Mixture Interpretation
Issues & Insights

Presentation Outline

- Review highlights from CODIS Oct 2006 mixture talk covering NIST MIX05 interlab study results
- Mixture interpretation protocol and report format variability across the community
- Propose several issues to discuss with a new SWGDAM mixture interpretation subcommittee

Mixtures: 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 other mixtures.

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

NIST and NIJ Disclaimer

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 or the National Institute of Standards and Technology.

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

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

- P. Gill 1,2, C.B. Bruun 1, J.S. Baker-Jones 1, A. Caracozza 1, M. Krawczyk 1, W.R. Mays 1, N. Mortimer 1, M. Prusa 1, F.M. Schneider 1, B.S. West 2

1 Forensic Science Laboratory, Kansas City, KS 66105-2805, USA
2 Research Scientist, Laboratory of Radiological Health, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

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

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 – leading to establishment of different thresholds of detection
- 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/NISTpub.htm
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?

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.

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)

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

Case #3 – balanced mixture (1F:1M)
  - Male lacked amelogenin X

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

Amelogenin X allele is missing in male perpetrator DNA sample for MIX05 Case #3

MIX05 Results on Multiple Kits

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

Case 1 evidence (mixture)

Profiler Plus

COifiler

Identifier

PowerPlex 16

SGM Plus

FMBO data was also made available upon request

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
**Example Mixture Data (MIX05 Study-Profiler Plus)**

<table>
<thead>
<tr>
<th>Loci</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>12,12</td>
<td>28,31.2</td>
<td>15,16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D18S51</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Single Source Sample (Victim)**
- **Evidence Mixture (Victim + Perpetrator)**

**Evidence Mixture**

- **Allele Counts**
  - D8S1179: 12,12
  - D21S11: 28
  - D18S51: 16
  - Amelogenin: 12

- **Obligate Alleles**
  - Y: 2

**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 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/COFlter
- 10 PowerPlex 16
- 7 PIP16 Bio
- 5 Identifiler
- 2 SOM Plus
- 1 All ABI kit data
- 9 Various combinations

Generally Identifier 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

**Summary of Some MIX05 Reported Results**

Case #2 has perpetrator as major component and thus is the easiest to solve...

**Some Mixture Ratios Reported in MIX05**

<table>
<thead>
<tr>
<th>LabID</th>
<th>Case</th>
<th>Calculated (Major/F)</th>
<th>Calculated (Minor/F)</th>
<th>Calculated (Major/M)</th>
<th>Calculated (Minor/M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>2</td>
<td>6</td>
<td>&lt;2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1.8-3.6</td>
<td>3.5-4.7</td>
<td>1.6-1.8</td>
<td>6.2-7.6</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>56 68%</td>
<td>32%</td>
<td>85% 15%</td>
<td>64% 36%</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>21</td>
<td>2.1</td>
<td>6.1</td>
<td>2.1</td>
<td>not determined</td>
</tr>
<tr>
<td>64</td>
<td>2.1</td>
<td>6.1</td>
<td>2.1</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>make3.39%</td>
<td>not determined</td>
<td>make64.71%</td>
<td>make64.71%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>4.1</td>
<td>5.1</td>
<td>1.4</td>
<td>-1.0</td>
</tr>
<tr>
<td>4</td>
<td>10.1</td>
<td>0.1</td>
<td>1.1</td>
<td>not determined</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>make50%</td>
<td>not determined</td>
<td>make50%</td>
<td>make50%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>make50%</td>
<td>make50%</td>
<td>make50%</td>
<td>make50%</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>1.2</td>
<td>3.4</td>
<td>2.1</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>2.1</td>
<td>6.1</td>
<td>1.621</td>
<td>4.451</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>-3.1</td>
<td>-2.1</td>
<td>-4.1</td>
<td>-2.1</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>2.1</td>
<td>5.1</td>
<td>2.1</td>
<td>10.1</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most calls were correct (when they were made)

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

<table>
<thead>
<tr>
<th>Lab ID</th>
<th>Kits Used</th>
<th>Caucasians</th>
<th>African Americans</th>
<th>Hispanics</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>ProPlus/ Ceiling</td>
<td>PE calculated</td>
<td>PE calculated</td>
<td>PE calculated</td>
</tr>
<tr>
<td>73</td>
<td>ProPlus/Ceiling</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>4</td>
<td>ProPlus/Ceiling</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>12</td>
<td>ProPlus/Ceiling</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>29</td>
<td>ProPlus/Ceiling</td>
<td>none provided</td>
<td>none provided</td>
<td>none provided</td>
</tr>
<tr>
<td>30</td>
<td>ProPlus/Ceiling</td>
<td>1.1E+15</td>
<td>2.13E+14</td>
<td>3.08E+15</td>
</tr>
<tr>
<td>34</td>
<td>ProPlus/Ceiling</td>
<td>2.4E+11</td>
<td>7.00E+09</td>
<td>9.86E+10</td>
</tr>
<tr>
<td>33</td>
<td>ProPlus/Ceiling</td>
<td>5.0E+09</td>
<td>3.0E+09</td>
<td>none provided</td>
</tr>
<tr>
<td>33</td>
<td>ProPlus/Ceiling</td>
<td>2.94E+08</td>
<td>1.12E+08</td>
<td>1.74E+09</td>
</tr>
<tr>
<td>6</td>
<td>ProPlus/Ceiling</td>
<td>40,000,000</td>
<td>3,500,000</td>
<td>200,000,000</td>
</tr>
<tr>
<td>9</td>
<td>ProPlus/Ceiling</td>
<td>4.1E+07</td>
<td>1.97E+07</td>
<td>1.54E+07</td>
</tr>
<tr>
<td>61</td>
<td>ProPlus/Ceiling</td>
<td>1.4E+07</td>
<td>230,000</td>
<td>2.49E+07</td>
</tr>
<tr>
<td>79</td>
<td>ProPlus/Ceiling</td>
<td>530,000</td>
<td>47,900</td>
<td>1,350,000</td>
</tr>
<tr>
<td>18</td>
<td>ProPlus/Ceiling</td>
<td>434,800</td>
<td>31,710</td>
<td>399,000</td>
</tr>
</tbody>
</table>

Different Detection Thresholds Used

<table>
<thead>
<tr>
<th>Lab ID</th>
<th>Kits Used</th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>ProPlus/Ceiling</td>
<td>1.1E+15</td>
<td>2.13E+14</td>
</tr>
<tr>
<td>34</td>
<td>ProPlus/Ceiling</td>
<td>2.4E+11</td>
<td>7.00E+09</td>
</tr>
<tr>
<td>33</td>
<td>ProPlus/Ceiling</td>
<td>2.94E+08</td>
<td>1.12E+08</td>
</tr>
<tr>
<td>6</td>
<td>ProPlus/Ceiling</td>
<td>40,000,000</td>
<td>3,500,000</td>
</tr>
<tr>
<td>9</td>
<td>ProPlus/Ceiling</td>
<td>4.1E+07</td>
<td>1.97E+07</td>
</tr>
<tr>
<td>79</td>
<td>ProPlus/Ceiling</td>
<td>530,000</td>
<td>47,900</td>
</tr>
<tr>
<td>18</td>
<td>ProPlus/Ceiling</td>
<td>434,800</td>
<td>31,710</td>
</tr>
</tbody>
</table>

~10 orders of magnitude difference (10^7 to 10^15) based on which alleles were deduced and reported

Remember that these labs are interpreting the same MIX05 electropherograms.
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)

Manually Solving Mixture Component Profiles

Lab 90 – correctly deduced all perpetrator alleles in Case #1
(highest of the 7 listed stats for Profiler Plus/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.

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...

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
Some Labs Do Not Attempt Mixture Interpretation

- A number of laboratories chose not to report anything in the MIX05 study citing that without a suspect, mixtures are not examined.

- Why does a National DNA Database such as CODIS exist and how can it be helpful and reach its full potential if casework mixtures are not examined and perpetrator alleles deduced (where possible)?

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 rerun 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)
  - Available for use at https://lsd.lit.net/

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

USACIL program developed by Tom Overson
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.”

Thank you for your attention…

Questions or Comments?

http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov
301-975-4049

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

MIX05 Acknowledgments

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

NIST 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 ID data review)
- Jan Redman (Access database entry, shipping)
- Dave Duewer (Virtual Mixture Maker 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