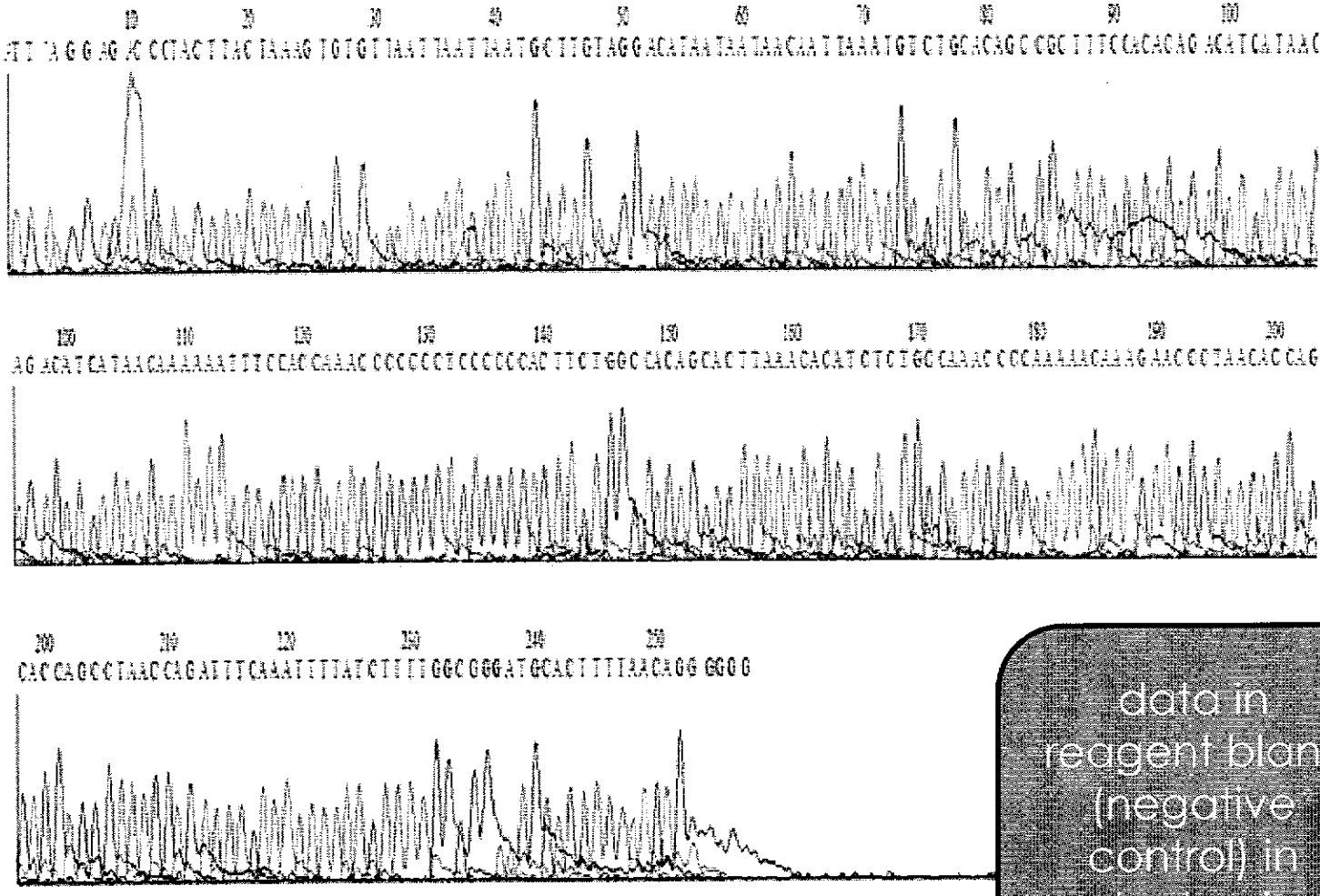
Item. No.	Description
1	Old Timer brand, pocket knife w/ 3 blades, used, scratched within bag with many other items **MFID 1433 18 May 12 AAN doc 165744 120-12**
1	SOG brand, 4.5 inch blade, used **MFID 2011 14 May 12 AAN doc 170855 061-12**
1	Gerber brand, 3 inch blade, serrated **MFID 1605 14 May 12 SBS doc 170855 066-12**
2	Box cutter, super knife brand, 2.5 inches, used **MFID 2020 14 May 12 AAN doc 170855 061-12**
2	Case, contains pocket style knife, Berkley brand, 4 inch blade **MFID 1606 14 May 12 SBS doc 170855 066-12** cross-reference to item. no. 23
3	Ozark Trail brand, 5 inch blade, used **MFID 2021 14 May 12 AAN doc 170855 061-12**
3	Scalpel, Miltex by Kai manufacturer, used **MFID 1607 14 May 12 SBS doc 170855 066-12**
4	Camillus brand, 3.5 inch blade, used **MFID 2022 14 May 12 AAN doc 170855 061-12**
5	Box cutter, Hesco-Bastion brand, 2.5 inch blade, used **MFID 2023 14 May 12 AAN doc 170855 061-12**
6	Unknown brand, knife, 3.75 inch blade, used **MFID 2024 14 May 12 AAN doc 170855 061-12**
6	Ti-Lite brand, 8 inch blade, scratched **MFID 1728 16 May 12 AAN doc 165744 100-12**
7	Mt Rainier brand, 2.75 inch blade, used **MFID 2025 14 May 12 AAN doc 170855 061-12**
8	Unknown brand, 2.75 inches, used **MFID 2026 14 May 12 AAN doc 170855 061-12**
9	Carson design brand, 4 inch blade, used **MFID 2027 14 May 12 AAN doc 170855 061-12**
10	Ax head, unknown brand, blade length 4 ¾ **MFID 2028 14 May 12 AAN doc 170855 061-12**
11	Unknown brand, 3.25 inches, used **MFID 2029 14 May 12 AAN doc 170855 061-12**
12	Previously marked as 11, knife 28, DN061-12-DNA/LP cross-reference to item no. 28
19	Carson design brand, 5 inch blade, rusty & scratched **MFID1149 18 May 12 AAN doc 165744 115-12**
19	Unknown brand, 2.75 inch blade, rusty **MFID 2038 14 May 12 AAN doc 170855 061-12**
20	Gerber brand, 4.5 inches, used **MFID 2039 14 May 12 AAN doc 170855 061-12**
23	Previously marked 22 and case, knife within case, item 1, DN 066-12-LP cross-reference to item no. 2
28	Gerber brand, 3.25 inch blade, red stains in internal portion of the handle **MFID 2042 14 May 12 AAN doc 170855 061-12**
31	Gerber brand, tool (utility), scratched, used, **MFID 1309 18 May 12 AAN doc 165744 115-12**

Other common errors:

confusing chain of custody

incorrect voucher numbers

multiples of similar or same items



060910 formamide 004_B06 contamination (trace)

data in reagent blank (negative control) in electronic data

RECENT ISSUES AT OCME - NYC

- Serologist failed to correctly identify semen stains (which fluoresce brightly under alternate light source) and placed evidence back into incorrect cases after examination.
- Body from morgue accidently dumped on street.
- Deputy Director, other supervisory staff, quality manager resigned/fired in wake of errors and lack of quality control at laboratory.
- Management issues are one issue; even more importantly quality assurance needs to be guaranteed for the legal system to be functioning correctly.
- If you have an issue or potential issue, please bring it forward so it can be addressed.
- Issues are being posted via the NY State Commission website – main goal is disclosure so errors can be prevented; and post-conviction issues can be addressed.