International Research Journal of Computer Science (IRJCS)
Issue 5, Volume 2 (May 2015)
ISSN: 2393-9842
www.irjcs.com

Sources of Computational Error in Probabilistic Genotyping Software Used for DNA Mixture Interpretation

Heather Miller Coyle*
Forensic Science Department, University of New Haven

Abstract— Use of DNA for human identification is considered a gold standard for criminal and civil casework. Although DNA is powerful and convincing technology, there is an inherent error rate associated with DNA mixture analysis methods whether computed manually or with software. The use of probabilistic genotyping software programs for the analysis of complex DNA mixtures is gaining momentum in certain regions of the United States and little information exists in the published literature on sources of error in establishing true contributors to DNA mixtures as compared to false positive matches from non-contributor reference DNA databases. On review of a forensic software program called forensic statistical tool or FST, some factors contributing to the high error rate have been identified as (a) percentages or peak height ratios used to establish major and minor components of a mixture, (b) choice of analytical thresholds and (c) the empirically derived allele drop-in rates (contamination events) and drop-out rates. All potential pair-wise comparisons of allele combinations must be considered for each locus and it is possible using certain computational parameters to artificially match an individual who is not the source of the evidence at some estimated probability which becomes the error rate for the method. Based on a brief survey of different computational programs, the error rates for two person DNA mixtures range from 0.005% (e.g. TrueAllele) to 0.02% (e.g. FST). With an increase in the number of contributors to the sample to three, there is a corresponding increase in the error rate (0.08%) with the FST software analysis as there are greater numbers of permutations or possible combinations of allele arrangements.

Keywords— probabilistic genotyping, DNA, human identification, error rates, mixture interpretation

I. INTRODUCTION

Computational analysis of DNA fragments derived from crime scene evidence requires a software package to identify the midpoint of detectable peak generated by fluorescence from a DNA fragment migrating through a gel polymer solution during capillary electrophoresis. The midpoint is based on peak height and peak area and once identified, the peak is placed into a virtual software bin and labelled with a size reflecting nucleotide bases. Peak morphology must be sharp and of high quality for sizing of the DNA fragment to be scientifically accurate. DNA profiles that are from a single individual are fairly straight forward to evaluate by eye and also with genotyping software programs such as GeneMarker HID (SoftGenetics LLC; State College, PA) and GeneMapper (Life Technologies; Grand Island, NY) that automatically size DNA fragments against co-injected commercially available and quality controlled size standards. DNA mixtures, however, are more challenging due to peak height differences that could originate from different ratios of DNA from each of the human contributors to the sample or from preferential PCR amplification efficiency in the enzymatic copying process that occurs before the software analysis step. Although sized accurately between 50 – 500 bases on the x-axis, the level of fluorescence or DNA quantity for each contributor is more challenging to accurately quantify on the y-axis and is a major contributing factor to the error rate for evaluating which DNA fragment is associated to each potential contributor. Older manual methods of interpreting DNA mixtures included establishing major and minor contributors based on peak heights when sufficient visual differences were observed on the y-axis; or labelled as inconclusive or uninterpretable when sufficient distinction between major and minor contributors did not exist.

The use of probabilistic genotyping software is a way of reducing cognitive bias which may have been present using manual methods by weighting each DNA fragment mathematically independent of examiner bias which tends to include rather than exclude an individual though the bias may be subconscious. The improvement in software interpretation should then be measured against estimates of source attribution error for methods which based on one study can range between 10-13% for major contributors and 13-33% for minor contributors to the DNA mixture [1]. In addition, the manner in which false positives or wrongful inclusion of noncontributors to a DNA mixture has been assessed is a DNA database of known human DNA genotypes is assembled and DNA mixtures are interpreted and then cross-referenced to the non-contributor database to establish how often a positive match using likelihood ratio statistics (LR) is achieved. This method will give an estimate of error as the database contains only individuals that are known not to be sources of the DNA and thus are included as potential contributors by coincidence (coincidental match rate). Paoletti et al. [2] estimates that on average 18% of alleles can be shared by coincidence when comparing DNA profiles of unrelated individuals. The published literature for software validation uses different reference DNA databases rather than one standardized set of knowns which is one explanation for the wide range in published error rates [3], [4]. Software packages also have different algorithms and assumptions made for the analysis which can result in variation in likelihood ratios for the same DNA evidence. Establishing the false positive rates for each software package is still, however, useful as a baseline benchmark of error for providing truth in testimony for DNA mixture analysis in the courts.

Identification of the analytical parameters that can contribute to errors in data interpretation is relevant for quality review of data and when attempting to cross-compare data that was analysed using the different software packages.

II. MATERIALS AND METHODS

Raw data .fsa files were collected from ABI Genetic Analysers and adjudicated casework (Identacode Consulting LLC). Data were re-analysed for the ability to effectively partition DNA mixtures into major or minor contributors using GeneMarker HID software v.2.6.0. This software analysis package has a flagging feature that color codes data that is subjective and needs to be assessed based on analysis criteria that specifies the peak height difference allowed for establishing a difference between major and minor contributors to a DNA mixture on the electropherogram image. For establishing peak height ratios, the first peak per genetic locus that is encountered by the software is scored as a value of 1.0 and subsequent peaks are expressed as a percentage of the first peak per any given marker using the first encountered peak height as the denominator. Peak height ratios and clustering based on percentages was evaluated as a mechanism for establishing true number of contributors independent of allele number. A second data set was analysed for peak inclusion percentages for several different analytical thresholds of 10 rfu, 25 rfu, 50 rfu and 100 rfu and compared for quality. A review of the validation of the FST software for allele calls in validation samples and associated controls was made for drop-in rates and for the empirically derived drop-out rates to establish potential for scientific inaccuracies that might explain the error rates.

III. RESULTS

After reviewing the re-analysed DNA data files, several sources of error were identified that can contribute to the scientific inaccuracies associated with the established error rates with probabilistic genotyping software programs. These sources of error include the analytical value used for establishing percentage difference between major and minor contributor, analytical threshold settings used above the baseline noise, estimates of data point loss (drop-out rates) and contamination events (drop-in rates).

Major-minor contributor percentage criteria used in analysis to establish number of contributors. Sample data with calculated peak height ratios and an analytical threshold set at 50 relative fluorescence units (rfu) above baseline is shown in Figure 1. One of the difficulties in establishing true number of contributors to a DNA sample relates to what percentage difference is used to ascertain different sources. In an evidentiary DNA sample, one may never know with certainty the true number of contributors, therefore, some DNA laboratories would evaluate (in the example) the D8S1179 marker and observe that 3 different peak height ratios exist and are scored as classifications of 1.0, 0.41 and 0.21. Assessing the next marker, D21S11, two peaks are detected with 1.0 and 0.27 peak height ratios. For the D7S820 marker, 3 peaks are detected with ratios of 1.0, 0.42 and 0.94. The final marker, CSF1PO, has a peak height ratio of 1.0 and 0.96 for two peaks. Therefore, when analysing this data, the major contributor to the DNA mixture would be in the 0.90 – 1.0 peak height ratio range and those peaks would be associated together as representative of the same DNA profile. However, the minor contributor is significantly more debatable as either one or two minor sources of DNA, respectively. Two interpretations are possible: data clustering all together in the 0.21-0.42 classification or appearing as two separate minor sources with data clustering in the 0.21-0.27 class separate from the 0.41-0.42 class. This is graphically depicted in Figure 2 which represents a DNA profile from two individuals collected from a bloodstain from within a vehicle. This debate over which fragment can be attributed to which individual is especially relevant for low template DNA samples and touched objects as DNA is easily deposited on a surface but also can be removed and transported through secondary transfer to additional objects. On heavily and regularly handled items such as door knobs and car door handles, multiple individuals may be detected and the DNA peak height ratios of greater than three individuals can be extraordinarily difficult to assess even with probabilistic genotyping software when present at low levels as in this example.

Sample	Dye	Size	Height	Ht_Ratio	Marker	Allele	Difference	Quality	Score
14-4995_#5-2S1_D09.fsa	Blue	143.6	329	1	D8S1179	13	0.2	Pass	18.8
14-4995_#5-2S1_D09.fsa	Blue	148.2	135	0.41	D8S1179	14	0	Pass	4
14-4995_#5-2S1_D09.fsa	Blue	152.6	69	0.21	D8S1179	15	0	Undetermined	1.4
14-4995_#5-2S1_D09.fsa	Blue	204.4	271	1	D21S11	29	0	Pass	15.5
14-4995_#5-2S1_D09.fsa	Blue	210.9	73	0.27	D21S11	30.2	0.1	Pass	1.1
14-4995_#5-2S1_D09.fsa	Blue	265.2	125	1	D7S820	8	0.1	Pass	3.3
14-4995_#5-2S1_D09.fsa	Blue	276.5	53	0.42	D7S820	11	0.2	Undetermined	0.6
14-4995_#5-2S1_D09.fsa	Blue	280.4	118	0.94	D7S820	12	0.1	Pass	2.6
14-4995_#5-2S1_D09.fsa	Blue	325	153	0.96	CSF1PO	11	0.1	Pass	3
14-4995_#5-2S1_D09.fsa	Blue	329.1	160	1	CSF1PO	12	0.1	Pass	3.2

Fig.1. Peak height ratios for establishing true number of contributors to a DNA mixture

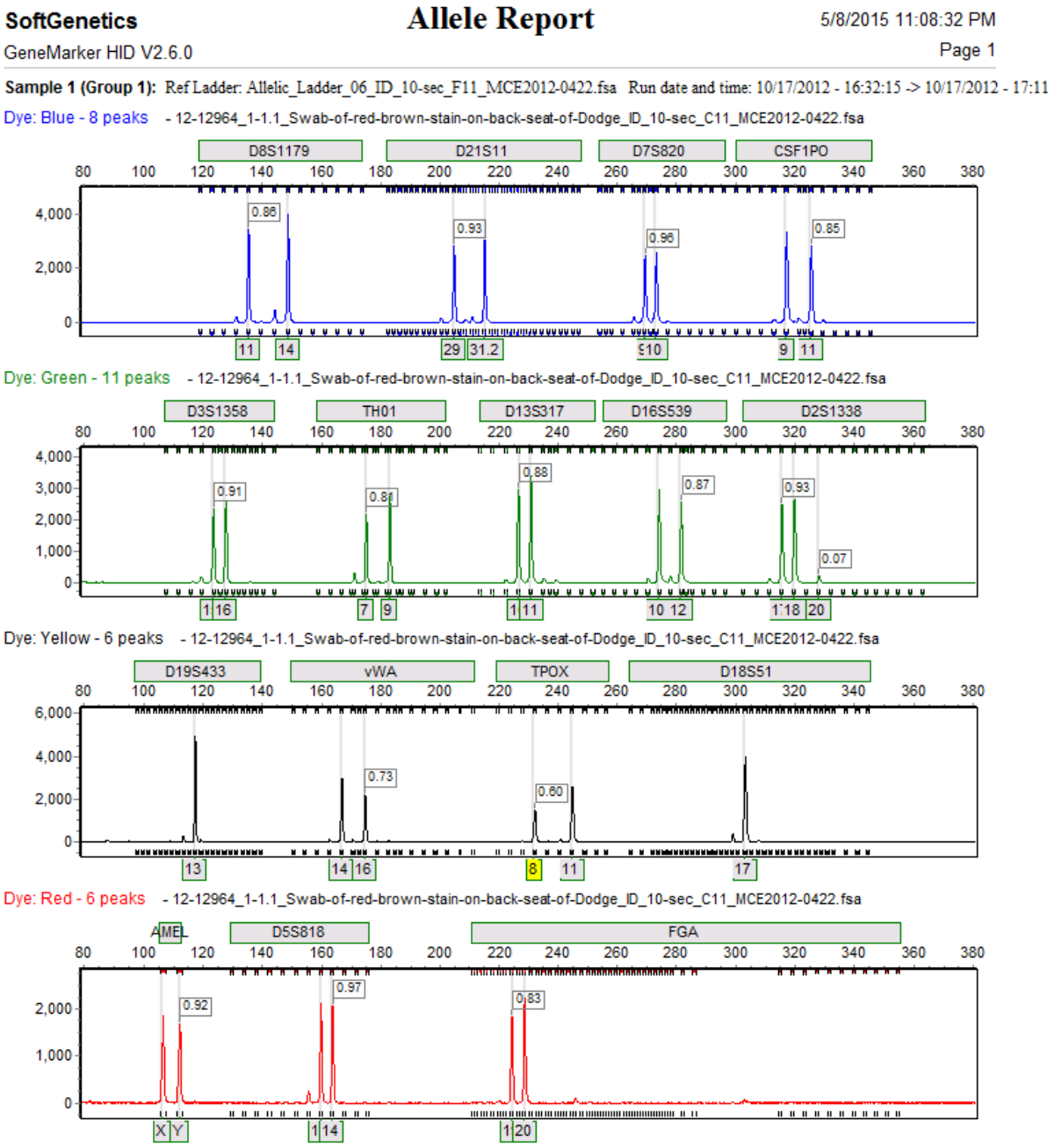

Fig. 2. A DNA electropherogram depicting a two person mixture detectable at D2S1338 with the 20 allele being significantly different in peak height ratio (0.07) as compared to the 17, 18 alleles (1.0, 0.93, respectively). At the TPOX marker, the debate is whether or not the 8 allele is from the major or minor contributor or if it belongs to a third person. Software such as FST that does assume number of contributors has an inherent error rate if the original assumption is incorrect because the data is forced to fit the model.

Analytical threshold settings used in analysis. The same concept holds true for analytical thresholds that are artificially set at some user defined level above the baseline of the instrument. A case example where data below the analytical threshold of 50 rfu was not utilized is illustrated in Figure 3. This case was a touch DNA case where glassine bags containing drugs were swabbed for the purpose of recovering DNA from the handler and there was discrepancy between whether or not one or two individuals could be correctly included in the DNA mixture. The software programs in the literature that are purported to have the least error rate are those that bring the analytical threshold closest to the baseline so as to recover the maximum amount of data and also assume no set number of contributors for analysis (e.g. TrueAllele).

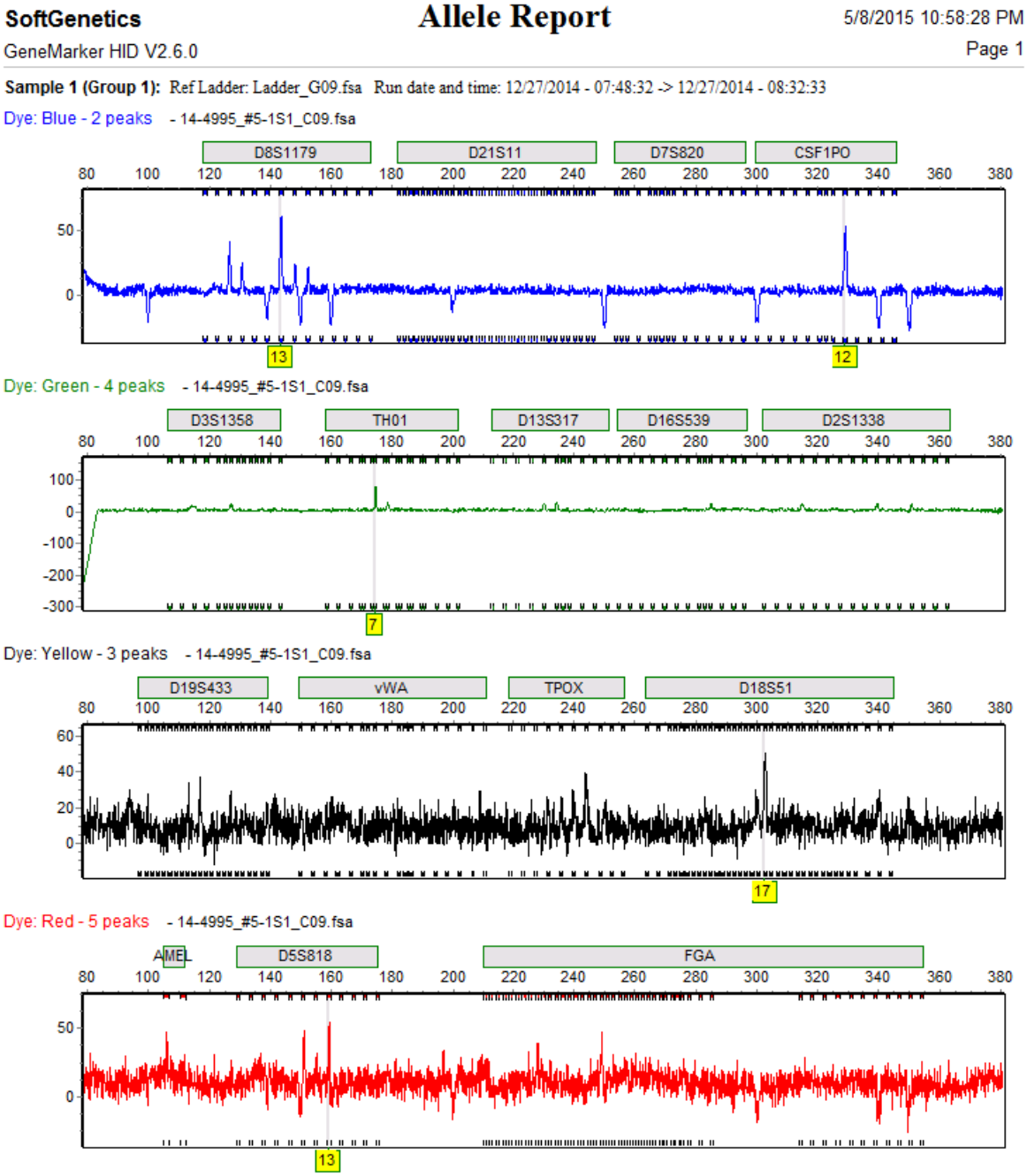

Fig.3.DNA mixture results recovered from a glassine bag and analysed for touch DNA results. All data called that are above the 50 rfu analytical threshold (AT) are flagged as yellow since they are close to the AT and require scrutiny on review. Some forensic examiners would argue there is additional data below the AT that should be noted or re-analysed to be more inclusive. A good example of the debate over AT settings is illustrated at the D5S818 marker where clearly two peaks are present but only one peak is above the AT of 50 rfu so only one peak is labelled. Four additional peaks below the 50 rfu AT are present at the D8S1179 marker and were not called by the software since they were below the user defined AT. Computational programs such as TrueAllele use all of the data close to the baseline and therefore have a reduced error rate.

Drop-in and drop-out rates. Estimates of contamination events for programming drop-in rates to software come from empirically derived validation studies. In one validation study, 9 of 305 samples were detected to have contaminant alleles in negative controls which gives an estimated contamination rate of 0.003% using 29 cycles of PCR amplification [5]. In another validation study, a range of 8-11% contamination was detected in negative extraction and amplification controls using 31 cycles of PCR [6]. The difficulty with inclusion of contaminant alleles in a computational evaluation of the data is that the true DNA profiles of the contributors are not accurately represented and the error rate for correctly establishing the true number of contributors to the sample can increase [4]. The contamination rates are anticipated to vary based on each forensic science laboratory and their corresponding methods, quality assurance, and quality control practices but is quantifiable. Most probabilistic software programs have the ability to subtract out the contaminant alleles

detected in the negative controls prior to calculating the likelihood ratios (LR). Empirically derived FST drop-out rates are purported to be more accurate by some examiners but in reality they reflect the drop-out rates for that particular data set based on DNA template quantity [7]. DNA quality estimates for drop-out rates are more theoretical in nature. The theoretical drop-out rate for degraded DNA samples is 50% as in a penny toss since DNA breakage is a random event although observationally larger DNA fragments may be more subject to shearing. It would be possible to predictively model increased chance of breakage events based on increased size of the DNA fragment which would better reflect the observation that larger DNA fragments appear to degrade or disappear from a DNA profile faster that smaller DNA fragments. The FST validation study models drop-out rates based on pristine DNA samples and does not account for random DNA degradation so only DNA quantity is considered in the calculation. Optimally, drop-out rates should be reflective of both loss of data through low template concentration as well as the random breakage of DNA observed in degradation patterns in a quantity by quality calculation.

IV. CONCLUSIONS

The most scientifically accurate method for computational analysis of DNA mixtures includes a highly defined approach to analysis with limited error. Currently, forensic science laboratories each set their own guidelines for DNA mixture interpretation and analyses leading to an inconsistent approach when comparing DNA data from state to state in the United States. To maintain consistency, a uniform national approach could be established and implemented in all forensic science DNA laboratories with some effort since most of the DNA typing reagents and equipment are calibrated and quality controlled already. The optimal analytical parameters would include a clearly defined difference between major and minor contributors before attempting to deduce sources, use of DNA data down to the baseline of the instrument, a standardized rationale based on both quantity and quality for drop-out rates (if utilized) in the calculations, and subtraction of any contaminant alleles from the data set that are detected in the negative controls. In addition, it would be beneficial to have false positive error rates computed across a consistent set of reference non-contributory DNA databases to correct for potential error in sampling and allow for a direct cross-comparison of error rates between probabilistic software programs.

ACKNOWLEDGMENT

Thank you to Attorneys Elaine Pourinski (Law Office of Elaine Pourinski, 13 Old South Street, Northampton, MA 01060-3840), Kyle Watters (Watters & Svetkey Law Offices, 286 Madison Avenue, New York, NY 10017) and The Legal Aid Society (Legal Aid Society Headquarters, 199 Water Street, New York, NY 10038) for sharing this data for review.

REFERENCES

[1] C. Ladd, H. C. Lee, N. Yang, F. Bieber, "Interpretation of Complex Forensic DNA Mixtures," *Croatian Medical Journal*, vol. 42(3), pp. 244-246, 2001.
[2] D. R. Paoletti, T. E. Doom, C. M. Krane, M. L. Raymer, D. E. Krane, "Empirical Analysis of the STR Profiles Resulting From Conceptual Mixtures, " *J. Forensic Sci.*, vol. 50, pp. 1361-1366, 2005.
[3] M. W. Perlin, K. Dormer, J. Hornyak, L. Schiermeier-Wood, S. Greenspoon, "TrueAllele Casework on Virginia DNA Mixture Evidence: Computer and Manual Interpretation in 72 Reported Criminal Cases," *PLoS ONE*, vol. 9(3), pp. 1-16, 2014.
[4] A. A. Mitchell, J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz, T. Caragine, "Validation of a DNA Mixture Statistics Tool Incorporating Allelic Drop-Out and Drop-In, " *Forensic Science International: Genetics*, vol. 6(6), pp. 749-761, 2012.
[5] The Institute of Environmental Science and Research Limited, "Estimation of STRMix Parameters for Erie County Central Police Services, "Validation Report, October 15, 2014.
[6] T. Caragine, R. Mikulasovich, J. Tamariz, E. Bajda, J. Sebestyen, H. Baum, M. Prinz, "Validation of Testing and Interpretation Protocols for Low Template DNA Samples Using AmpF*l*STR Identifiler, "*Croatian Medical Journal*, vol. 50, pp. 250-267, 2009.
[7] The Office of the Chief Medical Examiner of New York City (Department of Forensic Biology), "Forensic Statistical Tool Validation Summary Report, " vol. 19A, pp. 1-4, 2011.

