

The Use of the i-STReam Mixture Deconvolution Software Tool in the FSS-i³™ Suite for the Analysis of Casework Samples

Valerie K. Bostwick^{1*}, MSFS; Mallory K. Mest¹, BS; Cheryl A. Zakowski¹, BS; Terry W. Fenger¹, PhD; Rhonda K. Roby², PhD, MPH
¹Marshall University Forensic Science Center, Huntington, WV 25701; ²NIJ Technical Director, NEST Project, Fort Worth, TX 76107

ABSTRACT

During Phase I of the NEST Project, FSS-i³™ Software (Promega Corporation, Madison, Wisconsin) and other software programs were evaluated for their many features as expert systems for the analysis of single-source samples. In Phase II, different mixture deconvolution software tools are being evaluated. Mixture results pose an additional challenge in case interpretation and can be quite time-consuming, even for the experienced forensic scientist. The use of a mixture deconvolution tool can aid the forensic scientist in consistent unbiased evaluation of the data. These software programs and their tools are not expert systems, rather they are fancy genetic calculators designed to assist forensic scientists in mixture interpretation of casework STR data that include calculating mixing proportions and peak height ratios.

The focus of this presentation is the use of the i-STReam module in FSS-i³™ software to evaluate two-person mixtures. i-STReam has the ability to produce best-fit major and minor profiles. The i-STReam module calculates the mixing ratio of two donors in a mixture; it does not perform calculations for three or more contributors. The i-STReam module produces an overall mixing proportion for the samples and a list of possible candidate genotypes per locus. The overall mixing proportion is presented as a ratio, along with the mean, minimum, and maximum proportions for the combined loci in a sample lane. The mixture estimate provides guidance at each locus for the possible genotypes of each contributor. Based on the defined Preferential Amplification and the Mixing Proportion Rules, i-STReam reports the most likely genotype combinations for each locus.

Controlled mixture studies were conducted to produce simulated casework data. The design of the mixed samples included varying ratios of female and male DNA and varying input levels of DNA. These varying ratios and input quantities of DNA were amplified with PowerPlex[®] 16 System (Promega Corporation) and AmpFLSTR Identifier[®], Profiler Plus[®], and COfiler[®] PCR Amplification Kits (Applied Biosystems, Foster City, California). All samples were electrophoresed on an ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems).

The results of these studies demonstrate that the i-STReam module in the FSS-i³™ software can assist an analyst with challenging casework data. The peak height ratio calculations can be time-consuming when performed by hand and introduce the risk for human error or inconsistent calculations. An advantage to this new technology is that the peak height ratios for all combinations at each locus are calculated automatically and consistently. In addition, mixing proportions for the two contributors are calculated per locus as well as over the entire sample. As a result, the forensic analyst is freed to spend less time in performing mundane calculations and more time considering the various combinations produced by the given data.

INTRODUCTION

In a DNA laboratory, forensic analysts are tasked with the job of interpreting mixture results. With advancements in software interpretation tools for forensic DNA casework, rule firings and algorithms can be used to ensure consistency in interpretation from analyst to analyst. When working with samples at low levels with two contributors, mixture analysis becomes much more difficult and more variability may be seen in analyst interpretations. By incorporating the FSS-i³™ (Promega Corporation, Madison, Wisconsin) software into their workflow, analysts can evaluate mixture data more consistently within a laboratory.

Increasingly, expert systems and software tools have proved to be critical components during the analysis of both single source and mixture data. In mixture analysis, the i-STReam module of FSS-i³™ provides analysts an opportunity to examine all possible genotype combinations for major/minor contributors in two-person mixtures. The data are first analyzed in a precursor program, e.g. GeneMapper[®] ID (Applied Biosystems, Foster City, California), to export allele base pair size, height, and area. These values are imported into the FSS-i³™ software where calculations are performed. The time savings benefits of the program include peak height ratio and mixture proportion calculations for each marker. Most importantly, using the i-STReam software in no way replaces the analyst; it only assists in their interpretation of the data.

The NIJ Expert System Testbed (NEST) Implementation Team selected four unique profiles and created two sets of male/female mixtures. The information gathered from mixture analysis tools may prove to be a great advancement in forensic science technology.

MATERIALS & METHODS

A total of 280 samples were mixed resulting in two sets of two-person mixtures. The mixture sets were created from four specific profiles, two male and two female, to create two separate male/female mixtures. These profiles were chosen to demonstrate the effects of overlapping alleles, minor alleles which fall into stutter position, and other important mixture combinations. These mixture combinations include shared alleles, minor contributor alleles in the stutter position, major contributor alleles in the stutter position, no overlapping alleles, and other important mixture combinations.

These two-person sets were mixed at the following ratios: 30:1, 10:1, 3:1, 1:1, 1:3, 1:10, and 1:30. The samples were then amplified at varying DNA input levels: 1.50 ng, 1.00 ng, 0.50 ng, 0.25 ng, with five different multiplexes: AmpFLSTR Identifier[®], Profiler Plus[®], COfiler[®], SGM Plus[®] (Applied Biosystems, Foster City, California), and PowerPlex[®] 16 (Promega Corporation, Madison, Wisconsin). The samples were electrophoresed on an ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems) and analyzed with FSS-i³™ version 4.1.3 (Promega Corporation) in conjunction with GeneMapper[®] ID version 3.2 (Applied Biosystems).

ACKNOWLEDGEMENTS

The authors of this poster would like to acknowledge the staff of the National Institute of Justice's Expert System Testbed Project Implementation Project Team for their assistance with this project, specifically Chuck Heinrich, Program Manager to the NEST Project.

Marshall University Forensic Science Center is a partner with the Forensic Technologies Center of Excellence (FTCoE) as the host site of the NEST Project. This project is supported by Award No. 2008-MU-MU-K003 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice to the FTCoE administered by the National Forensic Science Technology Center.

The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

RESULTS

MIXTURE INTERP	
Pref Amp Threshold	50.0 %
Missing Tolerance	20.0 %
Homozygote Thresh	150 RFU
Major	Minor
D8S	8 13 14 F
D21	28 29 30 31,2
D7S	10 12 9 F
CSF	11 12 F F F
D3S	14 18 13 F
THO	F F F F F
D13	11 11 12 13
D16	11 12 9 F
D2S	20 24 21 23
D19	F F F F F
VWA	16 16 17 17
TPO	8 9 F F
D18	12 12 14 F
AME	X X X Y
D5S	11 12 13 F
FGA	23 23 22 F

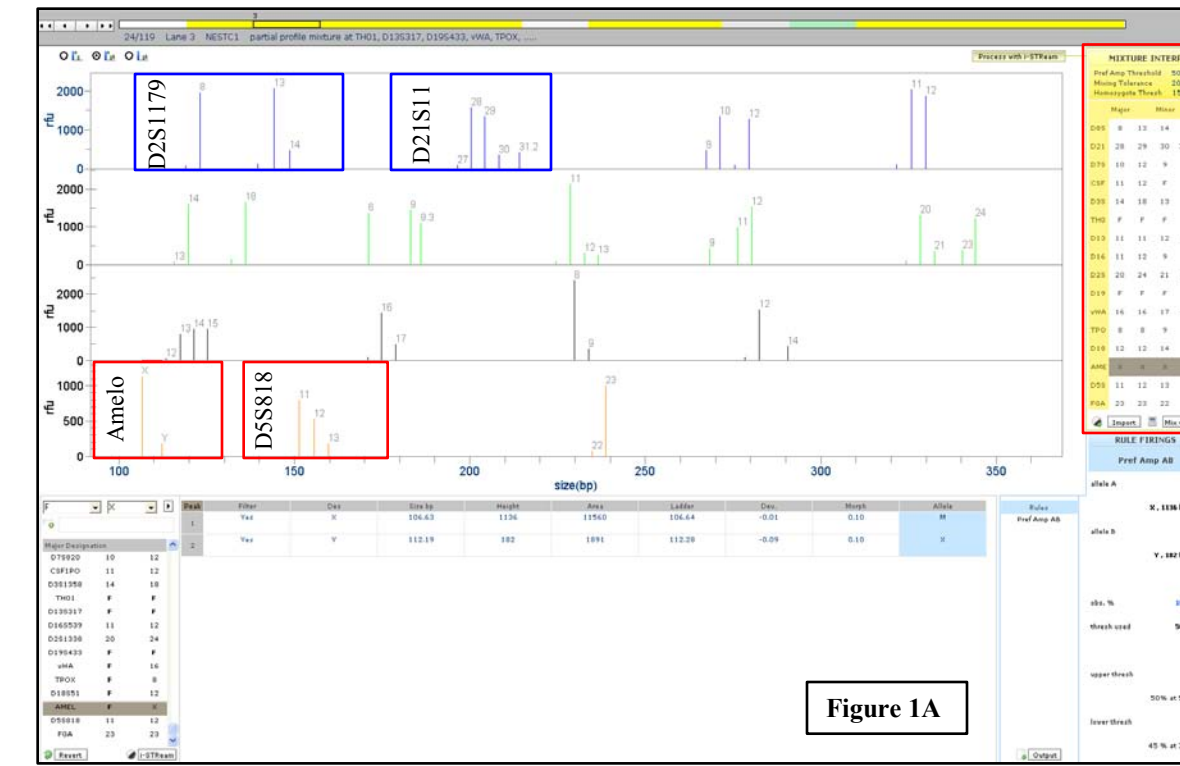


Figure 1: Two-Person Mixture of Samples A and X with Different Amounts of DNA. Two DNA samples were mixed at the same ratio (3F:1M) but with different input levels of DNA. Figure 1A shows an input level of 1.5 ng total DNA and Figure 1B has an input level of 0.25 ng total DNA. When comparing the Mixture Interop window of the 1.5 ng to the 0.25 ng samples, more F (or "failure") designations are present in the 0.25 ng results. Lower peak heights, dropout, and stochastic effects have interfered in the analysis and caused more ambiguity in the genotype determination for both major and minor profiles. Complete dropout of an allele occurred at Amelogenin (Y) and D5S818 (13). Stochastic effects at D5S818 in Figure 1B are represented in the Mixture Interop window where both major and minor genotypes have F designations.

MIXTURE INTERP	
Pref Amp Threshold	50.0 %
Missing Tolerance	20.0 %
Homozygote Thresh	150 RFU
Major	Minor
D8S	8 13 14 F
D21	F F F F F
D7S	10 12 9 F
CSF	11 12 F F F
D3S	14 18 17 F
THO	6 9 9,3 9,3
D13	11 11 12 13
D16	11 12 9 F
D2S	20 24 21 23
D19	F F F F F
VWA	16 F F F F
TPO	8 9 F F
D18	12 12 14 14
AME	F F F F
D5S	F F F F
FGA	23 F F F F

Pref Amp Tolerance	Mixing Proportion Tolerance	Homozygote
50%	20%	150
Weight: Maximum	Weight: Minimum	Weight: Mean
27% 3:1	19% 4:1	23% 3:1

Figure 2A: i-STReam has additional settings to be validated before using the mixture analysis tool. First, the **Pref Amp Tolerance** must be set. This is the minimum peak height ratio threshold for heterozygote peaks in both major and minor genotypes. A separate **Pref Amp Tolerance** value can be set for heterozygote peaks that fall below a specified RFU value. The **Mixing Proportion Tolerance** is used in conjunction with the Mix Prop Rule (see "Mixture Proportion Rule"). The **Homozygote** threshold is the RFU value in which there is enough confidence in a homozygote to give it a label of 11,11 vs. 11,F. The F indicates possible dropout.

The **Weight: Mean** is the MxP value, or the mixing proportion of the minor contributor across the entire sample (see "Weight: Mean"). The **Weight: Maximum** and **Weight: Minimum** refer to the minimum and maximum Mx_{obs} values used to calculate the mean for the MxP value.

Gen Number	Lane Number	Height	Area	Stutter	Peak Number	Stutter Number	Stutter Area
1	1	1000	1000		1		
2	2	1000	1000		2		
3	3	1000	1000		3		
4	4	1000	1000		4		
5	5	1000	1000		5		
6	6	1000	1000		6		
7	7	1000	1000		7		
8	8	1000	1000		8		
9	9	1000	1000		9		
10	10	1000	1000		10		
11	11	1000	1000		11		
12	12	1000	1000		12		
13	13	1000	1000		13		
14	14	1000	1000		14		
15	15	1000	1000		15		
16	16	1000	1000		16		
17	17	1000	1000		17		
18	18	1000	1000		18		
19	19	1000	1000		19		
20	20	1000	1000		20		
21	21	1000	1000		21		
22	22	1000	1000		22		
23	23	1000	1000		23		
24	24	1000	1000		24		
25	25	1000	1000		25		
26	26	1000	1000		26		
27	27	1000	1000		27		
28	28	1000	1000		28		
29	29	1000	1000		29		
30	30	1000	1000		30		
31	31	1000	1000		31		
32	32	1000	1000		32		
33	33	1000	1000		33		
34	34	1000	1000		34		
35	35	1000	1000		35		
36	36	1000	1000		36		
37	37	1000	1000		37		
38	38	1000	1000		38		
39	39	1000	1000		39		
40	40	1000	1000		40		
41	41	1000	1000		41		
42	42	1000	1000		42		
43	43	1000	1000		43		
44	44	1000	1000		44		
45	45	1000	1000		45		
46	46	1000	1000		46		
47	47	1000	1000		47		
48	48	1000	1000		48		
49	49	1000	1000		49		
50	50	1000	1000		50		

Locus	Allele	Height/Area	Contributor 1	Contributor 2	Pref Amp Rule	Mix Prop Rule	RC	Contributor 1	Contributor 2	
D8S1179	8	1001	13	14	8	442% N	-	-	-	
	13	2071	13	14	8	85% Y	85%	Y	83% 1.5	
	14	469	8	14	8	13	77% Y	77%	Y	86% 1.0
	-	-	14	14	8	13	-	Y	95%	Y
D18S11	8	13	13	14	77%	Y	77%	Y	14% 6.1	
	13	14	8	14	12%	N	12%	N	46% 1.1	
	13	13	8	14	12%	N	416%	N	54% 1.1	
	8	13	13	14	95%	Y	95%	Y	17% 5.1	
D5S818	8	14	13	14	12%	N	12%	N	52% 1.1	
	8	8	13	14	-	Y	442%	N	55% 1.1	
	8	14	13	13	418%	N	-	Y	46% 1.1	
	8	13	14	14	95%	Y	-	Y	10% 9.1	

Figure 2B: The i-STReam Summary Sheet breaks down the mixture analysis results by locus. It lists out all of the mixture specific settings used in analyzing the sample. Below the settings, results of all calculations are listed out and broken down by locus. For example, in Figure 2B results for the D8S1179 locus are shown. Up to four alleles per locus can be brought into i-STReam for analysis, along with their respective heights or areas. Each genotype combination is listed, considering every possibility for major and minor contributors.

Two rules are applied to the data, the Pref Amp Rule and the Mix Prop Rule (see below). If a combination passes a rule, it will get a Y (indicating, "yes, this rule passes"), and if it does not, it will get an N. Three Y results will "Include" the genotype combination, while even just one N result will exclude the combination. All included genotype combinations will be considered by the FSS-i³™ software, and consolidated in the final line, Database Consolidation. When there is ambiguity between multiple included genotypes, an F is used as a placeholder.

PREFERENTIAL AMPLIFICATION RULE

FSS-i³™ performs many calculations to sort through the many genotype combinations at each locus. Initially, a Preferential Amplification (Pref Amp) Rule is applied to each combination. A peak height ratio between heterozygote peaks is calculated for both the major and minor genotype combinations. Both of these peak height ratios must be above the user defined setting (50%) or below one divided by the setting (200%) (this will account for the larger value being on top).

The genotype combination will pass the Pref Amp Rule if the peak height ratios fall within this range, and be given a Y for yes. If the peak height ratio falls outside of this range, then the genotype combination will be given an N for no (Fig. 2B). In the case of a homozygote, the rule does not apply and will be given an automatic Y.

This rule is used to filter out nonsensical genotype combinations that contain imbalanced heterozygotes.

WEIGHT: MEAN (MxP)

The **Weight: Mean** is the best estimate of the proportion of the minor contributor across the entire sample (Fig. 2A). This value is also known as the MxP.

Once the Pref Amp Rule has been applied to every locus, an observed mixture proportion value (Mx_{obs}) is calculated for each genotype combination. The Mx_{obs} is a measure of the proportion of the minor contributor in each genotype combination at a locus. The equations used for each genotype pattern are shown in Table 1.

At each locus, the Mx_{obs} values for genotype combinations that passed the Pref Amp rule (combinations given a Y) are averaged. The averaged Mx_{obs} values across all loci in a sample are collected and also averaged together. The final value is known as your MxP, or expected Mx.

Multiple averaging steps ensures that the MxP value is as accurate as possible.

MIXTURE PROPORTION RULE

The second rule, Mixture Proportion (Mix Prop) Rule, considers the genotype combination's observed mixture proportion value (Mx_{obs}), the MxP (Weight: Mean), and the user defined Mixing Proportion Tolerance value.

FSS-i³™ takes the Mx_{obs} for each genotype combination and compares it to the MxP value for the overall sample. Using the range set forth by the Mixing Proportion Tolerance value (20%), each Mx_{obs} value must fall within plus or minus 20% of the MxP to pass the Mix Prop Rule.

Each genotype combination that falls in this range will get a Y for yes, and those that do not will get a N for no (Fig. 2B).

The purpose of this rule is to focus the analyst's attention on genotype combinations that have a minor contributor proportion that is consistent with the rest of the sample. Not taking into account the effects of degradation, the proportion of the minor contributor should be similar at each locus.

SOFTWARE CALCULATIONS

Locus	Genotype Pattern (Major/Minor)	Mx _{obs} Calculations
4 Alleles	AB, CD	$\frac{(C+D)}{(A+B+C+D)}$
3 Alleles No shared alleles	AA, BC	$\frac{(B+C)}{(A+B+C)}$
3 Alleles No shared alleles	AB, CC	$\frac{C}{(A+B+C)}$
3 Alleles One shared Allele	AB, AC	$\frac{[(3C)-B+A]}{2(A+B+C)}$
2 Alleles No shared alleles	AA, BB	$\frac{B}{(A+B)}$
2 Alleles One shared allele	AA, AB	$\frac{2B}{(A+B)}$
2 Alleles One shared allele	AB, BB	$\frac{(B-A)}{(A+B)}$
2 Alleles Two shared alleles	AB, AB	N / A

Mx_{obs} (mixture proportion) values are calculated below to demonstrate how the software uses these equations to determine the observed mixture proportions. Both calculations refer to the i-STReam summary sheet in Figure 2, and the loci are visible in the electropherogram in Figure 1A (blue boxes).

Mx_{obs} Calculation for D21S11, Figure 2:
 Genotype Pattern: AB, CD (4 alleles)
 Major: 28,29 Minor: 30,31,2
 Peaks: 28 (1592 RFU) 29 (1349 RFU) 30 (348 RFU) 31,2 (418 RFU)
 Alleles: A=28 B=29 C=30 D=31,2
 Equation: $\frac{(C+D)}{(A+B+C+D)} = Mx_{obs}$
 $\frac{(348+418)}{(1592+1349+348+418)} = Mx_{obs}$
 21% = Mx_{obs}

Mx_{obs} Calculation for D8S1179, Figure 2B:
 Genotype Pattern: AB, AC (3 alleles, one shared)
 Major: 13,8 Minor: 13,14
 Peaks: 8 (1961 RFU) 13 (2071 RFU) 14 (469 RFU)
 Alleles: A=13 B=8 C=14
 Equation: $\frac{[(3C)-B+A]}{2(A+B+C)} = Mx_{obs}$
 $\frac{[(3*469)-1961+2071]}{2(2071+1961+469)} = Mx_{obs}$
 17% = Mx_{obs}

Table 1: Chart depicting scenarios, genotype patterns, and equations used to determine the correct Mx_{obs} calculation performed by FSS-i³™. This table lists the equations used for calculating the observed mixture proportion for the minor contributor (Mx_{obs}) in each genotype combination.

CONCLUSIONS

- Peak height ratios and mixture proportions calculated consistently and accurately for every genotype combination.
 - Reduces possibility of human error.
- Every possible genotype combination is considered.
- Possibility of allelic dropout is considered.
- Two-person mixtures can be deconvoluted with the i-STReam software.
- No reference profiles can be applied in the software.
- Software cannot analyze tri-allele loci.
- Algorithms do not account for samples that are degraded or inhibited.
- An analyst *must* use their expertise to fully analyze each and every sample, and take ownership of all allele calls.

DISCUSSION

Unlike expert systems for single source analysis, mixture deconvolution tools such as FSS-i³™'s i-STReam are not meant to take the place of a trained forensic analyst. i-STReam is designed to be a tool used by analysts to assist them in their expert analyses. Calculations such as peak height ratios and mixture proportions are performed accurately and reproducibly every time the software is run. Based on user-defined parameters, genotype combinations are filtered for a more expedient review of mixture data. Once a genotype for a major and minor profile are determined, they can be manually compared to reference profiles or brought into a separate software program to perform mixture statistics.

REFERENCES

- Bill, M., Gill, P., et al. (2005). PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International*, 148, 181-189.
- Bill, M., Knox, C. (2005). FSS-i³ expert systems. *Profiles in DNA*, 8(2) 8-10.
- Christen, A., Roby, R. et al. (2006). The National Institute of Justice's Expert Systems Testbed Project: Phase II, Analysis of Casework Samples. Proceedings of the Seventeenth Symposium on Human Identification.