NIST National Institute of Standards and Technology

Email: amy.decker@nist.gov
Phone: 301-975-5205

Determining Contributor Profiles from DNA Mixtures of Varying Ratios

Poster #90 at 19th International Symposium on Human Identification, Hollywood, CA, October 14-16, 2008

Acknowledgments: Ann Gross for helping collect the case summary data. Sample data supplied by Todd Bille (ATF) and Amy Christian and Rhonda Roby (NEST Project Team). Funding support from the National Institute of Justice through interagency agreement 2008-DN-R-121 to the NIST Office of Law Enforcement Standards.

John M. Butler[1], Amy E. Decker[1], Timothy S. Kalafut[2], Michelle M. Burns[1], Angela M. (Dolph) Spessard[1], and Margaret C. Kline[1]

[1]National Institute of Standards and Technology (NIST), 100 Bureau Drive MS 8311, Gaithersburg, MD 20899-8311
[2]United States Army Criminal Investigation Laboratory (USACIL), 4930 N 31st Street, Forest Park, GA 30297

Mixtures exist whenever more than one individual contributes biological material to a DNA sample. Mixtures, which are commonly seen in evidentiary samples (typically with two or three contributors), are more challenging to interpret compared to single source DNA samples that come from known references. Deciphering mixtures to determine the individual contributor profiles can be a time-consuming task for a forensic DNA analyst that requires training and experience. Over the past decade, our group at the National Institute of Standards and Technology (NIST) has conducted several interlaboratory studies involving mixture interpretation with short tandem repeat (STR) markers in order to better characterize issues that arise when trying to decipher the contributors to a DNA mixture result [1].

Computer software tools to aid STR mixture interpretation (e.g., [2,3]) have become available in recent years. We have evaluated several software programs for DNA mixture deconvolution including the i-STReam portion of FSS-i3 v4.1.3 (Promega Corporation), the Web-based Least Squares Deconvolution (Web-LSD) available at https://lsd.lit.net/, and an Excel-based program developed at the U.S. Army Criminal Investigation Laboratory (USACIL) entitled DNA_DataAnalysis v2.1.3 (see also ref [4]). Multiple mixture data sets (either from previous NIST interlaboratory studies, collaborators, or new in-house generated sample sets) are being evaluated with these software programs to explore the impact of contributor ratios and STR allele peak height ratios on the ability to determine the contributors' DNA profiles. These data sets include mixture ratios ranging from 1:1 to 1:40 as well as two and three person mixtures. Other mixture interpretation programs to be explored in the future include TrueAllele (Cybergenetics), Genoproof Mixture (Qualitype AG), and GeneMapperID-X Mixture Interpretation (Applied Biosystems).

NIST Interlaboratory Studies

See http://www.cstl.nist.gov/biotech/strbase/interlab.htm

Over the past 15 years, NIST has conducted a variety of interlaboratory studies with many of them focused on various aspects of mixture interpretation. These interlab studies are not graded proficiency tests. They provide a big-picture view of the community that offers labs an opportunity to directly compare themselves to others in an anonymous fashion. From these interlab studies, we have learned that instrument sensitivities can vary significantly, the amount of input DNA plays an important role in the ability to detect minor contributors, and protocols and approaches are often different between forensic DNA labs.

Mixed Stain Study #1 (MSS1) – Apr-Nov 1997
(6 single-source, 4 two-source, 1 three-source stains)
Mixed Stain Study #2 (MSS2) – Jan-May 1999
(4 single-source, 1 two-source, 1 three-source stains)
Mixed Stain Study #3 (MSS3) - Dec 2000-Oct 2001
(1 single-source, 5 two-source, 1 three source DNA extracts)
Mixture Interpretation Study (MIX05) – Jan-Aug 2005
(4 two-person mixture "case" data with victim profiles supplied)

Mixture Case Summaries

See http://www.cstl.nist.gov/biotech/strbase/mixture.htm

During 2007 and early 2008, Ann Gross (MN BCA) from the SWGDAM Mixture Interpretation Committee coordinated the collection of case summary data from 14 different forensic labs who collectively reported on 4780 samples. A preliminary summary of this information is shown below divided by crime classifications: sexual assault, major crime (homicide), and high volume (burglary). **Over half of the samples examined were single source and ~75% of all reported mixtures were 2-person.**

	minimum # of contributors					
Crime Class	**1**	**2**	**3**	**4**	**>4**	**N**
Sexual Assault	884	787	145	11	0	1827
Major Crime	1261	519	182	32	0	1994
High Volume	344	220	140	11	5	720
Total	2489	1526	467	54	5	4541
	54.8%	33.6%	10.3%	1.2%	0.1%	

Data sets examined (kit, #contributors, input DNA, contributor ratios):

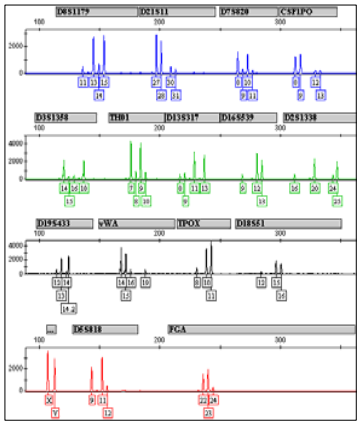
(Sources: ATF data supplied by Todd Bille, NEST data by Amy Christian and Rhonda Roby)

ATF (Identifiler, 2-person, 1ng DNA, 1:40 to 40:1, pristine:pristine and pristine:degraded samples)
NEST (Identifiler, PP16; 2-person, 0.25-1.5ng DNA, 1:30 to 30:1)
NIST MSS3 (Identifiler, PP16, 2-person, 1-4ng DNA, 3:1 to 10:1)
NIST MSS3 (Identifiler, PP16, 3-person, 3ng DNA, 4:2:1)
NIST MIX05 (ProPlus, 2-person, 1ng DNA, 1:1 to 1:8)
NIST additional 1 (Identifiler, 2-person, 1ng DNA, 1:3, 3:1, 1:5, 5:1)
NIST additional 2 (Identifiler, 3-person, 1ng DNA, 5:2:1)

Materials and methods for data generated at NIST:
Previously extracted NIST U.S. population samples
qPCR quantitation with Alu assay (SYBR Green) or Quantifiler
STR kits, commercial PCR kits amped at half-reactions, GeneAmp 9700, manufacturer cycling
ABI 3130*xl*, POP6, 36 cm array, 10s@3kV, 15 kV, 60 ºC
Data evaluation in GeneMapper*ID* v3.2

Note that the mixtures described here were created for research purposes and are synthetic mixtures of extracted DNA created in a controlled environment without PCR inhibitors or an unknown amount of degraded DNA as may be found in forensic casework situations.

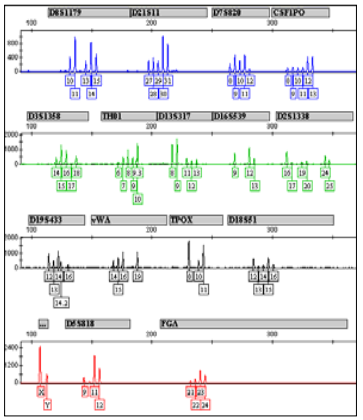
2-person mixture
(Identifiler data, 1ng DNA, 1:5)



1:5
Major- 7,9

5:1
Major- 8,10

3-person mixture
(Identifiler data, 1ng DNA, 5:2:1)



5:2:1
Major- 12,13
Contrib 2- 8,9
Minor- 10,11

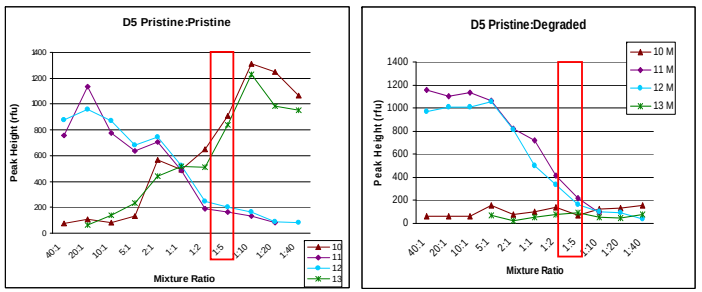
2:5:1
Major- 8,9
Contrib 2- 12,13
Minor- 10,11

2:1:5
Major- 10,11
Contrib 2- 12,13
Minor- 8,9

The 2- & 3-person mixtures displayed above were generated at NIST and are a mixture of a female and 1 or 2 male contributors. The samples contain pristine DNA and no degradation or dropout (peak below 50 RFUs) was detected. All stutter filters were turned off in GeneMapper ID v3.2. The peak threshold (limit of detection) was 50 RFUs. All artifacts were removed (i.e. bleed-through and –A). DNA concentrations: 2-person - 0.83 & 0.17ng/µL; 3-person – 0.625, 0.25, 0.125 ng/µL.

Peak Height Summaries of 2-person Pristine:Pristine and Pristine:Degraded mixtures

(ATF Identifiler data, 1ng total input DNA)

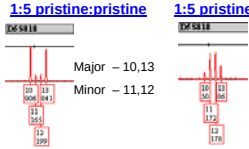


The electropherograms of the boxed data points are displayed in GeneMapper*ID* v3.2 (to the right).

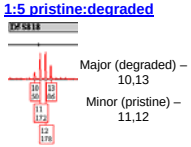
Data Summary: There are 4 unique alleles at D5S818 for both the pristine:pristine and pristine:degraded mixtures (ATF Identifiler data). The pristine:pristine mixture graph displays the peak heights increasing for the major contributor and decreasing for the minor contributor as expected. However, in the pristine:degraded example, even when more of the degraded major contributor is added, the peak heights do not increase as expected above the minor component. In this case, one might assume the 11,12 is the "major" and the 10,13 is the "minor". However, the apparent "minor" is actually the degraded "major" in this example.

The graphs (left) display all the 2-person mixture ratios (from 40:1 to 1:40) tested for these two sample types (pristine:pristine and pristine:degraded) from D5S818 of the Identifiler kit. Each line represents an allele and each of the points are the RFU values for that allele.

1:5 pristine:pristine



1:5 pristine:degraded



Poster available for download from STRBase: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2008poster.pdf

References

[1] Interlaboratory Studies on DNA: http://www.cstl.nist.gov/biotech/strbase/interlab.htm

[2] Bill, M., Gill, P., Curran, J., Clayton, T., Pinchin, R., Healy, M., and Buckleton, J. (2005) PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Sci. Int.* 148: 181-189.

[3] Wang T, Xue N, Birdwell J. Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. *J Forensic Sci.* 2006:51(6):1284-1297.

[4] Dolph, A.M. and Butler, J.M. (2008) Poster at AAFS Meeting (Washington, DC), "The DNA Mixture Conundrum: Sample Variation and Its Effects on Mixture Deconvolution Tools" Available at http://www.cstl.nist.gov/biotech/strbase/pub_pres/Dolph_AAFS2008poster.pdf

[5] Budowle, B., Shea, B., Niezgoda, S., and Chakraborty, R. (2001) CODIS STR loci data from 41 sample populations. *J.Forensic Sci.* 46: 453-489.

[6] Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., Morling, N., Prinz, M., Schneider, P.M., Weir, B.S. (2006) DNA commission of the International Society of Forensic Genetics. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci Int.* 160: 90-101.

Software Examined (see also [4])

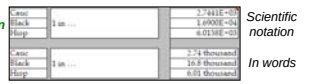
FSS-i3 (Promega) – makes decision without analyst input and is fairly conservative on assigning contributor allele combinations
Web-LSD (UTenn) and **DNA_DataAnalysis** (USACIL) – provides quantitative data behind multiple possibilities at examined loci, which analysts can use to make the final decision on contributor allele combinations

DNA_DataAnalysis

DNA_DataAnalysis is a Microsoft Excel based program developed at USACIL as a tool for DNA analysts to assist in analyzing STR mixture data (after appropriate allele calls are made, artifacts removed, and tabular allele calls, peak heights, and DNA sizes are imported from Genotyper or GeneMapperID). Features include control checks, contamination checks, stutter evaluation, CODIS functions, matching between samples and references and statistical calculations. The main focus of this study involved using the two or three contributor mixture interpretation tool to deconvolute mixture samples generated at NIST. 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 delineates the STR profiles according to lab protocol. Data sets examined include MSS3, NEST, and NIST 2- and 3-person mixtures.

The 3-person mixture (5:2:1) example from GMID (shown left) was evaluated with DNA_DataAnalysis. The probability of inclusion (PI) and exclusion (PE) statistics were calculated for each locus (e.g., CSF1PO) and the entire profile (see below). Allele frequencies are calculated using the same values as found in PopStats from the FBI population information for Caucasians, African Americans and Hispanics [5].

Probability of Inclusion from entire 3-person Identifiler Profile



Likelihood ratio (LR) calculations may also be performed with DNA_DataAnalysis so that laboratories may more easily perform LR analysis according to the International Society of Forensic Genetics (ISFG) recommendations on mixture interpretation [6].

USACIL is hoping to release this software to the forensic DNA community sometime in the near future.

Summary

Based on case summaries collected from 14 labs and almost 5000 samples, it appears that ~97% of mixtures observed contain 2 or 3 contributors. Various 2- or 3-person mixture data sets (generated in-house or obtained from collaborators) are being evaluated with different software programs to determine optimal ranges for reliable mixture interpretation.