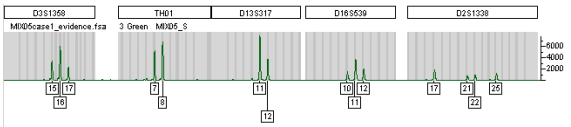


Mixture Interpretation:

An Open Discussion on Issues and Ways to Increase Casework Throughput

John M. Butler

NIST Biotechnology Division

HITA Workshop – September 30, 2005
Grapevine, TX

NIST and NIJ Disclaimers

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are mine 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.

Our publications and presentations are made available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

Presentation Format

Opening Remarks (~15-20 minutes):
- Mixtures: what are they and why are they challenging?
- Review of NIST mixture studies and lessons learned
- Observations from MIX05 Interlaboratory Study

Open Discussion (~15-20 minutes):
- **Questions** and **Responses** from audience

Summary (~5 minutes):
- Tools available and planned to aid mixture interpretation

Mixtures: Issues and Challenges

From J.M. Butler (2005) Forensic DNA Typing, 2nd Edition, p. 154

- Mixtures arise when two or more individuals contribute to the sample being tested.
- Mixtures can be challenging to detect and interpret without extensive experience and careful training.
- Differential extraction can help distinguish male and female components of many sexual assault mixtures.

Mixtures: Issues and Challenges

From J.M. Butler (2005) Forensic DNA Typing, 2nd Edition, p. 155

- The probability that a mixture will be detected improves with the use of more loci and genetic markers that have a high incidence of heterozygotes.
- The detectability of multiple DNA sources in a single sample relates to the ratio of DNA present from each source, the specific combinations of genotypes, and the total amount of DNA amplified.
- Some mixtures will not be as easily detectable as other mixtures.

When is a Sample a Potential Mixture?

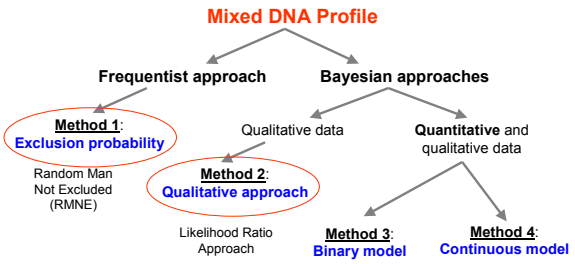
According to several MIX05 participant interpretation guidelines

- Number of Observed Peaks
  - Greater than two peaks at a locus
  - More than two alleles are present at two or more loci, although three banded patterns can occur
  - Presence of 3 alleles at a single locus within a profile
  - 4 peaked patterns (if observed at any locus), 3 peaked patterns (if observed at two or more loci), significant imbalances (peak height ratios <60%) of alleles for a heterozygous genotype at two ore more loci with the exception of low template amplifications, which should be interpreted with caution
- Imbalance of heterozygote alleles
  - thresholds range from 50-70%
- Stutter above expected levels
  - generally 15-20%

Two Parts to Mixture Interpretation

- **Deduction of alleles present in the evidence** (compared to victim and suspect profiles)
- **Providing some kind of statistical answer** regarding the weight of the evidence
  - An ISFG DNA Commission (Peter Gill, Bruce Weir, Charles Brenner, etc.) is evaluating the statistical approaches to mixture interpretation and will make recommendations soon

Approaches to Statistical Evaluation of Mixture Results



An ISFG DNA Commission chaired by Peter Gill will comment on these four methods for statistical mixture interpretation.

Figure 7.1 from Tim Clayton and John Buckleton, Chapter 7 "Mixtures" in *Forensic DNA Evidence Interpretation* (2005) CRC Press

Two Parts to Mixture Interpretation

- **Deduction of alleles present in the evidence** (compared to victim and suspect profiles)
- **Providing some kind of statistical answer** regarding the weight of the evidence
  - An ISFG DNA Commission (Peter Gill, Bruce Weir, Charles Brenner, etc.) is evaluating the statistical approaches to mixture interpretation and will make recommendations soon

NIST Initiated Interlaboratory Studies

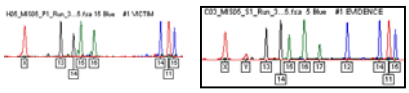
Studies involving STRs	# Labs	Publications
Evaluation of CSF1PO, TPOX, and TH01	34	Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M. (1997) Interlaboratory evaluation of STR triplex CTT. *J. Forensic Sci.* 42: 897-906
Mixed Stain Studies #1 and #2 (Apr–Nov 1997 and Jan–May 1999)	45	Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ. (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *J. Forensic Sci.* 46: 1199-1210
MSS3 **Mixed Stain Study** #3 (Oct 2000-May 2001)	74	Kline, M.C., Duewer, D.L., Redman, J.W., Butler, J.M. (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Anal. Chem.* 75: 2463-2469. Duewer, D.L., Kline, M.C., Redman, J.W., Butler, J.M. (2004) NIST Mixed Stain Study #3: signal intensity balance in commercial short tandem repeat multiplexes, *Anal. Chem.* 76: 6928-6934.
DNA Quantitation Study (Jan-Mar 2004) QS04	80	Kline, M.C., Duewer, D.L., Redman, J.W., Butler, J.M. (2005) Results from the NIST 2004 DNA Quantitation Study, *J. Forensic Sci.* 50(3):571-578
Mixture Interpretation Study (Jan - Aug 2005)	**69**	MIX05 ***Data analysis currently on-going ...***

Overall Lessons Learned from NIST MSS 1,2,&3

- Laboratories have instruments with different sensitivities
- Different levels of experience and training plays a part in effective mixture interpretation
- Amount of input DNA makes a difference in the ability to detect the minor component (labs that put in "too much" DNA actually detected minor components more frequently)

Mixture Interpretation Interlab Study (MIX05)

- **Only involves interpretation of data – to remove instrument detection variability and quantitation accuracy issues**
- **94 labs enrolled** for participation
- **69 labs have returned results** (17 from outside U.S.)
- Four mock cases supplied with "victim" and "evidence" electropherograms (GeneScan .fsa files – that can be converted for Mac or GeneMapper; gel files made available to FMBIO labs)
- Data available with Profiler Plus, COfiler, SGM Plus, PowerPlex 16, Identifiler, PowerPlex 16 BIO (FMBIO) kits



Perpetrator Profile(s) ??

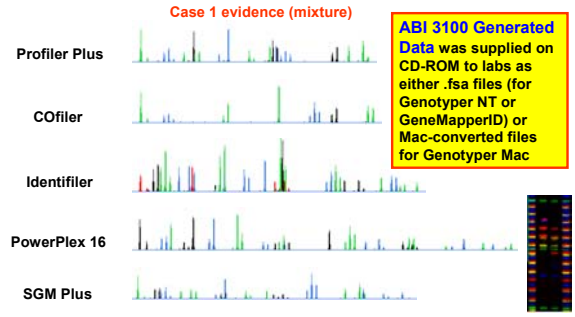
Along with reasons for making calls and any stats that would be reported

MIX05 Study Design and Purpose

- Permit a large number of forensic practioners to **evaluate the same mixture data**
- Provide multiple cases representing a range of mixture scenarios
- Generate data from multiple STR kits on the same mixture samples to compare performance for detecting minor components
- The primary variable should be the laboratory's interpretation guidelines rather than the DNA extraction, PCR amplification, and STR typing instrument sensitivity
- **Are there best practices in the field that can be advocated to others?**

MIX05 Results on Multiple Kits

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm



ABI 3100 Generated Data was supplied on CD-ROM to labs as either .fsa files (for Genotyper NT or GeneMapperID) or Mac-converted files for Genotyper Mac

FMBIO data was also made available upon request

Requests for Participants in MIX05

Mixtures representing four different case scenarios have been generated at NIST with multiple STR kits and provided to laboratories as electropherograms.

We would like to receive the following information:

1) **Report the results as though they were from a real case** including whether a statistical value would be attached to the results. *Please summarize the perpetrator(s) alleles in each "case" as they might be presented in court—along with an appropriate statistic (if warranted by your laboratory standard operating procedure) and the source of the allele frequencies used to make the calculation.* Please indicate which kit(s) were used to solve each case.

2) **Estimate the ratio for samples present in the evidence mixture** and how this estimate was determined.

3) **Provide a copy of your laboratory mixture interpretation guidelines** and a brief explanation as to why conclusions were reached in each scenario

Summary of MIX05 Responses

94 labs enrolled for participation

69 labs returned results (17 from outside U.S.)

50 labs made allele calls
39 labs estimated ratios
29 labs provided stats

STR kit results used

- **34 ProfilerPlus/COfiler**
- **10 PowerPlex 16**
- **7 PP16 BIO**
- **5 Identifiler**
- **2 SGM Plus**
- **1 All ABI kit data**
- **9 Various combinations**

Summary of Some MIX05 Reported Results

CASE #2		D3S1358	VWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO
True Perp	Z779619	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
LabID	Kit Used														
16	ProPlus/Cofiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
6	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
91	SGM Plus	15	15	20,24	X,Y	11,13	28,32.2	17,18				10,11	7,9.3		
46	PP16	--	--	--	--	--	--	--	--	--	--	--	--	--	--
37	ProPlus/Cofiler	--	15	20	X,Y	13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
2	PP16	15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	INC	8,10	10,11	7,9.3	9,10	7,10
13	PP16 & Identifiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
34	ProPlus/Cofiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
70	Identifiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
55	ProPlus/Cofiler	15	15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
21	ProPlus/Cofiler	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
73	ProPlus/Cofiler	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
29	Identifiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
54	All Kits	15,15	15,15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
90	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
9	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
4	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
33	ProPlus/Cofiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
12	ProPlus/Cofiler	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
67	PP16	15	15,16	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
86	ProPlus/Cofiler	15,15	15,15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
79	ProPlus/Cofiler	15,15	15,15	20,24	--	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
77	Identifiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--
60	PP16	15	15	20,24	X,Y	11,13	28,32.2	17,18	8,13	12,14	8,10	10,11	7,9.3	9,10	7,10
61	Identifiler	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Most calls were correct (when they were made)

Some Mixture Ratios Reported in MIX05

Many labs do not routinely report the estimated ratio of mixture components

LabID	Case1 (F:M)	Case2 (M:F)	Case3 (M:F)	Case4 (F:M)
13	2	5	<2	10
34	1.8--3.6	3.9--6.7	1.6--1.8	6.2--7.6
70				
55	68%:32%	85%:15%	64%:36%	
21				
73	2:1	6:1	2:1	not determined
29				
54	2:1	6:1	2:1	6:1
90	male23-39%	not determined	male64-71%	
9	3 or 4:1	4 or 5:1	1.4:1	~10:1
4	10:1	6:1	1:1	not determined
33	male60-78%	male80-90%	male58-71%	victim86%
12	male25%	male85%	male40-45%	unknown10%
67	1:2.3	6.4:1	2:1	1:6.8
86	2:1	6-6.5:1	1.6-2:1	4-4.5:1
79	~3:1 to ~2:1	~6:1 to ~4:1	~2:1*	a lot of victim
77				
60	2:1	5:1	2:1	10:1
61				

Some Reported Stats for MIX05 Case #1

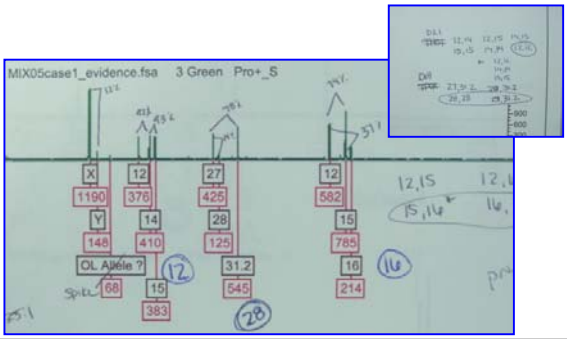
		Case1		
LabID	Kits Used	Caucasians	African Americans	Hispanics
77	Identifiler	PE calculated	PE calculated	PE calculated
73	ProPlus/Cofiler	none provided	none provided	none provided
4	ProPlus/Cofiler	none provided	none provided	none provided
12	ProPlus/Cofiler	none provided	none provided	none provided
29	Identifiler	none provided	none provided	none provided
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.00E+09	9.80E+10
46	PP16	5.60E+09	3.80E+11	none provided
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
61	Identifiler	1.50E+06	260,000	2.40E+07
79	ProPlus/Cofiler	930,000	47,900	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

Some Differences in Reporting Statistics

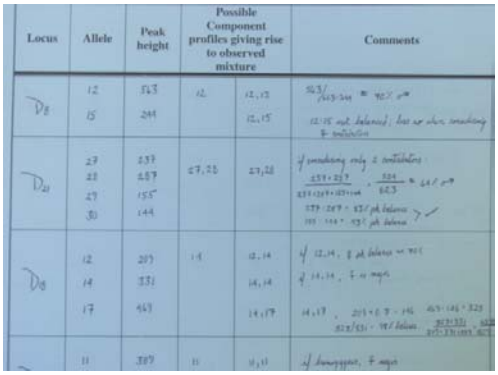
		Case1		
LabID	Kits Used	Caucasians	African Americans	Hispanics
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.00E+09	9.80E+10
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
79	ProPlus/Cofiler	930,000	47,900	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

Remember that these labs are interpreting the same MIX05 electropherograms

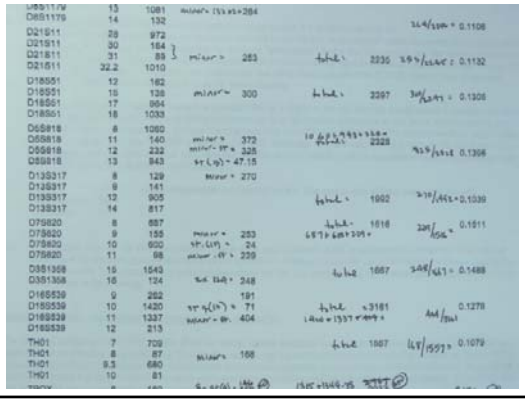
Manual Solving of MIX05 Peak Ratios and Possible Mixture Combinations



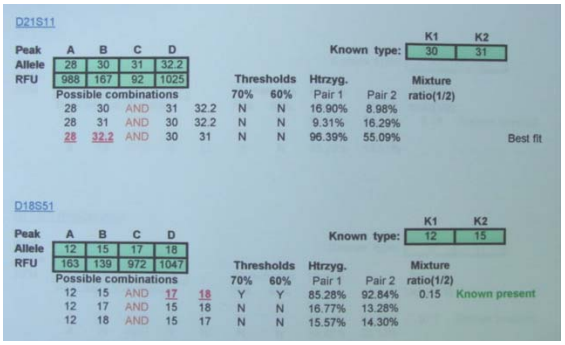
Manual Solving of Mixture Component Profiles



Another MIX05 Participant Manually Solving A Mixture



Semi-Automated Locus-by-Locus Interpretation Performed by One MIX05 Participant



Excel spreadsheet used to examine possible component combinations

Different Reporting Formats for MIX05 Data

Different Reporting Formats for MIX05 Data

Table 1 SUMMARY OF DNA TYPING RESULTS: Alleles Detected

Locus	Victim P Reference	Item S Questioned Sample
D3S1358	15,16	15,16,(17)
vWA	17	15,bt,17
FGA	19,21	19,20,21,22
Amelogenin	X	X(Y)
D8S1179	14,15	12,14,15
D21S11	27,31.2	27,(28),31.2
D18S51	12,15	12,15,(16)
D5S818	11	11
D13S317	11	11,12
D7S820	9,10	9,10
D16S359	11,12	10,11,12
TH01	8	7,8
TPOX	8	8
CSF1PO	11,12	11,12

( ): unbalanced/minor allele
bt: below laboratory threshold of 100
inc: inconclusive

No attempt to deduce perpetrator alleles (foreign profile)

Different Reporting Formats for MIX05 Data

Profile that would be put into CODIS

LOCI	CODIS ENTRY * obligate allele	OTHER ALLELE'S IN SUSPECT'S POSSIBLE PROFILE
D3S1358	17	16,17
VWA	15*	15,17
FGA	20,22	20,22
D8S1179	12	12,12
D21S11	28*	28,31.2
D18S51	15*	15,16
D5S818	--	
D13S317	12	12,12
D7S820	--	10
D16S539	10,11*	10,11
THO1	7*	7,8 maybe
TPOX	8	8 maybe
CSF1PO	--	11,12 maybe

Kits – Profiler Plus and Cofiler
Ratio - 1:2 (perpetrator:victim)

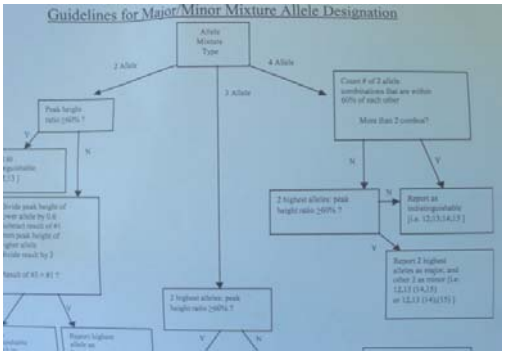
Different Reporting Formats for MIX05 Data

Case 1:

Item: description	D3S1358	VWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO
Pro+/CO_S: evid 1	15,16 (17)	15|17	19|20| 21|22	X,X (Y)	12|14|15	27,31.2 (28)	12,15 (16)	11,11	11|12	9|10	10|11 |12	7|8	8,8	11,12
Pro+/CO_P: victim 1 reference	15,16	17,17	19,21	X,X	14,15	27,31.2	12,15	11,11	11,11	9,10	11,12	8,8	8,8	11,12
Male interpreted from evidence 1	17	15,15 15,17	20,22	X,Y	12,12	28	16	11,11	12,12	Nd	10,11	7,7 7,8	Nd	Nd

Two allele values separated by a comma represent a genotype. Genotype calls assume biallelic donors with no null alleles.
( ) Indicates minor allele detected.
Single numbers and numbers separated by "|" represent an allele only designation rather than a genotype.
Interpreted profile assumes that the victim is present in the evidence mixture of two people. More than one genotype may be listed where a single genotype could not be conclusively determined. Nd=not determined due to level of results.

Some Protocols Have Flow Charts to Help Make Decisions in Mixture Resolution



OPEN DISCUSSION TIME

Any questions from the audience?

A High Degree of Variability Currently Exists with Mixture Interpretation

- "If you show 10 colleagues a mixture, you will probably end up with 10 different answers"
  - Peter Gill, Human Identification E-Symposium, April 14, 2005
- Is this a problem?

Question

- Without a suspect, should a mixture be considered (by show of hands, Y or N)?

Question

- Can mixture reports be standardized (to make them easier for review)?

Question

- Do you look at the evidence data first without considering the suspect's profile?

Question

- Do you have a decision point whereby you consider a mixture too complicated and do not try to solve it?
- If so, is the case declared inconclusive?

Question

- What elements would be needed in an expert system in order to automatically evaluate mixtures?

Question

- Are composite profiles acceptable – e.g., high injection for minor component and low injection for major component allele identification?

Question

- Should two amplifications be done – e.g., one at 1 ng to type the major component and one at higher concentration to move the minor component out of the low-copy number regime?

Question

- What kind of training materials would be beneficial to help your laboratory more effectively solve mixtures?

Question

- What are the biggest obstacles you face in your lab in terms of mixture interpretation?

Question

- How do you report mixture statistics in court?

Question

- What percentage of time is spent in a case trying to deduce the mixture components?

Question

- Do more loci in a multiplex result in less minor component alleles detected?...
- Is there an optimal multiplex for mixtures?

Question

- Will improved training information and software tools aid in mixture interpretation (or will lab policies prevent examination of these cases no matter what tools are brought to bear on this problem)?

WRAP-UP AND SUMMARY

Some Final Thoughts…

- It is of the highest importance in the art of detection to be able to recognize out of a number of facts, which are incidental and which vital. Otherwise your energy and attention must be dissipated instead of being concentrated (Sherlock Holmes, *The Reigate Puzzle*).
- "**Don't do mixture interpretation unless you have to**" (Peter Gill, Forensic Science Service, 1998).

Purpose of MIX05 Study

- **Goal is to understand the "lay of the land" regarding mixture analysis across the DNA typing community**
- One of the primary benefits we hope to gain from this study is **recommendations for a more uniform approach to mixture interpretation** and training tools to help educate the community

Value of the MIX05 Study

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm

- Data sets exist with multiple mixture scenarios and a variety of STR kits that **can be used for training purposes**
- A wide variety of approaches to mixture interpretation have been applied on the **same data sets evaluated as part of a single study**
- **Interpretation guidelines from many laboratories are being compared to one another for the first time in an effort to determine challenges facing future efforts to develop "expert systems" for automated mixture interpretation**
- **We are exploring the challenges of supplying a common data set to a number of forensic laboratories** (e.g., if a standard reference data set was ever desired for evaluating expert systems)

Additional Thoughts on Mixtures

From J.M. Butler (2005) Forensic DNA Typing, 2nd Edition, p. 166

- Some forensic DNA laboratories may decide not to go through the trouble of fully deciphering the genotype possibilities and assigning them to the major and minor contributors.
- An easier approach is to simply include or exclude a suspect's DNA profile from the crime scene mixture profile. If all of the alleles from a suspect's DNA profile are represented in the crime scene mixture, then the suspect cannot be excluded as contributing to the crime scene stain.
- Likewise, the alleles in a victim's DNA profile could be subtracted out of the mixture profile to simplify the alleles that need to be present in the perpetrator's DNA profile.

NIST Software Programs to Aid Mixture Work

Excel-based programs developed by David Duewer (NIST)

- **mixSTR** *(developed at request of Palm Beach Sheriff's Office)*
  - Does not interpret data (relies on user inputted alleles following STR data review)
  - Aids in the organization of STR mixture information
  - Considers only the presence/absence of alleles (no peak heights used)
- **Virtual MixtureMaker** *(developed to aid MIX05 sample selection)*
  - Creates mixture combinations through pairwise comparisons of input STR profiles
  - Returns information on the number of loci possessing 0,1,2,3,4,5, or 6 alleles in each 2-person mixture (also reports number of loci in each sample with 0,1,2, or 3 alleles)
  - Useful for selection of samples in mixture or validation studies with various degrees of overlapping alleles in combined STR profiles
  - Useful in checking for potentially related individuals in a population database

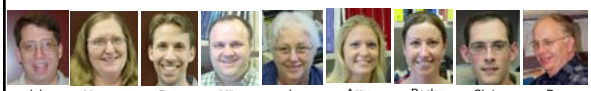
Programs can be downloaded from NIST STRBase web site:
http://www.cstl.nist.gov/div831/strbase/software.htm

Conclusions

- We plan to develop training information based on lessons learned from the MIX05 study.
- We intend to create other useful software tools like ***mixSTR*** and ***Virtual MixtureMaker*** to increase mixture interpretation capabilities of the forensic DNA typing community.

Acknowledgments

Funding from interagency agreement 2003-IJ-R-029 between NIJ and the NIST Office of Law Enforcement Standards



John Butler, Margaret Kline, Pete Vallone, Mike Coble, Jan Redman, Amy Decker, Becky Hill, Chris DeAngelis, Dave Duewer

Role in MIX05
- Margaret Kline (sample prep, running study)
- John Butler (study design and data review)
- Becky Hill (GeneMapper*ID* data review)
- Jan Redman (Access database entry, shipping)
- Dave Duewer (*Virtual MixtureMaker* to aid sample selection; **mixSTR program**)
- Chris Tomsey & Frank Krist (FMBIO Mac data)
- Kermit Channel & Mary Robnett (FMBIO NT data)

The many forensic scientists and their supervisors who took time out of their busy schedules to examine the MIX05 data provided as part of this interlaboratory study