### Mixture Interpretation

**Topics and Techniques for Forensic DNA Analysis**

**Houston DNA Training Workshop**

**Dr. John M. Butler**

National Institute of Standards and Technology

john.butler@nist.gov

**Scientific Content**

- **Common Casework Challenges**
  - Degraded DNA: Loss of signal at larger size loci. miniSTRs can help.
  - Mixtures: More than two alleles at multiple loci.

- **Forensic Casework DNA Profiles**
  - Profiler Plus STR kit results
  - Crime scene evidence (MIXTURE of victim and perpetrator)
  - All of the suspect’s alleles are included in the evidentiary DNA profile.

- **Differential Extraction**
  - Remove a portion of the mixed stain.
  - SDS, EDTA, and proteinase K (cell lysis buffer).
  - Incubate at 37°C.
  - DTT (male sperm)
  - Female sperm pellet.
  - REMOVES supernatant.

- **Evidence Enrichment Through Physical Capture of Sperm Cells**
  - Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides.
  - Laser capture microdissection permits collection of individual sperm. Typically >50 sperm heads enable collection of full diploid DNA profile.

- **Evidence (female fraction)**
  - Evidence from female sources.

- **Evidence (male fraction)**
  - Evidence from male sources.

- **Suspect**
  - Male

- **Victim**
  - Female

---

[Image links to the NIST website](http://www.cstl.nist.gov/biotech/strbase/training.htm)
Mixtures: Issues and Challenges

• 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.

How Can DNA Mixtures Arise?

• Two (or more) individuals contribute to the biological evidence examined in a forensic case (e.g., sexual assault with victim and perpetrator or victim, consensual sexual partner, and perp)

• Contamination of a single source sample from
  – evidence collection staff
  – laboratory staff handling the sample
  – Low-level DNA in reagents or PCR tubes or pipet tips

Y-chromosome markers can help here in some cases...

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

Contamination

• Systematic
  – e.g., Contaminated water or PCR buffer

• Sporadic
  – e.g., individual PCR tube contamination

• To reduce risks of contamination:
  – Careful lab cleanliness
  – Constant monitoring of reagents and consumables

• Contaminants are more likely to show up in the low molecular weight STR loci because they amplify more efficiently (miniSTRs will have a greater chance of detecting contaminating DNA)

• A negative control can detect systematic contamination but may not detect sporadic contamination, such as could be found in a single PCR tube

Potential Impact of Contamination on Cold Cases or Post-Conviction Testing

• While this contamination possibility might only rarely impact a careful forensic DNA laboratory, it can have potential significance on old cases under review including the Innocence Project. For example, if biological evidence from a 20-year old case was handled by ungloved police officers or evidence custodians (prior to knowledge regarding the sensitivity of modern DNA testing), then the true perpetrator’s DNA might be masked by contamination from the collecting officer. Thus, when a DNA test is performed, the police officer’s or evidence custodian’s DNA would be detected rather than the true perpetrator. In the absence of other evidence, the individual in prison might then be falsely declared “innocent” because his DNA profile was not found on the original crime scene evidence. This scenario emphasizes the importance of considering DNA evidence as an investigative tool within the context of a case rather than the sole absolute proof of guilt or innocence.

Impact of Contamination on Casework

• Use negative controls to predict the level of overall contamination in a lab

• Conclude that most likely outcome of a contamination event is a false exclusion...

Principles of Mixture Interpretation

Most mixtures encountered in casework are 2-component mixtures arising from a combination of victim and perpetrator DNA profiles.

Torres et al. (2003), Forensic Sci. Int. 134:180-186 examined 1,547 cases from 1997-2000 containing 2,424 typed samples of which 163 (6.7%) contained a mixed profile with only 8 (0.3%) coming from more than two contributors.

95.1% (155/163) were 2-component mixtures

Ratios of the various mixture components stay fairly constant between multiple loci enabling deduction of the profiles for the major and minor components

Some mixture interpretation strategies involve using victim (or other reference) alleles to help isolate obligate alleles coming from the unknown portion of the mixture
Example Mixture Data (MIX05 Study-Profiler Plus)

Single Source Sample (Victim)

Evidence Mixture (Victim + Perpetrator)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele 1</th>
<th>Allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>X</td>
<td>12.12</td>
<td>28.31.2</td>
</tr>
<tr>
<td>Y1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Y2</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

MIX05 Case #1; Profiler Plus green loci
Victim = major
Perpetrator = minor

MIX05: Issues and Challenges

- Artifacts of PCR amplification such as stutter products and heterozygote peak imbalance complicate mixture interpretation
- Thus, only a limited range of mixture component ratios can be solved routinely

Setting thresholds for the ABI 310/3100

- Where do current ideas on instrument thresholds for the ABI 310/3100 come from?
- How do I set these values in my laboratory?
- Why might they vary from one instrument to the next?
- How do these thresholds affect data interpretation?

Future defense attacks will likely focus on detection thresholds – can you defend your current threshold (e.g., 50 RFU or 150 RFU)?

The Scientific Reasoning behind the Concept of an Analytical Threshold (limit of detection)

- This is fundamentally an issue of reliability
- For a peak intensity three times the standard deviation of the noise there is a limited chance that such a signal is the result of a random fluctuation
- This is because 99.7 percent of all noise signals fall below this value (from the definition of a Gaussian curve)
- Below this point the very real possibility exists that what you think is a peak is simply a statistical fluctuation in the baseline noise.

Threshold Settings for the ABI 310/3100

- Detection Limit: 3x the standard deviation of the noise. Estimated using 2x peak to peak noise. (approximately 35 - 50 RFUs)
- Limit of Quantitation: 10x the standard deviation of the noise. Estimated using 7x peak to peak noise (150-200 RFUs). Below this point estimates of peak area or height are unreliable.
- Dynamic Range: The range of sample quantities that can be analyzed from the lowest to the highest (linear range is also important)
- Stochastic Threshold: Level of quantifiable DNA below which peaks can show severe imbalance (peak height ratios below 60%). Approximately 150-200 RFUs. Enhanced stutter also occurs at these signal levels.

What is a true peak (allele)?

- GeneScan function
  - Peak detection threshold
  - Peak height ratio (PHR)
  - Stutter percentage
- Genotyper function
  - Signal > 3x sd of noise
  - Allele 1
  - Allele 2
  - PHR consistent with single source
  - Typically above 60%
  - Stutter location above 15%
  - True alleles

Will be covered more in the low copy number section of this workshop...
How Does Your Laboratory Derive Its Interpretation Rules?
From your Validation Studies or Others?
• Peak detection threshold – set to 50 RFU or 150 RFU based on your lab data or what FBI or manufacturer has done? Do you use S/N >3 for determining if something is a true peak?
• Peak height ratio threshold – Set at 70% due to suggestion by manufacturer? Or 50-70% based on other data?
• Stutter product threshold – Genotyper macros set to 15%, manufacturer values, or adjusted based on your validation? Does it matter? How do these values play into your mixture interpretation guidelines?
• Sample Cleanup - Post PCR concentration a sample may also remove salts artificially enhancing injection. Will this move results into stochastic range?

Is a DNA Profile Consistent with Being a Mixture?

If the answer to any one of the following three questions is yes, then the DNA profile may very well have resulted from a mixed sample:

• Do any of the loci show more than two peaks in the expected allele size range?
• Is there a severe peak height imbalance between heterozygous alleles at a locus?
• Does the stutter product appear abnormally high (e.g., >15-20%)?

Steps in the Interpretation of Mixtures
(Clayton et al. 1998)

Step #1 Identify the Presence of a Mixture
Step #2 Designate Allele Peaks
Step #3 Identify the Number of Potential Contributors
Step #4 Estimate the Relative Ratio of the Individuals Contributing to the Mixture
Step #5 Consider All Possible Genotype Combinations
Step #6 Compare Reference Samples

Mixture: Issues and Challenges

• 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 others mixtures.

Mixtures: Issues and Challenges

Steps in the Interpretation of Mixtures
(Clayton et al. 1998)

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 has made recommendations

Our discussions have highlighted a significant need for continuing education and research into this area.

DNA commission of the International Society for Forensic Genetics:
Recommendations on the interpretation of mixtures
P. Gill, C.-B. Brown, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Marks, N. Martin, M. Paggi, P.M. Schneider, B.S. Weir
Promega Corporation, Madison, WI 53711. Available at: www.promega.com/geneticidproc/ussymp17proc/oralpresentations/Perlin.pdf

Some of Mark Perlin’s Recent Statements

Different laboratories follow different mixture interpretation guidelines. Moreover, different examiners within the same laboratory who are following the same guidelines often infer different STR profiles.

Therefore, there is no concordance in current forensic practice on what constitutes a “correct” mixture solution. Thus, it is not possible to conduct a mixture interpretation concordance study in order to validate a mixture interpretation method.

DNA mixture evidence currently fails the general acceptance test of both Frye and Daubert, since there are no generally accepted methods for interpreting mixed stains.

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

Interlaboratory studies help to better understand why variability may exist between laboratories

Most analysts are only concerned about their own lab protocols and do not get an opportunity to see the big picture from the entire community that can be provided by a well-run interlaboratory study

Overall Lessons Learned from NIST MSS 1,2,83

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)

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

http://www.cstl.nist.gov/biotech/strbase/training.htm
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, CODiFer, SGM Plus, PowerPlex 16, Identifiler, PowerPlex 16 BIO (FMBIO) kits
- Summary of results will involve training materials to illustrate various approaches to solving mixtures

Requests for Participants in MIX05

Mixture 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

A MIX05 Participant Noted…

“Things we do not do:"

- Calculate mixture ratios for casework
  - Calculation used for this study: Find loci with 4 alleles (2 sets of sister alleles). Make sure sister alleles fall within 70%, then take the ratio of one allele from one sister set to one allele of the second sister set, figure ratios for all combinations and average. Use peak heights to calculate ratios.
- Provide allele calls in reports
- Provide perpetrator(s) alleles or statistics in court without a reference sample to compare to the DNA profile obtained from the evidence. We will try to determine the perpetrator(s) profile for entry into CODIS."

We recognize that some of the information requested in this interlab study may not be part of a lab’s standard operating procedure

MIX05 Case Scenarios

Genomic DNA samples with specific allele combinations (“evidence”) were mixed in the following ratios:

Case #1 – victim is major contributor (3F:1M)
- Male lacked amelogenin X allele is missing in male perpetrator DNA sample for MIX05 Case #3

Case #2 – perpetrator is major contributor (1F:3M)

Case #3 – balanced mixture (1F:1M)
- Male lacked amelogenin X
- Male contained tri-allelic pattern at TPOX

Case #4 – more extreme mixture (7F:1M)
- Male contained tri-allelic pattern at TPOX

Female victim DNA profile was supplied for each case

Labs asked to deduce the perpetrator DNA profile – suspect(s) not provided

MIX05 Study Design and Purpose

Interlab studies provide a “big picture” view of the community

- Permit a large number of forensic practitioners 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

Case 1 evidence (mixture)

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

Profiler Plus
COffler
Identifiler
PowerPlex 16
SGM Plus

FMBIO data was also made available upon request

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/COffler
10 PowerPlex 16
7 PP16 BIO
5 Identifiler
2 SGM Plus
1 All ABI kit data
9 Various combinations

All participants were supplied with all data and could choose what kits to examine based on their experience and lab protocols

Generally Identifiler data was of poorer quality in the electropherograms we provided...which caused some labs to not return results (they indicated a desire for higher quality data through sample re-injection to reduce pull-up prior to data interpretation)

What MIX05 Participants Have Received Back from NIST...

• Certificate of participation in the interlab study
• Copy of the poster presented at the Promega Sept 2005 meeting displaying "correct" results for the perpetrator in each case scenario as well as an explanation of study design and preliminary results

http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05/MIX05poster.pdf

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 or 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%

These protocol differences can lead to variation in reported alleles and therefore the deduced profile and resulting statistics

Summary of Some MIX05 Reported Results
Case #2 has perpetrator as major component and thus is the easiest to solve...

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

Lab ID | Case (F M) | Case (M F) | Case (M F) | Case (F M) |
--- | --- | --- | --- | --- |
13 | 2 | 6 | <2 | 10 |
34 | 1 | 3 | 3 | 4 |
55 | 2 | 6 | 4 | 6 |
70 | 3 | 1 | 2 | 1 |
73 | 1 | 6 | 1 | 2 |
79 | 3 | 4 | 1 | 4 |
54 | 1 | 2 | 1 | 2 |
56 | 3 | 1 | 2 | 1 |
90 | 2 | 1 | 2 | 1 |
12 | 2 | 4 | 1 | 2 |
67 | 2 | 4 | 1 | 2 |
32 | 4 | 2 | 4 | 2 |
61 | 4 | 2 | 4 | 2 |
60 | 1 | 2 | 1 | 2 |
61 | 1 | 2 | 1 | 2 |

http://www.cstl.nist.gov/biotech/strbase/training.htm
Some Reported Stats for MIX05 Case #1

<table>
<thead>
<tr>
<th>LabID</th>
<th>Kits Used</th>
<th>Caucasians</th>
<th>African Americans</th>
<th>Hispanics</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>Preplus/Colter</td>
<td>PE calculated</td>
<td>PE calculated</td>
<td>PE calculated</td>
</tr>
<tr>
<td>73</td>
<td>Preplus/Colter</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>4</td>
<td>Preplus/Colter</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>12</td>
<td>Preplus/Colter</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>20</td>
<td>Preplus/Colter</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>80</td>
<td>Preplus/Colter</td>
<td>1.1E+15</td>
<td>2.1E+14</td>
<td>3.8E+16</td>
</tr>
<tr>
<td>34</td>
<td>Preplus/Colter</td>
<td>2.4E+11</td>
<td>7.0E+09</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>45</td>
<td>Preplus/Colter</td>
<td>0.0E+00</td>
<td>0.0E+00</td>
<td>0.0E+00</td>
</tr>
<tr>
<td>33</td>
<td>Preplus/Colter</td>
<td>1.2E+08</td>
<td>3.0E+07</td>
<td>1.7E+09</td>
</tr>
<tr>
<td>6</td>
<td>Preplus/Colter</td>
<td>0.0E+00</td>
<td>3.5E+02</td>
<td>2.0E+08</td>
</tr>
<tr>
<td>9</td>
<td>Preplus/Colter</td>
<td>1.8E+02</td>
<td>1.9E+02</td>
<td>1.5E+09</td>
</tr>
<tr>
<td>61</td>
<td>Preplus/Colter</td>
<td>1.2E+08</td>
<td>2.5E+07</td>
<td>2.4E+07</td>
</tr>
<tr>
<td>79</td>
<td>Preplus/Colter</td>
<td>0.0E+00</td>
<td>4.7E+00</td>
<td>1.3E+00</td>
</tr>
<tr>
<td>18</td>
<td>Preplus/Colter</td>
<td>4.3E+00</td>
<td>3.1E+00</td>
<td>3.9E+00</td>
</tr>
</tbody>
</table>

~10 orders of magnitude difference (10^5 to 10^15)
based on which alleles were deduced and reported

Remember that these labs are interpreting the same MIX05 electropherograms

Which loci are included in each calculation?

<table>
<thead>
<tr>
<th>Case</th>
<th>Some Differences in Reporting Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>34</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>33</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>6</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>9</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>79</td>
<td>Preplus/Colter</td>
</tr>
<tr>
<td>18</td>
<td>Preplus/Colter</td>
</tr>
</tbody>
</table>

Solved loci listed?

Case 1: ASCLD-LAB accredited?
Yes

Solved loci listed?

Case 1: Lab 9

• 4.14 x 10^7 used 1/CPI

Case 1: Lab 6

• 4.0 x 10^7 used selected loci and summed all possible genotypes for loci not completely deduced

Case 1: Lab 90

• 1.18 x 10^15 used theta value of 0.03 and deduced alleles at all 13 loci (correctly deduced all perpetrator alleles)

Different Detection Thresholds Used

Case 1: Lab 90 has specific, detailed mixture interpretation guidelines with worked examples and a fabulous flowchart

Case 1: Lab 16 has vague guidelines that begin with "mixture interpretation is not always straightforward. Analysts must depend on their knowledge and experience…"

http://www.cstl.nist.gov/biotech/strbase/training.htm
Questions for Consideration

• Do you look at the evidence data first without considering the suspect’s profile?

• Without a suspect, does your lab proceed with mixture interpretation?

• 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?

• What kind of training materials would benefit your lab in improving consistency in mixture interpretation?

Examples of MIX05 Report Formats

All examples with Case #1
(~3:1 mixture with female victim as the major component – and victim profile is provided)

Manual Solving of MIX05 Peak Ratios and Possible Mixture Combinations

Manually Solving Mixture Component Profiles

Lab 90 – correctly deduced all perpetrator alleles in Case #1
(highest of the 7 listed stats for ProPlus/COfiler at 1.18 x 10^{15})
Also prepared a CODIS Search/Upload Request with the deduced profile

A Model Report of Analysis…

• “The Profiler Plus and COfiler sample files were evaluated by four different analysts, using both NT and MAC analysis platforms. The analysts checked for concordance, and a single conclusion for each mock case has been issued.”

• They detailed all assumptions made outside the course of routine casework:
  – Assumed intimate samples
  – That a comparison of deduced “foreign” alleles had been made with the perpetrator’s known standard in order to calculate the significance of the inclusion with the evidentiary profile

• For Case #4: “A Combined Probability of Inclusion was calculated and reported for only those loci where all the alleles were above threshold [75 RFUs]. However, a minor profile(s) could not be deduced from this sample. Please note that our laboratory may employ strategies to gain more information from the sample, such as a 10 second injection of the CE and Y-STR analysis.”

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

No attempt to deduce perpetrator alleles (foreign profile)

The community would benefit from more uniform reporting formats and mixture solving strategies...
Some Protocols Have Flow Charts to Help Make Decisions in Mixture Resolution

Quotes from One Lab’s MIX05 Report

- Case 1: STR typing results from the Evidence sample indicate a DNA mixture profile. The victim cannot be excluded as a possible donor of the genetic material in the Evidence sample. No statistics will be generated at this time.
- The Evidence samples would have to be re-run in order to verify any alleles called in the final profiles. This is true for any mixed sample profiles as per our laboratory guidelines.
- Our laboratory does not “pull out” any profile from a mixture for interpretation or statistical purposes. The exception to this is for CODIS profiles where the alleles that can be unambiguously attributed to the victim are removed.
- We currently do not calculate and report statistics on mixture samples.

The Same Lab’s “Mixture Interpretation Grid”

The Mixture Interpretation Grid provides an objective summary of how many alleles the two profiles have in common. The results will fall into one of the following categories:

- “Can not be excluded”
  - If the majority of alleles from the exemplar specimen are not present and/or a number of alleles foreign to the exemplar specimen are present
- “Excluded”
  - If the majority of alleles from the exemplar specimen are not included in the mixture profile
- “No conclusion can be made”
  - Cases where the mixture profile is limited

See laboratory mixture interpretation guidelines for further explanation. All the cases in the study fell into the “can not be excluded” category.

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)

Conclusions (Opportunities for Improvement)

- It is worth taking a closer look at protocol differences between labs to see the impact on recovering information from mixture data
- Training should help bring greater consistency
- Expert systems (when they become available and are used) should help aid consistency in evaluating mixtures and help produce more uniform reporting formats

Software Programs (Expert Systems) for Mixture Deconvolution

These programs do not supply stats (only attempt to deduce mixture components)

- **Linear Mixture Analysis (LMA)**
  - Part of TrueAllele system developed by Mark Perlin (Cybergenetics)

- **Least Squares Deconvolution (LSD)**
  - Described by T. Wang (University of Tennessee) at Oct 2002 Promega meeting
  - Available for use at [https://lsd.lit.net/](https://lsd.lit.net/)

- **PENDULUM**
  - Part of FSS i-3 software suite (i-STReam)

**USACIL program developed by Tom Overson**

http://www.cstl.nist.gov/biotech/strbase/training.htm
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

**mixSTR Program**

Comparisons are made between
- suspect and evidence (S/E) alleles,
- suspect and suspect (S/S) alleles (to look for potential close relatives),
- evidence and other evidence (E/E) sample(s) alleles (to see how various evidentiary samples compare to one another), and
- controls to evidence (C/E) and controls to suspect (C/S) alleles (as a quality control contamination check).

**Virtual MixtureMaker Output**

When the STR profiles for these two individuals are combined to create a 2-person mixture, the mixture profile will contain 1 locus with a single allele, 7 loci with two alleles, 4 loci with three alleles, and 3 loci with four alleles (and no loci with 5 or 6 alleles, which is only possible if one or both samples possess tri-allelic patterns at the same STR locus).

**Future Plans**

- Develop training information based on lessons learned from the MIX05 study
- Create other useful software tools like **mixSTR** and **Virtual MixtureMaker** to increase mixture interpretation capabilities of the forensic DNA typing community
- Conduct mixture interpretation training
  - Develop worked specific examples
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).

• Mixture interpretation consumes a large part of DNA analysts’ time – software tools that improve consistency in analysis will speed casework reporting and hopefully cases solved

Conclusion

“Mixture interpretation theory is well established and used in forensic laboratories. Most mixtures detected in casework are satisfactorily solved. But from this revision we can conclude that the behaviour of each mixed sample can be different and multifactorial and occasionally its interpretation turns out to be complicated—sometimes paralleling the importance of the evidence in the resolution of the case. In some casework mixtures our experience has proved that theoretical assumptions from studies with laboratory samples, albeit very useful, can turn out to be impracticable. We consider that more sharing of day to day forensic laboratory problems is needed to refine our technical procedures in the resolution of specially difficult evidence.”

Acknowledgments

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

NST Human Identity Project Team – Leading the Way in Forensic DNA...

Role in MIX05
- Margaret Kline (running study, sample prep, data review)
- John Butler (study design and data review)
- Becky Hill (GeneMapper® 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