The DNA Mixture Conundrum: Sample Variation and its Effects on Mixture Deconvolution Tools

Angela M. Dolph*[1,2] and John M. Butler[2]

[1] Marshall University, Forensic Science Department, 1401 Forensic Science Drive, Huntington, WV 25701
[2] National Institute of Standards and Technology, Biochemical Science Division, 100 Bureau Drive, Gaithersburg, MD 20899

Email: dolph@marshall.edu

Disclaimer: This project was funded in part by the National Institute of Justice through interagency agreement 2003-IJ-R-029 to the NIST Office of Law Enforcement Standards. Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Introduction

For many in the forensic community, DNA mixture interpretation is a dreaded and confounding task. Not only does mixture interpretation involve a manual calculation method to find the correct genotypes, but it can be challenging to detect and interpret mixtures without extensive experience and training. From the NIST MIX05 study, it was shown that the participating labs had different methods of reporting mixture ratios, statistics, solving possible mixture combinations, and reporting. Inconsistencies have emerged because no national guidelines exist yet on how to perform mixture interpretation and statistical analysis. The inter-laboratory variation also illustrates that the forensic community would benefit from more uniform DNA mixture solving strategies, statistics, and reporting formats. The present study evaluates some DNA mixture deconvolution tools and assesses if these programs may be utilized to aid forensic DNA analysts in solving two-person mixtures. The mixture deconvolution tools analyzed are FSS-i3® v4.1.3 (i-STReam), Least-Square Deconvolution (LSD), and USACIL's DNA_DataAnalysis v2.1.3. An example of a mixture electropherogram is shown in Figure 1. This electropherogram is from GeneMapper® ID v3.2 and is the same data shown in the FSS-i3®, LSD, and DNA_DataAnalysis replicate and ratios examples discussed below.

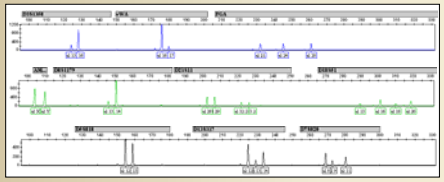

Figure 1. GeneMapper® ID electropherogram of Profiler Plus® 1:2 mixture ratio replicate 1a.

Materials and Methods

Experiment 1 – MIX05 *(all 2-person mixtures)*

- MIX05 data sent through FSS-i3® i-STReam and LSD
- MIX05 data from **http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm**
  - Case 1 – 3 parts female : 1 part male
  - Case 2 – 1 part female : 3 parts male
  - Case 3 – 1 part female : 1 part male
  - Case 4 – 7 parts female : 1 part male
- Web-LSD program accessed at **https://lsd.lit.net**
- STR amplification kits: SGM Plus®, Profiler Plus®, Identifiler®, COfiler®, PowerPlex® 16

Experiment 2 – Replicates and Ratios

- Mixtures were created by combining genomic DNA samples at different major and minor contributor ratios: 1:2, 1:3, 1:5, 1:8
- AmpFlSTR® STR kits: Identifiler®, COfiler®, Profiler Plus®
- The samples were amplified in replicate (n = 7) in order to test PCR variation and to observe how this variation affects the mixture deconvolution tool's ability to reliably solve DNA mixtures. An example of the peak height variation across replicates can be seen in Table 1.
- The data was collected on a 3130xl and initially analyzed with GeneMapper® ID v3.2 then evaluated with the tools FSS-i3® v4.1.3, Web-LSD, and DNA_DataAnalysis v2.1.3.
- FSS-i3® v4.1.3 was purchased from Promega Corporation.
- DNA_DataAnalysis v2.1.3 was obtained from Tom Overson at USACIL (US Army Crime Laboratory, Forest Park, GA).

Locus	Alleles	Replicate Peak Heights (RFUs)						
		1a	1b	2a	4a	6a	6b	7a
D3	15	244	680	564	585	652	228	457
	16	948	2526	2476	2307	2979	2096	1549
vWA	16	1221	3683	3117	2809	3874	2488	1992
	17	178	556	388	431	582	475	297
FGA	21	330	1050	764	747	1244	714	594
	24	309	1164	1100	949	1253	814	604
	28	327	1420	1131	1059	1216	944	633
Amel	X	756	1720	1422	1463	1706	1303	979
	Y	660	1575	1433	1272	1534	1459	896
D8	13	203	762	662	465	861	531	419
	14	1167	3361	2891	2416	3562	2563	2065
D21	28	407	1493	1402	1071	1691	1071	703
	29	407	1464	1163	1152	1472	1073	716
	32.2	158	417	589	373	741	373	289
	33.2	142	512	385	357	591	346	256
D18	13	120	673	488	464	454	394	227
	16	289	1380	1125	877	1240	842	573
	18	98	370	465	319	420	351	264
	20	229	1058	959	856	1298	798	567
D5	12	577	1785	1443	1382	1919	1206	941
	13	488	1519	1589	1383	1630	1364	866
D13	12	470	1539	1581	1304	1828	1164	954
	13	122	590	444	367	533	392	274
	14	295	1281	955	914	1124	896	682
D7	8	266	1365	1361	1053	1377	980	762
	9	196	425	352	260	446	224	204
	11	198	956	759	630	813	820	474

Table 1. Profiler Plus® 1:2 mixture ratio peak heights showing variation across replicates.

FSS-i3® i-STReam

Introduction

FSS-i3® is a suite of three software programs, i-STRess, i-STReam, and i-ntegrity, created by the Forensic Science Service. It is able to determine the genotypes of single source and mixture samples and detect for contamination. The main interface of the software that determines the genotypes is i-STRess, and another program, i-STReam, works on top of i-STRess to deconvolute mixture samples. i-STReam uses the heterozygote balance and mixture proportion guidelines to eliminate unreasonable genotype combinations. If multiple genotype combinations are calculated as possibilities at a single locus, the program gives F designations. The FSS-i3® spikeogram of the Profiler Plus® 1:2 mixture ratio replicate 1a is illustrated in Figure 2.

Figure 2. FSS-i3® spikeogram of Profiler Plus® 1:2 mixture ratio replicate 1a and the i-STRess program interface.

MIX05 Results

The FSS-i3® MIX05 results can be seen in Figure 3. Case 2 obtained the best results with 82% of the genotypes called with 100% accuracy. Not all of the genotypes were called because i-STReam allows for conservativeness in its F designation. Alternatively, Case 3 performed the worst with only 68% of the genotypes called and a 83% accuracy.

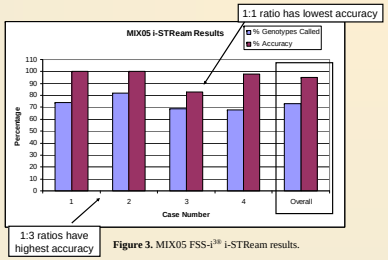

Figure 3. MIX05 FSS-i3® i-STReam results.

Replicates and Ratios Results

An output data file was created with GeneMapper® ID v3.2 then analysis was performed with FSS-i3® v.4.1.3. Some initial observations when trying to get the data into FSS-i3® included: the GMID minus A and stutter filters needed to be set to zero in order to allow all the alleles to enter FSS-i3®, i-STReam allows for conservativeness with its F designations, and stutter becomes a problem in the 1:8 mixture ratio. Table 2 shows the total number of alleles and the allele compositions of the loci used for the replicate and ratio study.

The results encompassing the entire replicates and ratios study can be viewed in Figure 4. Overall, i-STReam called 68% of the genotypes with only a 0.64% error; the remainder of the percentage was given as F designations.

The replicates and ratios i-STReam results according to mixture ratio are located in Figure 5. The 1:2 mixture ratio gave the worst results with 56% of the alleles being called correctly; however, the 1:3 ratio showed the best results with 78% of the alleles being called correctly. Drop-out was observed in the 1:5 and 1:8 mixture ratios.

i-STReam incorrectly called 26 / 4080 alleles. These incorrect calls are explained by PCR variation across the replicates. Fluctuations in peak height ratios allowed i-STReam to pass some incorrect genotypes and an example of the peak height ratio variation can be seen in Figure 6. Once the peak height ratio of the 11 allele achieved a certain threshold, i-STReam's calculations allowed for the incorrect genotype to be listed as probable.

Total # Alleles	% 4 Allele	% 3 Aallele	% 2 Allele	% 1 Allele
4080	25	33	38	4

Table 2. Number of alleles and loci composition for replicates and ratios study.

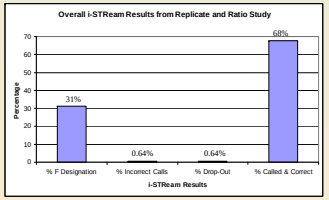

Figure 4. Overall replicate and ratio FSS-i3® i-STReam results.

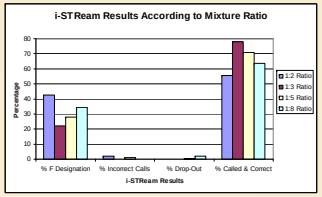

Figure 5. FSS-i3® i-STReam replicate and ratio results according to mixture ratio.

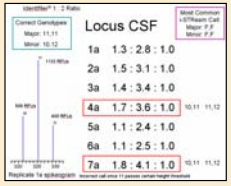

Figure 6. Peak height ratios of Identifiler® 1:2 ratio at locus CSF. The incorrect calls are boxed in red.

Web-LSD

Introduction

MIX05 analysis was conducted with Web-LSD. This version of LSD requires correct allele calls with the only input being the loci, alleles, and RFU values. LSD calculates best fit mass proportions and error residuals for all possible genotype combinations. The analyst then applies heuristic criteria, which include having a consistent mass proportion across all loci and small error residuals, to determine the correct genotypes. An example of the final LSD output is illustrated in Figure 7 utilizing the Profiler Plus® 1:2 mixture ratio replicate 1a.

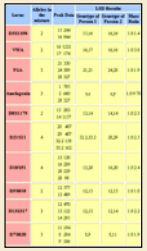

Figure 7. LSD output for Profiler Plus® 1:2 mixture ratio replicate 1a.

MIX05 Results

The MIX05 LSD results can be seen in Figure 8. Case 2 showed the best results with 96% accuracy and Case 3 had the worst results with 70% accuracy. All of the genotypes were called because a choice was always made according to the correct genotype based on the given calculations. The results were broken-down according to allele composition in order to illustrate LSD effectiveness in different allelic situations. The 4-allele loci illustrated the most success in achieving the correct genotypes.

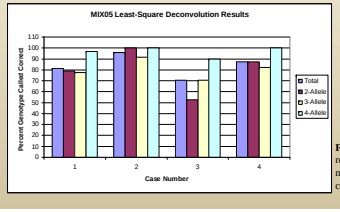

Figure 8. MIX05 LSD results according to case number and allele composition.

Replicates and Ratios Results

Three to four replicates per mixture ratio from the Identifiler® replicates and ratios data were analyzed utilizing Web-LSD. A correct genotype choice was always made and the results can be seen in Figure 9. The best results were obtained from the 1:8 mixture ratio showing a 99% accuracy. On the other hand, the 1:2 mixture ratio gave the worst results with 87% accuracy. Overall, these results, with a 95% average accuracy, were better than the results obtained from the MIX05 data and comparable to the i-STReam results.

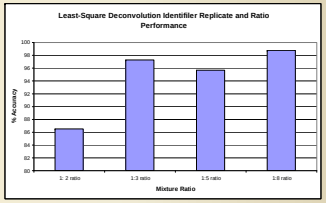

Figure 9. Web-LSD Identifiler® replicate and ratio study results.

DNA_DataAnalysis

Introduction

DNA_DataAnalysis was created by USACIL and is intended to be a DNA analyst tool that performs matching between samples and references, contamination checks, control checks, stutter evaluation, CODIS functions, two or three contributor mixture interpretation, etc. The program runs through Microsoft Excel 2003 and requires proper allele calls and removal of all artifacts with GMID prior to import. The mixture interpretation tool gives the most likely genotype combinations based on peak proportions and peak height ratios, and based on these calculations, the analyst deconvolutes the profiles according to lab protocol. An example of the 2-component mixture interpretation tool and its output calculations is illustrated in Figure 10.

The replicate and ratio data was sent through DNA_DataAnalysis, but no extensive evaluation was conducted because the program is more user driven and the mixture interpretation tool does not give explicit deconvoluted genotypes, it only provides the mixture calculations.

Figure 10. Locus D7 from Profiler Plus® 1:2 mixture ratio replicate 1a in the DNA_DataAnalysis mixture interpretation tool.

Conclusions

Overall, FSS-i3®, LSD, and DNA_DataAnalysis proved capable at solving two-person mixtures. LSD was not as accurate as i-STReam, but that could be because i-STReam allows for conservativeness in its F designations and in this study when utilizing LSD, a genotype was always chosen. In general, both Web-LSD and i-STReam were greater than 90% accurate with making mixture interpretation allele assignments.

It was also illustrated that amplification variation can lead to different and/or incorrect calls. Therefore, analysts should be aware of how confident they are in the results given by a mixture interpretation tool and optimization of the software parameters is very important.

From the experiments performed here, it appears that there is an optimum window of mixture ratios that can be confidently solved. This window of opportunity seems to be around 1:3 to less than 1:8, where stutter products begin to be in similar peak height ranges as minor contributor alleles.

References

Butler JM. Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, 2nd ed. Burlington, MA: Elsevier, 2005.

Bill M et al. PENDULUM – a guideline-based approach to the interpretation of STR mixtures. *Forensic Sci. Int.* 2005; 148:181-189.

Clayton et al. Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 1998; 91:55-70.

Wang T et al. Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. *J Forensic Sci.* 2006; 51(6):1284-1297.

Acknowledgments

The members of the Human Identity Project Team at NIST for their assistance and Tom Overson and Tim Kalafut for their insight into DNA_DataAnalysis. Also Dr. Pam Staton and Dr. Terry Fenger, my advisors at Marshall University.