STRmix™ collaborative exercise on DNA mixture interpretation

Kevin Cheng et al.
Accuracy and Precision
Forensic DNA Analysis

- Crime Sample
- DNA Extraction
- DNA Amplification
- Separation of Analyte Signals
- Match Statistic Calculation
- Analyst Interpretation
- Signal Visualization Algorithm
- Signal Detection
Previous Inter-laboratory Studies

• The Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles (L.Prieto et al., 2014)

• GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06) (P.A.Barrio et al., 2018)

• Investigating a common approach to DNA profile interpretation using probabilistic software (Early STRmix™ Study) (Stuart Cooper et al., 2015)

• NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): Variation observed and lessons learned (NIST MIX13) (John M. Butler et al., 2018)
Euroforgen-NoE & GHEP-MIX06

- Euroforgen-NoE Case 1
- Euroforgen-NoE Case 2
- GHEP-ISFG

Graph showing log(LR) on the y-axis and categories on the x-axis.
Euroforgen-NoE & GHEP-MIX06

![Graph showing peak heights and unresolved peak signal.]

- **D12S391**
- Unresolved Peak Signal

**Peak Height (rfu)**

- 0
- 100
- 200
- 300
- 400
- 500
- 600
- 700
- 800
- 900
- 1000
- 1100
- 1200

**Time (Minutes)**

- 18
- 19
- 19.3
- 21
- 25
Early STRmix™ Study
NIST MIX05 & MIX13

Mostly undertaken using manual interpretation methods

Profile Generation

Signal Visualization
- Analytical Threshold
- Stutter Filter

Analyst Interpretation
- Number of Contributors
- Propositions
- Stochastic Threshold

Match Statistic Calculation
- Calculation Method
- Allele Frequencies
Lessons Learnt

• Probabilistic Genotyping Software automate some decisions

• There is still some subjective decisions impacting the precision

• Fuelled discussions for more standardization
STRmix™ Collaborative Exercise

• An intra- and inter-laboratory study using the probabilistic genotyping (PG) software, STRmix™.

• Aim to identify the sources of variation in the reported LR.

• Forty-two (42) laboratories
• One-hundred seventy-four (174) participants
  • Multiple submissions
  • Or did not interpret
**STRmix™ Collaborative Exercise**

Two complex mixtures from the PROVEDIt set, analysed on an Applied Biosystems™ 3500 Series Genetic Analyzer. (Lauren E. Alfonse et. al, 2018)

Corresponding STRmix™ parameter files.

- Profile Generation
- Signal Visualization
  - Analytical Threshold
  - Stutter Parameters
- Analyst Interpretation
  - Number of Contributors
  - Propositions
  - Stochastic Threshold
- Match Statistic Calculation
  - Calculation Method - LR
  - Allele Frequencies
The DNA profile was obtained from a semen stained anal swab after an alleged sexual assault. The male complainant alleges he has been sexually assaulted by two male individuals.

DNA swabs from the complainant and a person of interest have been taken for analysis.
Sample Two

- High-template profile, 0.75 ng
- Designed as a 3-person mixture
- 1:4:4 Mixture Ratio
Sample Two

- Total of 176 submissions
- Three-contributors (151/176)
- Four-contributors (25/176)
Sample Two

- Lowest $LR$ submission ignored CSF1PO
  $\log(LR) = 28.3$

- Maximum $LR$ submission used a bespoke method to manage artefacts
  $\log(LR) = 29.4$
Sample Two

- Excluding the outliers $\log(LR)$ range from 29.0 to 29.1
  - Regardless of $NoC$

- When excluding the outliers the intra-laboratory $\log(LR)$ range = 0.05
Sample Two

• STRmix™ utilizes an MCMC (Markov chain Monte Carlo) interpretation method.

• MCMC involves Random Sampling.

• $LR$ variation for each method can be attributed to the MCMC sampling.
Sample One

The DNA profile was obtained from a semen stained sample from underwear collected from the complainant after an alleged sexual assault. The male complainant alleges he has been sexually assaulted by two male individuals.

DNA swabs from the complainant and a person of interest have been taken for analysis.
Sample One

- Low-template profile, 0.105 ng
- Designed as a 4-person mixture
- 4:1:1:1 Mixture Ratio
Sample One

- Five (5) participants did not interpret
- Total of 173 submissions
- Three-contributors (11/173)
  - Nine exclusions ($LR = 0$)
  - Two inclusions by ignoring D18S51 during interpretation.
D18S51, Assuming 3 Contributors

Drop-in Cap
Pr(C) = 0.0001
Sample One

- Four-contributors (162/173)

- One submission did not condition on the complainant log($LR$) = 5.7

- log($LR$) range from 4.3 to 6.6

- Intra-laboratory log($LR$) Range = 2.09
The large range of $LR$s of 2.3 may be too broad to be explained by MCMC variation.
Sample One

- Large differences in the \( LR \) was due to differences in analysis methods, causing differences in labelled peaks and peak heights.

<table>
<thead>
<tr>
<th>Method</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothing</td>
<td>Light</td>
<td>Light</td>
<td>Light</td>
<td>No</td>
</tr>
<tr>
<td>Normalization</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Baseline Window</td>
<td>51</td>
<td>51</td>
<td>33</td>
<td>51</td>
</tr>
</tbody>
</table>
Analysis Methods – TPOX
Sample One

- Remaining variability observed in the $LR$ can be attributed to MCMC variation.
Challenges

Sample One

• Complex Profile
• Low-template, four-contributors
• Several participants commented that due to the complexity of the profile, in a casework situation they would not interpret the profile
• Some would like a PCR replicate to aid in interpretation of the profile

Sample Two

• Less ambiguous profile. Slight differences in interpretation
STRmix™ Collaborative Exercise

Aim to identify the sources of variation in the reported LR

- Varying the number of contributors assumed when interpreting a profile
- Exclusion of some loci when interpreting a profile
- Differences in CE data analysis methods, leading to variation in peak heights
- Run-to-run variation due to random sampling inherent to the MCMC method
Research paper

STRmix™ collaborative exercise on DNA mixture interpretation

References


Kevin Cheng

E: kevin.cheng@esr.cri.nz