Advanced Topics in Forensic DNA Analysis

Mixture Interpretation

New Jersey State Police Training Workshop
Hamilton, NJ December 5-6, 2006

Dr. John M. Butler
National Institute of Standards and Technology
john.butler@nist.gov

CODIS Conference – October 23, 2006

Presentation Outline

• Mixtures: issues and challenges
• MIX05 interlaboratory study (initiated at CODIS Conference Nov 15, 2004)
• Mixture interpretation variation – future role of expert systems
• Opportunities for community improvement and standardization regarding mixture interpretation

Other Session Speakers
Elizabeth Johnson – software demo of USACIL 2-component mixture ratio program
Angelo Della Manna – case examples and CODIS search strategies with mixtures

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.

Even more challenging with poor quality data when degraded DNA is present...

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

<table>
<thead>
<tr>
<th>Single Source Sample (Victim)</th>
<th>Evidence Mixture (Victim + Perpetrator)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amelogenin</td>
<td>D8S1179, D21S11, D18S51</td>
</tr>
<tr>
<td>Y</td>
<td>12, 28, 15</td>
</tr>
<tr>
<td>X,Y</td>
<td>12, 12</td>
</tr>
</tbody>
</table>

Obligate Alleles (not present in the victim reference)

<table>
<thead>
<tr>
<th>True &quot;Perpetrator&quot; Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
</tr>
<tr>
<td>X,Y</td>
</tr>
</tbody>
</table>

Mixtures: 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.

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

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

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

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.

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

### NIST Initiated Interlaboratory Studies

<table>
<thead>
<tr>
<th>Studies involving STRs</th>
<th>&quot;Other&quot;</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture Interpretation Study (Jan - Aug 2005)</td>
<td>69</td>
<td>Poster at 2005 Promega meeting (Sept 2005); available on STRBase</td>
</tr>
</tbody>
</table>
**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)

**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

**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, CODfiler, SGM Plus, PowerPlex 16, Identifier, PowerPlex 16 BIO (FMBIO) kits

- Summary of results will involve training materials to illustrate various approaches to solving mixtures

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

http://www.cstl.nist.gov/biotech/strbase/training.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
COflfer
Identifiler
PowerPlex 16
SGM Plus

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

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

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

STR kit results used
- 34 ProfilerPlus/COFiler
- 10 PowerPlex 16
- 7 PP16 BIO
- 5 Identifier
- 2 SGM 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

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

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

Some Reported Stats for MIX05 Case #1

Many of the 29 labs providing statistics used PopStats 5.7
Some Differences in Reporting Statistics

<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>30</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>34</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>6</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>9</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>79</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</td>
</tr>
<tr>
<td>15</td>
<td>PrePlux/Coffler</td>
<td>2.4E+11</td>
<td>7.0E+10</td>
<td>9.8E+10</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

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

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)
Different Reporting Formats for MIX05 Data

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

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)?

Value of the MIX05 Study

- 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

• 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

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](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).

mixSTR S/E output

Example of suspect to evidence (S/E) comparisons made in this case. Note that the suspect is 21,23 at FGA while the evidence contains 23,24* (* indicates that allele 24 is a minor component). Thus this suspect has allele 23 in common and is missing allele 24 in the evidence.

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).
Virtual MixtureMaker Output

One tri-allelic locus
One locus with 3 alleles in this 2-person mixture
16 loci examined with 31 distinguishable alleles
No locus failures in this profile
13 heterozygous loci
2 homozygous loci

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 another interlab study in 2007 (MIX07)?
  – To try and capture improved knowledge regarding mixture interpretation and capabilities of expert systems

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

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 MixtureMaker to aid sample selection; mixSTR program)
- Chris Tomsey & Frank Krist (FMBIO Mac data)
- Kermit Channel & Mary Robnett (FMBIO NT data)
- Mandy Sozer for early discussions on study design

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