Introduction to STRmix™ and Likelihood Ratios

INDIANA STATE POLICE LABORATORY

BIOLOGY SECTION

Introduction

- On November 1, 2017, the Indiana State Police (ISP) Laboratory Biology Section began utilizing STRmix™ probabilistic genotyping software to aid in the interpretation and statistical evaluation of DNA profiles including some previously uninterpretable mixed DNA samples.
- Goals of this presentation:
  - Educate on the motivation for change
  - Lay foundation for acceptance in court
  - Introduce the Likelihood Ratio

Introduction

What is NOT changing?

- Serological testing
- Process for developing a DNA profile
- Our understanding of the way DNA profiles behave
- Other types of DNA testing (i.e. Y-STRs, relationship comparisons)

What is changing?

- Interpretation methodology of DNA profiles, particularly mixtures
- Statistical evaluation of DNA profiles (i.e. Likelihood Ratios)
- Ability to interpret more complex samples

Why are we changing?

- Advances in science and technology are always improving
- Desire to provide more informative and relevant results to our customers

When will these changes take effect?

- Started Nov. 1, 2017

What is STRmix™?

STRmix™ is a fully-continuous probabilistic genotyping software that interprets and evaluates complex DNA profiles.

- "Probabilistic Genotyping" refers to the use of biological modeling, statistical theory, computer algorithms, and probability distributions to infer genotypes and calculate likelihood ratios for the DNA profiles developed from forensic samples.

STRmix™ was created in 2011 jointly by forensic scientists at the New Zealand Institute of Environmental Science and Research (ESR) and Forensic Science South Australia (FSSA).

- Primary developers are Dr. John Buckleton and Dr. Jo-Anne Bright from ESR and Dr. Duncan Taylor from FSSA.
- Initially intended to be utilized exclusively in New Zealand and Australia, STRmix™ has been adopted and used in forensic casework in more than 25 local, state, federal, and private laboratories all across the U.S., as well as labs in England, Ireland, Scotland, and Canada.

What is STRmix™?

- A short animated video explaining STRmix™ is available at:
  - http://strmix.esr.cri.nz/

Motivation for Change

Why is STRmix™ necessary?

Traditional DNA interpretation methods waste information due to the complexity of DNA mixtures and millions of potential explanations, forcing the DNA analyst to simplify assumptions and discard considerable identification information.

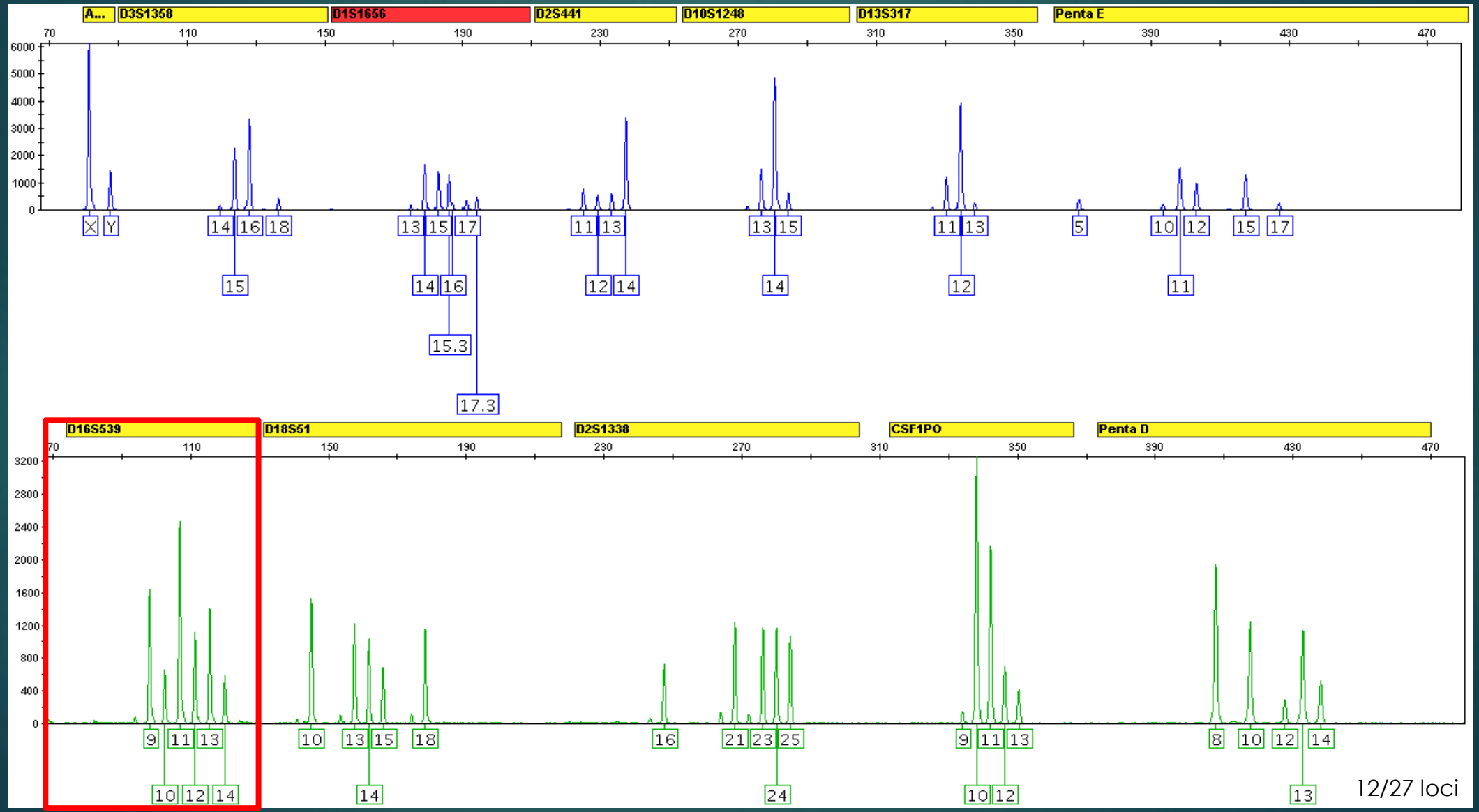
- Due to these limitations, many complicated DNA profiles could not be interpreted.
  - Profiles with 3 or more contributors; low level/poor quality profiles; profiles impacted by environmental factors causing degradation – most often associated with touch DNA samples
- Those samples that are deemed interpretable consider all genotypes to be equally probable and reduces the discriminatory power of the system.
- Training and experience can also affect interpretations such that the same profile may result in slightly different conclusions among different analysts and laboratories.

Why is STRmix™ necessary?

STRmix™ uses all of the genetic information available (including peak heights, stutter percentages, and mixture proportions) to make full use of the data and provide weightings to different genotypes such that some are deemed more probable than others.

- This enhances the ability to distinguish true donors and non-donors to the DNA profile in question.
- The only limitations to the software are the analyst's ability to make a reasonable assumption regarding the number of contributors (NoC) and the processing power of the computer used for interpretation.
- The software reduces profile interpretation variability and analyst bias.

Example evidence profile

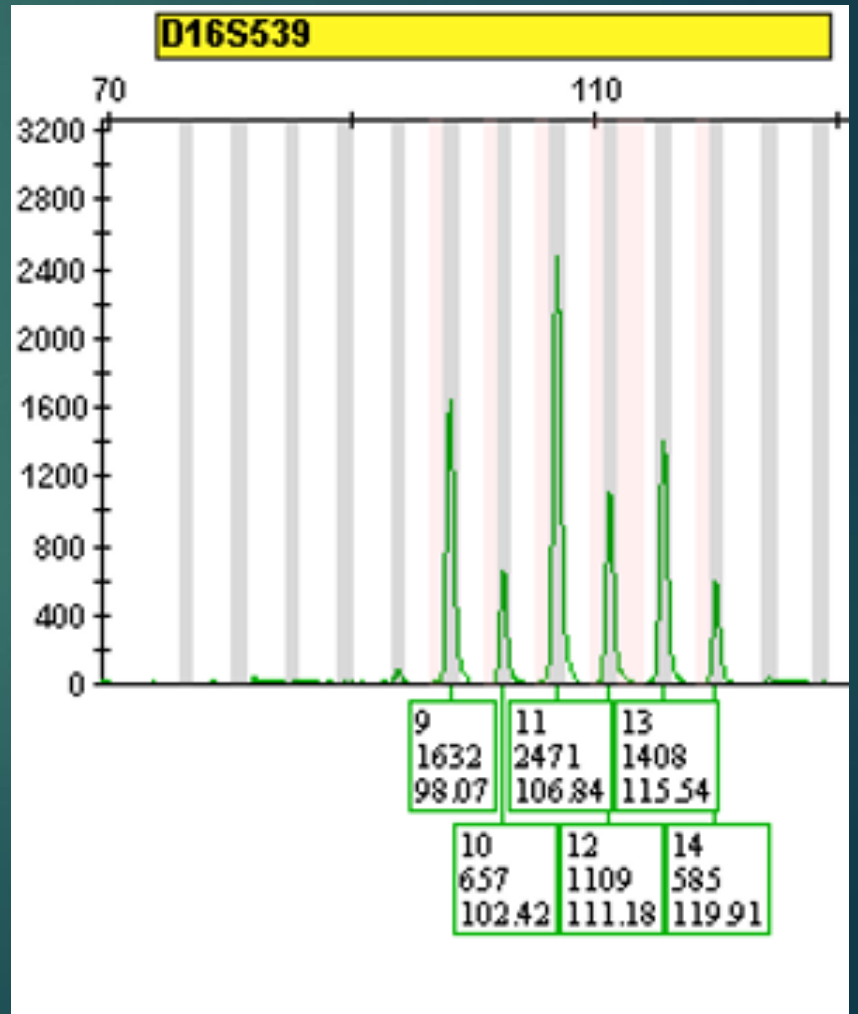


Traditional methods

Possible genotype combinations

Assuming 3 contributors (no allelic drop-out)

9,10	11,12	13,14	N
9,10	11,13	12,14	N
9,10	11,14	12,13	N
9,11	10,12	13,14	Y
9,11	10,13	12,14	Y
9,11	10,14	12,13	Y
9,12	10,11	13,14	Y
9,12	10,13	12,14	Y
9,12	10,14	11,13	Y
9,13	10,11	12,14	Y
9,13	10,12	11,14	N
9,13	10,14	11,12	Y
9,14	10,11	12,13	N
9,14	10,12	11,13	Y
9,14	10,13	11,12	N



Traditional methods

Limitations

Due to low level data and complexity of the mixture, the analyst may not be comfortable assuming 3 contributors.

- Unable to appropriately apply statistical estimate without considering NoC

Some genotype combinations would seem more reasonable than others

- Traditional interpretation methods are considered "binary", analyst must decide "yes" or "no"

Genotype combinations become more complicated when:

- Alleles are potentially shared
- Data is low/poor quality
- Number of contributors increases

Statistical estimates are restricted to 1-, 2-, and some 3-person samples.

Traditional methods

Conclusions

"The DNA profile demonstrated the presence of a mixture in which the number of contributors cannot reasonably be assumed. Therefore, no further conclusions were drawn."

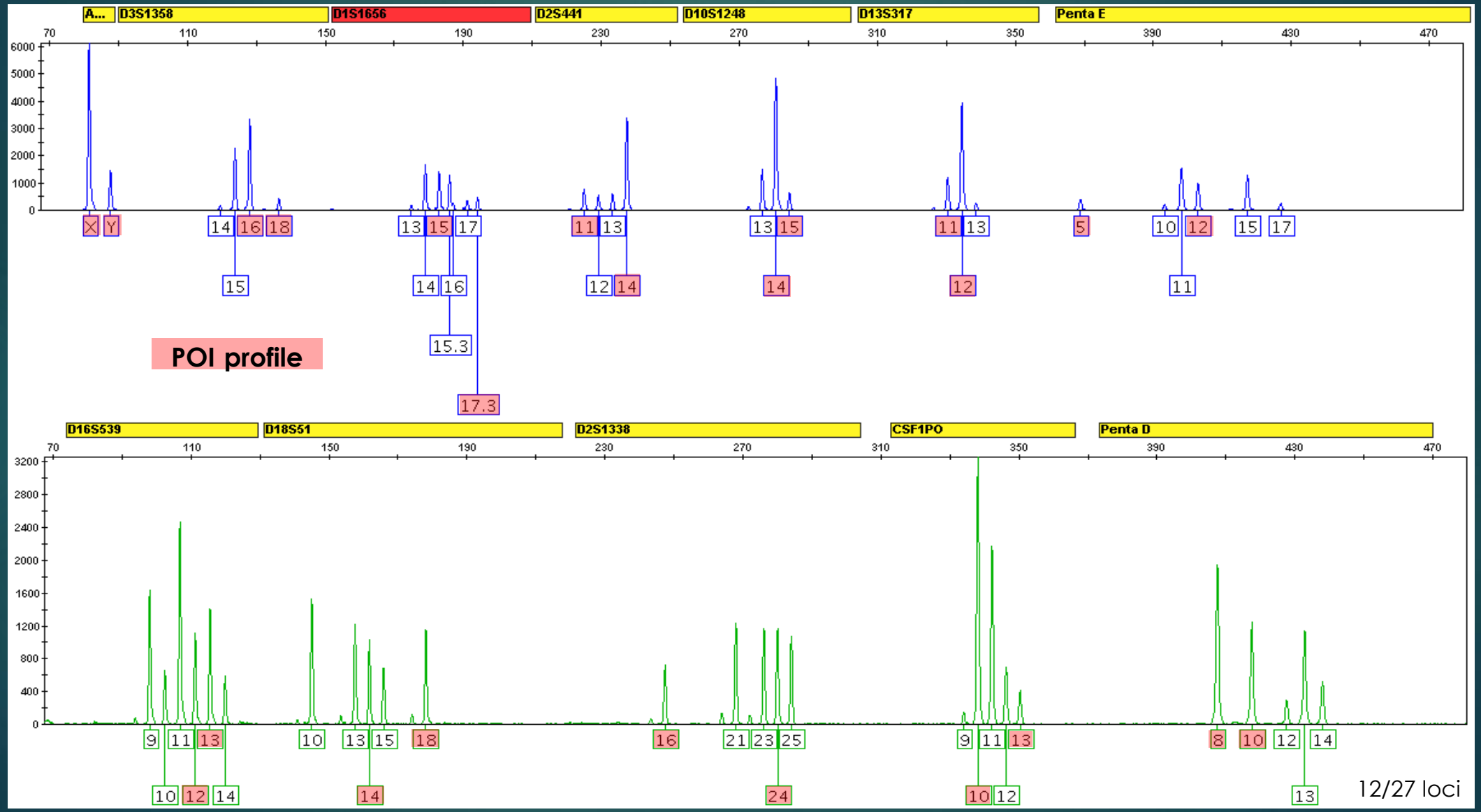
- Due to low quality and complexity of the observed profile, the DNA analyst may feel uncertain as to the number of contributors. Traditional DNA interpretation methods do not handle uncertainty well.

"The DNA profile demonstrated the presence of a mixture of at least three individuals. The results do not qualify for statistical calculations; therefore, no further conclusions were drawn."

- The DNA analyst may be comfortable with making a reasonable assumption as to the number of contributors; however, due to the low quality and complexity of the observed profile, traditional DNA statistical evaluations of the evidence cannot be responsibly applied without advanced software.

This sample is considered **INCONCLUSIVE** and the DNA analyst is unable to provide any information (inclusion or exclusion) to the court.

Example evidence profile



12/27 loci

Traditional methods

Limitations

By allele-centric, visual comparison alone, Person of Interest (POI) appears represented

Analysis of evidence profile must be performed independent of any reference profiles except for intimate body samples.

- Reduces bias by eliminating "painting the target around the arrow"
  - Drawing conclusions based on the opinion that the POI is included

Must have a valid method for calculating statistical estimates of the evidence profile

- Limitations in the ability to accurately and appropriately apply stats may result in samples with which we cannot draw any conclusions, including exclusions.
- National standards require providing statistical weight to present DNA conclusions.

Probabilistic Genotyping

Advantages

More elimination profiles may be used for interpretation of evidence profiles thanks to software that does the complex comparisons.

- Increases information resulting in more appropriate interpretation and relevant statistical evaluation

Methods for statistical estimates are far more robust.

- Only limitations are:
  - The analyst's ability to make a reasonable assumption regarding the number of contributors
  - The computer processing power to deconvolute/resolve the DNA profile into separate components
    - Currently able to process up to 4-person mixtures

Probabilistic Genotyping

Advantages

Analyst may be unsure whether to interpret as a 3-person or a 4-person mixture.

- Allow STRmix™ to interpret as both and evaluate deconvolution diagnostics to determine which is most appropriate.

STRmix™ assigns a probability to different genotype combinations.

- Software assigns weights, those that are most reasonable are given the highest weight.

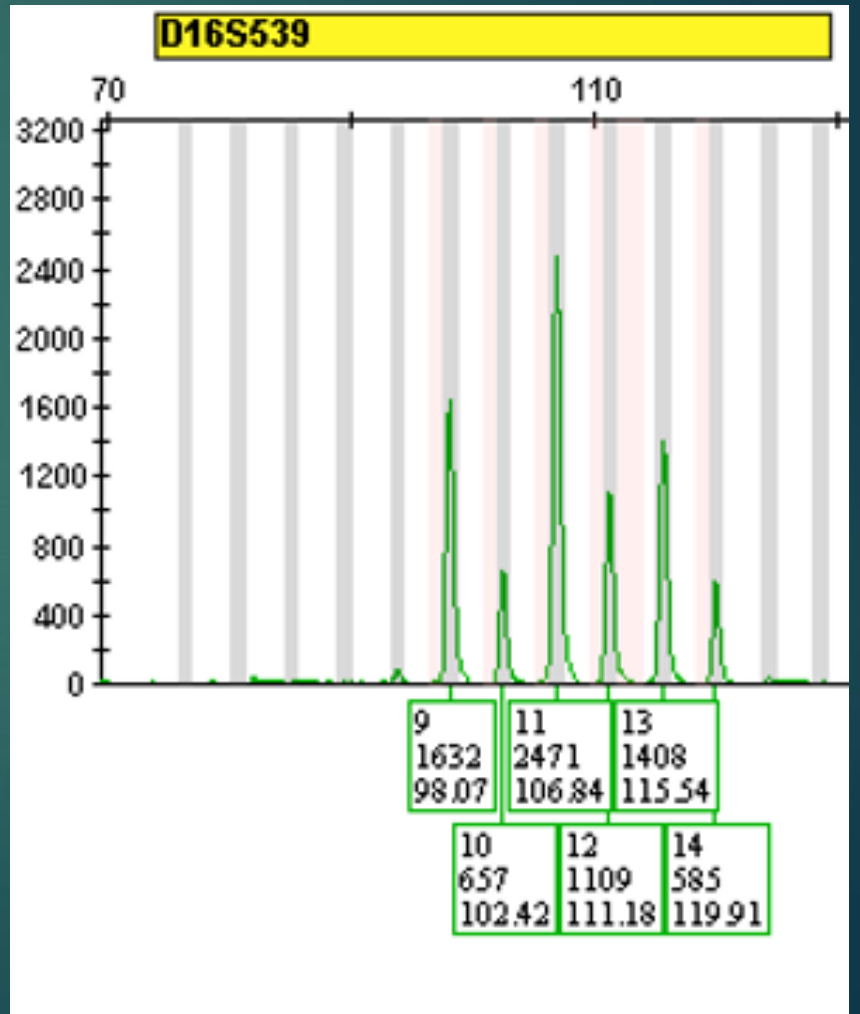
STRmix™ can quickly and easily handle very complex mixtures that the human brain cannot process.

Probabilistic genotyping

Possible genotype combinations

Assuming 3 contributors (not considering contributor order)

9,10	11,12	13,14	**3%**
9,10	11,13	12,14	**4%**
9,10	11,14	12,13	**2%**
9,11	10,12	13,14	**9%**
9,11	10,13	12,14	**7%**
9,11	10,14	12,13	**45%**
9,12	10,11	13,14	**1%**
9,12	10,13	12,14	**1%**
9,12	10,14	11,13	**16%**
9,13	10,11	12,14	**3%**
9,13	10,12	11,14	**1%**
9,13	10,14	11,12	**4%**
9,14	10,11	12,13	**--**
9,14	10,12	11,13	**2%**
9,14	10,13	11,12	**2%**



Probabilistic genotyping

Conclusions

"The DNA profile was interpreted as originating from three individuals. The probability of the evidence has been calculated by considering the following propositions:

H1: The evidence originated from John Doe and two unknown individuals.

H2: The evidence originated from three unknown, unrelated individuals.

The DNA profile is at least ***LR*** times more likely if it originated from John Doe and two unknown individuals than if it originated from three unknown individuals. This analysis provides ***weak/moderate/strong*** support for the proposition that John Doe is a contributor to the DNA profile."

Foundation for acceptance

Is STRmix™ validated?

Yes

- STRmix™ has undergone rigorous developmental validation and is internally validated by each laboratory utilizing it in casework, including the Indiana State Police, for their specific laboratory system (i.e. chemistry, equipment, and procedures).
  - Developmental validation and FBI internal validation are both published in peer-reviewed scientific publications
- The Indiana State Police validation of STRmix™ followed the recommendations of the program developers as well as the 2015 SWGDAM Guidelines for Validation of Probabilistic Genotyping Systems. https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf

Is the theory behind STRmix™ generally accepted?

Yes

- Biological model (the way DNA profiles behave) is well-known in the forensic DNA community and has been utilized for years even with traditional interpretation methods.
  - Genotypes, DNA amount, degradation, amplification efficiency, stutter, peak variance
- Papers describing the biological model, mathematics, performance and validation of STRmix™ have been published in various peer-reviewed forensic journals.
- Probabilistic genotyping is also the recommended approach for interpretation of low-level DNA profiles and complex mixtures.

Published references

1. D.A. Taylor, J.-A. Bright, J. S. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics. 7(5) (2013) 516-528.
2. J.-A. Bright, D.A. Taylor, J. M. Curran, J. S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Science International: Genetics. 7(2) (2013) 296-304.
3. J.-A. Bright, D.A. Taylor, J. Curran, J.S. Buckleton, Degradation of forensic DNA profiles, Australian Journal of Forensic Sciences. 45(4) (2013) 445-449.
4. J.-A. Bright, D.A. Taylor, J. M. Curran, J. S. Buckleton, Searching mixed DNA profiles directly against profile databases Forensic Science International: Genetics. 9 (2014) 102-110.
5. D.A. Taylor. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics, 2014. 11: 144-153.
6. J.-A. Bright, J.M. Curran and J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. Forensic Science International: Genetics, 2014. 12: 208-214.
7. J.-A. Bright, I.W. Evett, D.A. Taylor, J.M. Curran and J.S. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics, 2015. 14: 125-131.
8. J.-A. Bright, K.E. Stevenson, J.M. Curran and J.S. Buckleton, The variability in likelihood ratios due to different mechanisms. Forensic Science International: Genetics, 2015. 14:187-190.
9. D.A. Taylor, J.-A. Bright and J.S. Buckleton, Considering relatives when assessing the evidential strength of mixed DNA profiles. Forensic Science International: Genetics, 2014. 13: 259-263.
10. T.W. Bille, S.M. Weitz, M.D. Coble, J.S. Buckleton and J.-A. Bright, Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. Electrophoresis, 2014. 35:3125-33.

Published references

11. D.A. Taylor, J.-A. Bright and J.S. Buckleton, The 'factor of two' issue in mixed DNA profiles. Journal of Theoretical Biology, 2014. 363: p. 300-306.
12. D.A. Taylor and J.S. Buckleton, Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics, 2015. 16:13-6.
13. D.A. Taylor, J.S. Buckleton and I. Evett, Testing likelihood ratios produced from complex DNA profiles. Forensic Science International: Genetics, 2015. 16:165-171.
14. S.J. Cooper, C.E. McGovern, J.-A. Bright, D.A. Taylor and J.S. Buckleton, Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics, 2015. 16:121-131.
15. J.-A. Bright, D.A. Taylor, C.E. McGovern, S.J. Cooper, L.J. Russell, D.V. Abarno and J.S. Buckleton, Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics, 2016. 23:226-239.
16. D.A. Taylor, J.-A. Bright, C.E. McGovern, C. Hefford, T. Kalafut, J.S. Buckleton, Validating multiplexes for use in conjunction with modern interpretation strategies. Forensic Science International: Genetics, 2016. 20:6-19.
17. D.A. Taylor, J.S. Buckleton, J.-A. Bright, Factors affecting peak height variability for short tandem repeat data. Forensic Science International: Genetics, 2016. 21:126-133.
18. T.R. Moretti, R.S. Just, S.C. Kehl, L.E. Willis, J.S. Buckleton, J.-A. Bright, D.A. Taylor, Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics, 2017. 29:126-144.
19. D.A. Taylor, J.S. Buckleton, J.-A. Bright, Does the use of probabilistic genotyping change the way we should view sub-threshold data? Australian Journal of Forensic Sciences, 2017. 49(1):78-92.

Have STRmix™ results been accepted in U.S. courts?

Non-exhaustive list of U.S. cases per https://johnbuckleton.wordpress.com/strmix/

1. Michigan v Elamin Muhammad,

   Daubert - December 3, 2015
2. New York v Vincent Bullard-Daniel

   March 10, 2016
3. State of Texas v Michael Shane Clack

   Daubert/admissibility hearing
4. State of Texas v Roy Edward Smith
   - Appeal Roy Edward Smith v Texas

     Judgement affirmed – April 28, 2017
5. Henry Watkins Skinner v State of Texas

   June 8, 2016
6. Michigan v Irby

Have STRmix™ results been accepted in U.S. courts?

7. Michigan v Sayers
8. Michigan v Herbert Maurice Alford
   Daubert - November 28, 2016
9. Florida v Marc Regisme
   Daubert hearing denied
10. United States v Pettway (New York)
    https://casetext.com/case/united-states-v-pettway-7
11. New York v Oral Nicholas Hillary
    STRmix™ found to be generally accepted, but not internally validated by the testing laboratory
12. Michigan v Larry David Smith
    Daubert – May 3, 2017
13. Florida v Dwayne Cummings
    Daubert – May 12, 2017
14. Michigan v Marlon Anthony Burns
    Daubert – July 27, 2017

Have STRmix™ results been accepted in Indiana courts?

Not yet

- STRmix™ was implemented for use in the Indiana State Police Laboratory system on November 1, 2017.
- Unless admitted within last ~2 weeks (time this PowerPoint was submitted to IPAC and presented today).

Have any probabilistic genotyping results been accepted in Indiana courts?

Yes. There are several other probabilistic genotyping systems available. One of them is Cybergenetics TrueAllele® system.

- Cybergenetics will perform a free evaluation of some evidence profiles upon request, official results will need to be paid for, but there have been a few counties in the state that taken it upon themselves to pursue this analysis.

1. **Indiana v Dugniqio Dishay Forest** (6/3/2016)
   82D03-1501-F2-000566 – Vanderburgh County
   Defendant withdrew objection prior to conclusion of Daubert – court found TrueAllele® scientifically reliable
2. **Indiana v Malcolm Bryan Wade** (8/1/2016)
   53C02-1411-F3-001042 – Monroe County
   No apparent challenge
3. **Indiana v Randal L. Coalter** (8/3/2017)
   62C01-1703-MR-000192 – Perry County
   Motion to exclude denied

Introduction to Likelihood Ratios

Likelihood Ratios

Definition

A mathematical relationship between 2 different/competing explanations for the evidence profile

- Explanations are generally referred to as **Hypotheses** or **Propositions**
  - H1: Inclusionary (Prosecutor's) Hypothesis
    - "The profile originated from John Doe"
  - H2: Exclusionary (Defense) Hypothesis
    - "The profile originated from an unknown, unrelated individual"
- LR is the relationship between these 2 propositions
  - Divide Probability of Evidence profile if H1 by Probability of Evidence profile if H2
    - $LR = Pr(E|H1)/Pr(E|H2)$
  - LR = 1 is indicates both explanations are equally probable
  - LR > 1 favors H1
  - LR < 1 favors H2

Likelihood Ratios

Comparison to traditional statistics (RMP)

- **Random Match Probability (RMP)**
  - Estimates how often we expect to find someone in the general population who could have the evidence profile.
  - Expressed as a frequency
    - "The profile is estimated to occur once in **#** unrelated individuals."
  - Value is *completely independent* of any reference profiles
- **Likelihood Ratio**
  - Value is *entirely dependent* upon the possible contributors and specific propositions being considered
    - Appropriate case information and reference profiles are critical to develop most relevant LR(s)
- **Both provide weight to a DNA profile**
  - Distinguish between profiles with a lot of information (discriminatory power) and profiles with very little information

Likelihood Ratio

1-Person Example

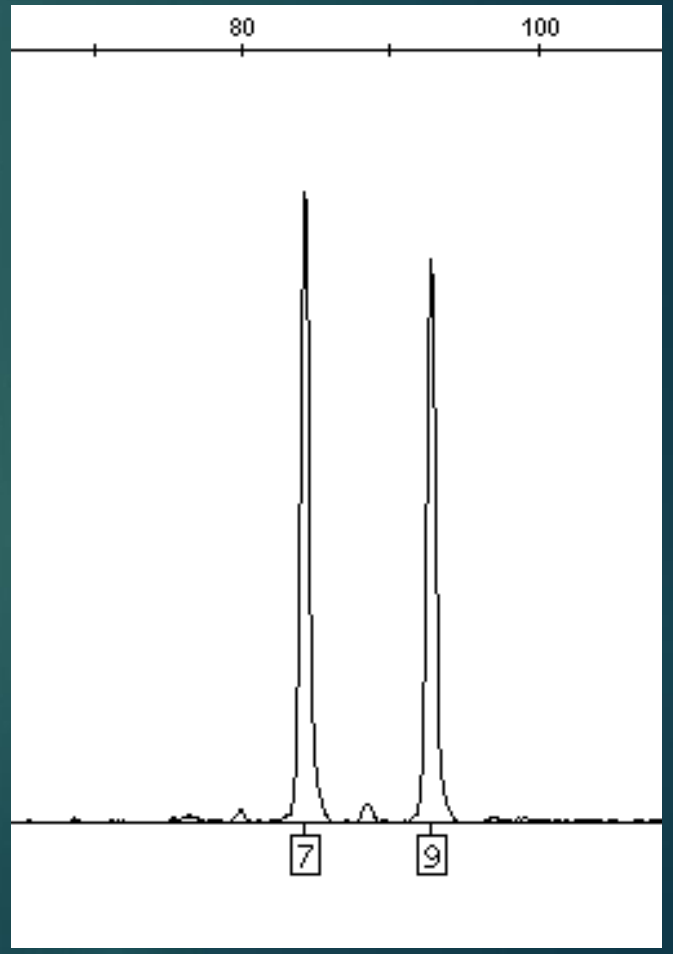
- Evidence Profile: 7, 9
- Person Of Interest: 7, 9

- H1: Probability of Evidence Profile if POI is the source

  = **1**
- H2: Probability of Evidence Profile if UNK is the source

  = Frequency of the 7, 9 genotype ≈ **5.75%**

- LR = 1/.0575 = **17.4**



Likelihood Ratios

Reporting

- LR > 1

  "The DNA profile is at least ***LR*** times more likely if it originated from John Doe than if it originated from an unknown individual."

- LR < 1

  "The DNA profile is at least ***LR*** times more likely if it originated from an unknown individual than if it originated from John Doe."

  - LR has been inverted to be more easily understood and Propositions re-ordered

- LR will be capped at 1 trillion to avoid reporting incomprehensible numbers
  - What is a duodecillion?

Likelihood Ratios

Verbal Equivalent

- Used to assist with understanding the magnitude of the LR.
- Scale determined by evaluating literature sources, SOPs from other labs, and from internal validation of STRmix™.
- "This analysis provides ***weak/moderate/strong*** support for the proposition that John Doe is a contributor to the DNA profile."

Likelihood Ratio (or 1/LR)	Verbal equivalent
$1 \leq LR < 10$	uninformative
$10 \leq LR < 100$	provides weak support
$100 \leq LR < 1000$	provides moderate support
$1000 \leq LR$	provides strong support

Likelihood Ratio

Conclusions

"The DNA profile was interpreted as originating from a single individual. The probability of the evidence has been calculated by considering the following propositions:

H1: The evidence originated from John Doe.

H2: The evidence originated from an unknown, unrelated individual.

The DNA profile is at least ***LR*** times more likely if it originated from John Doe than if it originated from an unknown individual. This analysis provides ***weak/moderate/strong*** support for the proposition that John Doe is a contributor to the DNA profile."

Likelihood Ratio

Alternate Propositions

- The LR is dependent upon the Propositions
  - More than one proposition set may be proposed/calculated/reported
- Example: 2-Person Mixture
  - Neither Suspect 1 nor Suspect 2 are excluded

H1: DNA profile originated from Suspect 1 and Suspect 2

$H2_a$: DNA profile originated from 2 unknown individuals

OR

$H2_b$: DNA profile originated from Suspect 1 and an unknown individual

OR

$H2_c$: DNA profile originated from an unknown individual and Suspect 2

Likelihood Ratio

Exclusions

DNA Analyst may exclude POI visually.

- "John Doe has been excluded as a possible contributor to this profile."
  - Any reference profiles listed as "used for comparison purposes" have been compared and may not be explicitly stated with the evidence result as listed above.

Complex mixtures may use STRmix™ to assist with interpretation

- LR = 0

  "… Statistical analysis of these two propositions provided no scientific support that the mixed DNA profile originated from John Doe and two unknown individuals. Based on the propositions detailed above, John Doe is excluded as a contributor to the DNA profile."

- 0 < LR < 1

  "… The DNA profile is at least ***LR*** times more likely if it originated from an unknown individual than if it originated from John Doe. This analysis provides ***weak/moderate/strong*** support for the proposition that John Doe is excluded as a contributor to the DNA profile."

Likelihood Ratio

Stating the LR

LRs are to be expressed as stated in the previous slides.

- "The DNA profile is at least ***LR*** times more likely if ***H1*** than if ***H2***."

Verbal equivalent is used to assist with understanding.

- "This analysis provides ***weak/moderate/strong*** support for the proposition ***H1***".

Any attempt to restate the LR in another form is a misrepresentation of the analysis that was performed.

- Converting to a frequency:

  "The probability of observing this profile is 1 in #."

  "If there are 6.5 million people in Indiana, we could expect # of them to also be included in this mixture."

- Stating the probability of the proposition:

  "It is # times more likely that John Doe is a contributor."

Re-evaluation of previously analyzed cases

Implementation of STRmix™ does not invalidate previously reported results. It is only expected to aid interpretation of select sample types.

- Inconclusive samples or samples with low statistical weight will benefit the most from STRmix™ analysis

Non-property crime cases will be re-evaluated by the reporting analyst back to ~August 1, 2016.

- Corresponds to the switch in chemistry used to develop DNA profiles which has been validated for use with STRmix™
- Expect results from any re-analysis to be completed by March 1, 2018.

Cases analyzed prior to August 1, 2016 would need a request from the submitting agency or prosecutor's office.

- Requires re-analysis using the new chemistry
- Contact the reporting analyst to initiate the re-evaluation

Questions?

- Evansville Regional Laboratory
  19411 Highway 41 North
  Evansville, IN 47725
  812-867-3157

- Fort Wayne Regional Laboratory
  5811 Ellison Road
  Ft. Wayne, IN 46804
  260-436-7522

- Indianapolis Regional Laboratory
  550 W. 16th St.
  Indianapolis, IN 46202
  317-921-5300

- Lowell Regional Laboratory
  1550 East 181st Ave
  Lowell, IN 46356
  219-696-1835

- Bobb Dilley
  rdilley@isp.in.gov
  (317)921-5355