

PRESENTING STRMIX RESULTS IN COURT

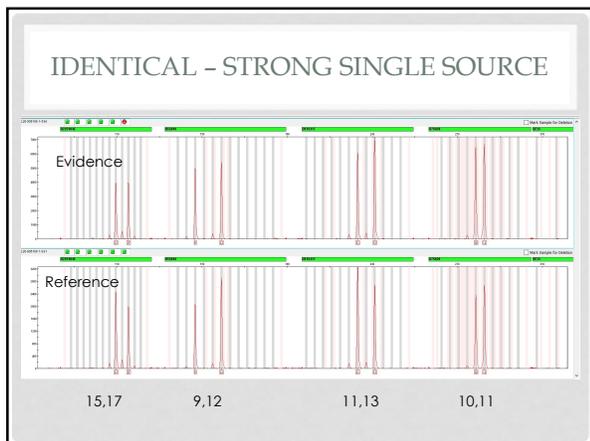
SUZETTE SANDERS, SUPERVISING FORENSIC SCIENTIST / TECHNICAL LEADER
VERMONT COURTESI SHERIFFS OFFICE FORENSIC SERVICES BUREAU
FORENSIC BIOLOGY SECTION

OVERVIEW

- DNA Evidence Examination
- STRmix DNA Profile Interpretation
- Likelihood Ratios
 - Analogy
 - Verbal Scales
 - Conservative LR
 - Common misrepresentation of Likelihood Ratios
 - Possible ways to rephrase the Likelihood Ratio
- Pretrial Conferences
- Questions/Answers

DNA EVIDENCE EXAMINATION

- Locate biological material for testing
 - Examples, clothing, body fluid testing, touch DNA swab
- Extract DNA
- Quantify DNA
 - How much DNA?
 - Is male DNA present?
 - Male : female ratio of DNA
 - Degraded?
 - Inhibited?
- Amplify DNA
- Visualize DNA types by analysis on an instrument





WHAT IS STRMIX?

Software tool used to assist the Forensic Scientist with the interpretation of DNA profiles and the calculation of statistics

- 1) Attempts to deconvolute (untangle) mixed DNA profiles from more than one individual into their individual components (or contributors/donors)
- 2) Calculates *likelihood ratio statistics* which answers the question how many times more likely is it to observe the evidence DNA profile given two different competing scenarios



STRMIX DECONVOLUTION

- STRmix uses the Markov chain Monte Carlo (MCMC), an algorithm that assigns a likelihood for each DNA type combination
- MCMC is used in many other scientific fields including:
 - Physics: Predicting behavior of radiation particles
 - Genetics: Creating phylogenetic trees
 - Engineering: Predicting behavior of buildings in earthquakes
 - Cryptography: Code breaking

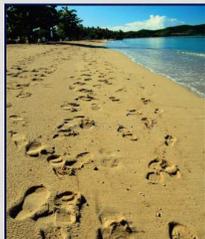
ANALOGY FOR MCMC

- Imagine yourself on a beach with a metal detector and your mission is to find the buried treasure...but where do you begin?



ANALOGY FOR MCMC

- We randomly pick a direction to start walking and use the frequency and speed of the beeps to tell us whether we are going in a "good" direction or a "bad" direction.
- The slower the beeps the further we are from finding the treasure
- As the beeps speed up, we attempt to continue walking in the direction that increases the frequency



ANALOGY FOR MCMC

- Sometimes we may take a few steps and the beeps slow down, rather than speed up
- We turn around or take small steps from side-to-side to re-orient ourselves onto a "better" path
- We ultimately wander in the direction of the beeps of the highest frequency, and then we dig!



METAL DETECTOR VS. MCMC

- The MCMC process is very similar to this analogy of using a metal detector on a beach:
- Allows us to randomly sample a relatively large probability space (e.g., beach)
- Uses a comparison method to identify "good" guesses from "not-so-good" guesses (e.g., beeping)
- More quickly homes in on our area of interest and finds the best DNA combinations that explains our observed DNA evidence profile (e.g., buried treasure)

ALTERNATIVE MCMC ANALOGY

- Similarly, another analogy that's commonly used is the childhood game of hot and cold to find a hidden item.
- Child moves in the direction of the "hot" guesses and away from "cold" guesses until the child ends up at the location of the hidden item.

GENOTYPE COMBINATION WEIGHTS

- When we first start looking for the treasure, we will be making many changes in direction and taking larger steps trying to find a good signal, but once we get closer to the treasure, we will be making many smaller moves and far more of them will be "hot" guesses.
- Similarly, once STRmix enters the sampling space where the software is narrowing in on which DNA type combinations best explain the evidence, it starts tracking how often a particular DNA type combination is accepted.
- Then the probability of those proposed DNA type combinations are expressed as a normalized weight. That is all the weights of the DNA type combinations that are proposed add up to 100%.

STRMIX - DNA PROFILE INTERPRETATION

Single Source

LOCUS	CONTRIBUTORS	WEIGHT
	1 (100%)	(HIGHLIGHT ≥ 0.99)
D3S1358	14, 16	1

Mixture

LOCUS	CONTRIBUTORS		WEIGHT
	1 (64%)	2 (36%)	
D3S1358	16, 17	16, 17	9.3861E-1
	16, 16	17, 17	5.6572E-2
	17, 17	16, 16	3.7574E-3
	16, 17	16, 16	9.9350E-4
	16, 17	17, 17	6.6233E-5
	16, 16	16, 17	1.1404E-5

STRMIX - LIKELIHOOD RATIOS

- After STRmix is used to interpret possible DNA profiles for the individual contributors, STRmix can be used to compare a reference sample to those interpretations and generate a statistic called a Likelihood Ratio (LR).
- Note.....

TESTIFYING TO LIKELIHOOD RATIOS

- Likelihood ratios represent the probability of observing the DNA evidence profile, specifically when comparing two different explanations of how the DNA profile may have been created.
- Answers the question:
"How likely is it to observe this piece of DNA evidence if Explanation 1 is true vs. if Explanation 2 is true?"

TEETER TOTTER ANALOGY

- This comparison concept can be understood by picturing a teeter totter (or seesaw) on a children's playground



TEETER TOTTER ANALOGY

- If a woman who weighs 130 pounds sits on one side of the teeter totter while a 200 pound man sits on the other side, it tips towards him.
- In other words, the teeter totter favors him, because compared to her, he carries more weight

TEETER TOTTER ANALOGY

- Similarly, likelihood ratios (LRs) place the probability of observing the DNA evidence profile if Explanation 1 is true on one side of the teeter totter and the probability of observing this same DNA evidence profile if Explanation 2 is true on the other side.



TEETER TOTTER ANALOGY

- The "weight" of the probabilities are determined by how well (or not well) the particular explanation fits the data observed in the DNA evidence profile
- Explanations that fit the data well have high probabilities and will weigh more
- Poor explanations for the observed DNA data have low probabilities and therefore weigh less
- The teeter totter tips towards the explanation with the higher probability, meaning the explanation is a better fit for the DNA evidence...
- We are more likely to observe this DNA evidence profile using this explanation as opposed to the other.

TEETER TOTTER ANALOGY

- How far and how fast the teeter totter tips represents the overall weight of the comparison
- The higher the LR result, the more support for that explanation of the DNA evidence
- When the LR is very high, it indicates that a random coincidental association is an unlikely explanation for the DNA evidence

TESTIFYING TO LIKELIHOOD RATIOS

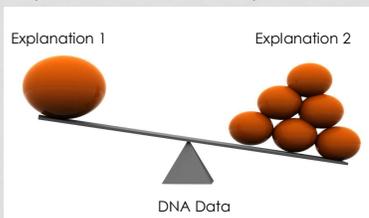
- It is important to understand that even though the teeter totter hinges upon the DNA evidence profile's data, the result ultimately depends only on the two specific explanations being compared
- If you change one or more of the explanations being compared, you will get a different likelihood ratio result
- This is why more than one likelihood ratio can be calculated for the same DNA evidence profile

TESTIFYING TO LIKELIHOOD RATIOS

- Going back to the playground example...
- If the woman in the last example, trades places with an elephant, the teeter totter will suddenly tip away from the 200 pound man.
- The man carried more weight when compared to the woman, but does not carry more weight when compared to the elephant.
- This results in a different LR calculation than previously performed for the same evidence DNA profile

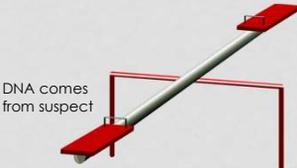
CONTINUING LIKELIHOOD RATIO TESTIMONY

- As the testimony progresses, use this strategy for explaining the different LR results by relating back to who is sitting on each side of the teeter totter based on the specific LR calculations performed.

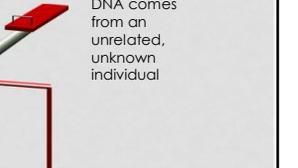


REPORT WORDING

- Obtaining this profile is at least 10 million times more likely if the DNA came from the suspect than if the DNA came from an unrelated, unknown individual.



DNA comes from suspect



DNA comes from an unrelated, unknown individual

EXPLAINING INDIVIDUAL LR

- Person of interest (POI) is on one side and a randomly selected unknown unrelated person is on the other
- Inclusions:** Teeter totter tips towards the POI
 - It is more likely to observe the DNA evidence if the POI is a contributor vs. a random unknown person
- Exclusions:** Teeter totter tips away from the POI and towards a random person
- Uninformative:** Teeter totter is balanced and is not tipping towards either explanation
 - Neither explanation is favored, so we cannot tell one way or the other

EXPLAINING SMALL LR

- A common criticism of LR is that they are difficult to explain without the use of a verbal scale
- Analogy can be used to convey the degree of support for very small likelihood ratios vs. very large ones insofar as how far and how fast the teeter totter tips
- Small LR will only tip ever-so-slightly in the direction of the explanation which supports the observed data
- As probabilities increase, the teeter totter will tip further and much more quickly in the favored explanation's direction



VERBAL SCALE

- A verbal scale is used in conjunction with the numerical LR value to convey in words the level of support that the LR provides for the explanation it favors.
- Obtaining this mixture is at least 10 million times more likely if the DNA came from the suspect and an unrelated, unknown individual than if the DNA came from two unrelated, unknown individuals.
- Verbal scale: Provides very strong support for inclusion of the suspect in the mixture.

VERBAL SCALE - SWGDAM

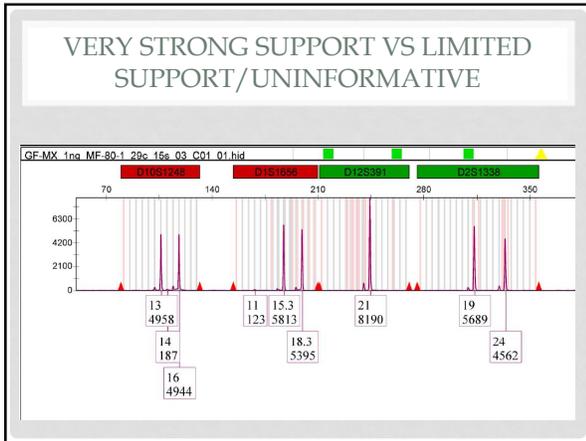
- The Scientific Working Group on DNA Analysis Methods (SWGAM) is currently comprised of dedicated forensic scientists, from international, federal, state and local forensic DNA laboratories as well as guests representing academia and other Federal agencies. These forensic scientists serve as the DNA technical leaders or Combined DNA Index System (CODIS) Administrators for their laboratories, and are able to offer the perspectives of practitioners in the areas of nuclear and mitochondrial DNA technologies. SWGDAM is also fortunate to have invited guests attend each meeting representing academia, other Federal laboratories and international agencies to provide their specific expertise in areas such as mitochondrial DNA, population genetics, statistics, and YSTRs.

The responsibilities of SWGDAM are: (1) to recommend revisions, as necessary, to the Quality Assurance Standards for Forensic DNA Testing Laboratories and the Quality Assurance Standards for DNA Databasing Laboratories; (2) to serve as a forum to discuss, share, and evaluate forensic biology methods, protocols, training, and research to enhance forensic biology services; and (3) to recommend and conduct research to develop and/or validate forensic biology methods.

- Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios, July 12, 2018.

VERBAL SCALE

1 Million Plus	Very Strong Support for Inclusion
10,000 to 999,999	Strong Support for Inclusion
100 to 9,999	Moderate Support for Inclusion
2 to 99	Limited Support for Inclusion
Approximately 1 (between 0.5 and 2)	Uninformative
Between 0.1 and 0.5	Limited Support for Exclusion
0 to 0.1	Exclusion



VERBAL SCALE

<p>Advantages</p> <ul style="list-style-type: none"> • Gives perspective on the range of values typically seen • Example, 5 units tall • Provides a consistent way to describe the LR support. 	<p>Disadvantages</p> <ul style="list-style-type: none"> • Somewhat arbitrary • An LR of 100 is not much different than an LR of 99, but one is considered moderate support and the other limited support. • An LR of 100 and an LR of 9,000 are quite a bit different, but they are both considered moderate support
--	--

VERBAL SCALE AND NUMERICAL LR

- The best way to understand the weight of an LR is to consider both the verbal scale and the numerical value.
- The verbal scale can put the LR into general perspective (is the LR limited, moderate, strong, very strong?)
- The numerical LR can pinpoint more specifically where the LR is within that verbal scale category. (is the LR near the low end of that category or nearer the high end of that category?)

WHAT IS TRUE?

- An LR of 100 (moderate support) is only 1 away from an LR of 99 (limited support). There is not much difference between the LR of 100 and LR of 99 even though it has crossed over this arbitrary verbal scale line.
- An LR of 1 million (1,000,000) and 1 quintillion are both in the verbal scale with the strongest support (very strong support) yet an LR of 1 quintillion (1,000,000,000,000,000,000) is 1 trillion (1,000,000,000,000) times more than 1 million. The numbers reflect that there is a big difference between the two numbers.

CONSERVATIVE LR

- A low estimate of the LR is the reported LR
- Report the lowest HPD LR of four populations (African-American, Asian, Caucasian, and Hispanic)
- Highest Posterior Density (HPD) LR is a lower LR than the original calculated LR, taking into account uncertainty in the LR calculation.
- This is why our reports state "at least" for inclusionary LRs.

COMMUNICATING LRS

- Transposing the conditional – an error that can occur in communicating the likelihood ratio if it is not worded correctly

Occurs when the wording changes from the probability of observing the DNA evidence given two opposing explanations for who has donated the DNA,

To

The probability that it was donated by that person given the DNA evidence,

TRANSPOSING THE CONDITIONAL

Probability you have 4 legs



IF

you're an elephant



Probability you're an elephant

IF

you have 4 legs



≠



INCORRECT WAYS TO RESTATE THE LR

<p>Correct</p> <ul style="list-style-type: none"> • Obtaining this profile is at least 10 million times more likely if the DNA came from the suspect than if the DNA came from an unrelated, unknown individual. • Replace "Obtaining" with "Observing" • Replace "profile" with "DNA profile" or "evidence profile" or "DNA evidence" • Replace "individual" with "person" 	<p>Incorrect</p> <ul style="list-style-type: none"> • It is at least 10 million times more likely that the suspect is the DNA donor. • It is at least 10 million times more likely that the suspect is the DNA donor than someone else.
--	--

INCORRECT WAYS TO RESTATE THE LR

<p>Correct</p> <ul style="list-style-type: none"> • Obtaining this profile is at least 10 million times more likely if the DNA came from the suspect than if the DNA came from an unrelated, unknown individual. 	<p>Incorrect</p> <ul style="list-style-type: none"> • The DNA profile is at least 10 million times more likely to be from the suspect than from an unrelated, unknown individual. • The probability that the suspect is the contributor of this DNA profile is at least 10 million times higher than the probability that an unrelated, unknown individual is the contributor. • It is at least 10 million times more likely that the suspect is the contributor of this DNA profile than it is that an unrelated, unknown individual is the contributor.
--	---

POSSIBLE WAYS TO REPHRASE LR

Original

- Obtaining this mixture is at least 10 million times more likely if the DNA came from the suspect and an unrelated, unknown individual than if the DNA came from two unrelated, unknown individuals.

Correct Rephrasing

- This DNA profile is at least 10 million times more likely to be observed in a mixture of the suspect and an unrelated, unknown individual than in a mixture of two unrelated, unknown individuals.

POSSIBLE WAYS TO REPHRASE THE LR

Original

- Obtaining this mixture is at least 10 million times more likely if the DNA came from the suspect and an unrelated, unknown individual than if the DNA came from two unrelated, unknown individuals.

Correct Rephrasing

- This DNA profile is at least 10 million times better explained as/by a mixture of the suspect and an unrelated, unknown individual than as/by a mixture of two unrelated, unknown individuals.
- Other options "easier to explain if" or "better explained if"

POSSIBLE WAYS TO REPHRASE LR

Original

- Obtaining this mixture is at least 10 million times more likely if the DNA came from the suspect and an unrelated, unknown individual than if the DNA came from two unrelated, unknown individuals.

Correct Rephrasing

- A mixture of the suspect and an unknown, unrelated person fits the evidence at least 10 million times better than a mixture of two unknown, unrelated people.

IMPORTANCE OF PRE-TRIAL CONFERENCE

- Talk with the scientist ahead of time
- Talk about analogies that you want to use
- Talk about how you plan to re-phrase the LR
- If you plan to point out that the LR is X times more than the 1 million starting point, discuss it with them.
- Other questions – Is it identical, etc.

QUESTIONS?

- Next up.....Conditioning on Assumed Contributors
