# Introduction to Familial Searching

## Kristen Lewis O'Connor, Ph.D.

Senior Consultant, Booz Allen Hamilton

(Former Postdoctoral Fellow at NIST)

Promega ISHI

Washington, DC

October 3, 2011

# Outline

- Definition of familial searching

- Case example

- World wide perspective

- Validation points to consider

# What is a familial search?

- An additional search of a law enforcement DNA database conducted after a routine search has been performed and no profile matches are identified

- Deliberate search of a DNA database to potentially identify close biological relatives (parent, child, full sibling) of an unknown forensic profile obtained from crime scene evidence

- Typically used in no suspect cases, cold cases, violent crimes to develop **investigative leads**
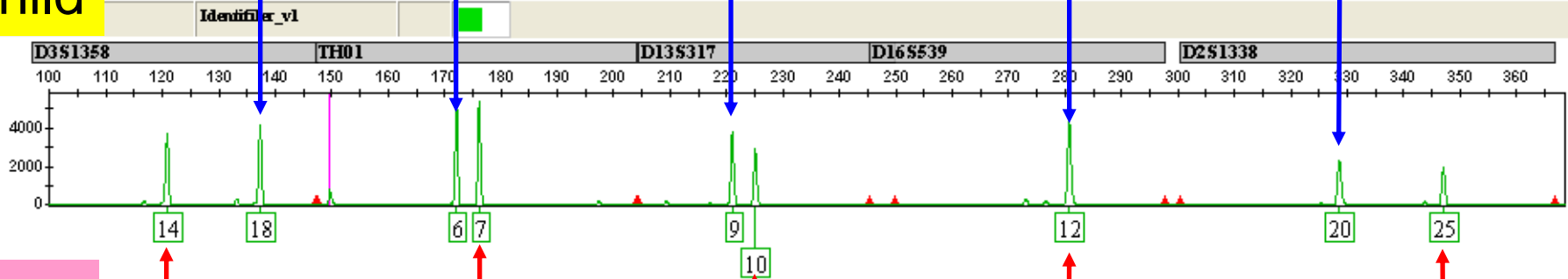
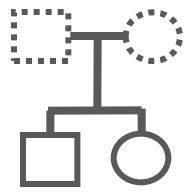# Fundamentals of Paternity Testing
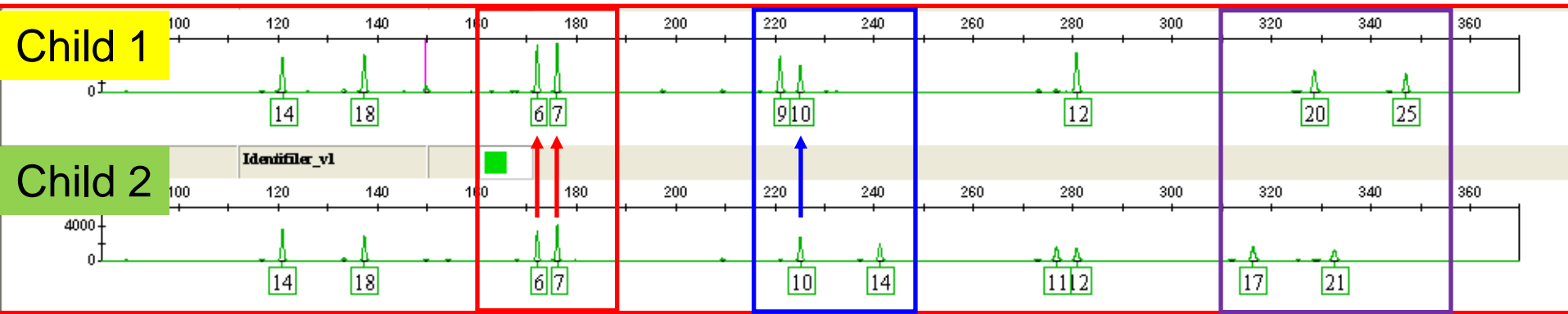
Focus on 5 markers…
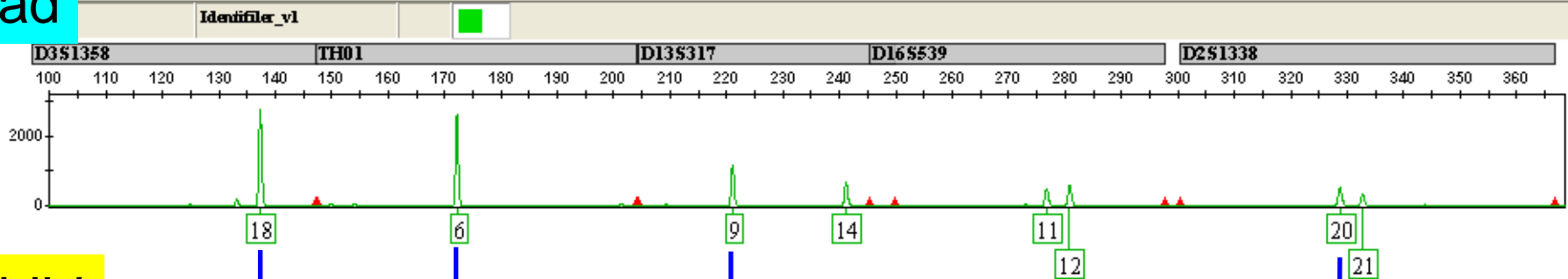


Parent-offspring will share **one** allele at every locus

# Kinship Analysis: Full Siblings

Focusing on 5 markers…



Full siblings may share **two**, **one**, or **zero** alleles at a locus

↓

**For more distant familial relationships, allele sharing decreases**

↓

**Uncertainty Increases**

# Basis of Searching for Relatives

- First-order relatives will share more genetic data than unrelated individuals

- Child inherits half of his DNA from each parent

- Siblings may share:
  - no alleles in common with 25% chance
  - one allele in common with 50% chance
  - two alleles in common with 25% chance

# How is kinship assessed?

## Likelihood Ratio (LR)

Describes how strongly the genotypes support one relationship versus the other relationship

Expresses the likelihood of obtaining the DNA profiles under two mutually exclusive hypotheses

$$LR = \frac{\text{Probability of genotypes if individuals are related as claimed}}{\text{Probability of genotypes if individuals are unrelated}}$$

The LR takes into account:
- the **probability of allele sharing** for individuals with a specific relationship
- the **allele frequency** of alleles
- a possible mutation event (if necessary)

# Likelihood Ratio (LR)

The LR is also called the relationship index (RI) or kinship index (KI).

Each independent locus tested produces its own relationship index, which can be multiplied by those of other independent loci to calculate a combined relationship index (CRI).

$$\textbf{CRI} = \frac{\text{Probability of genotypes if A,B are full siblings}}{\text{Probability of genotypes if A,B are unrelated}}$$
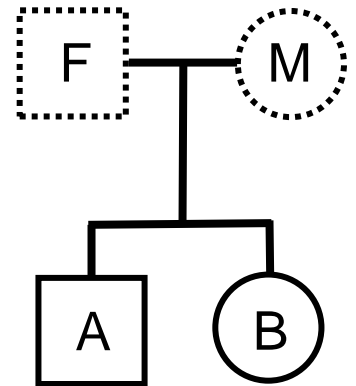


By the definition of a LR:
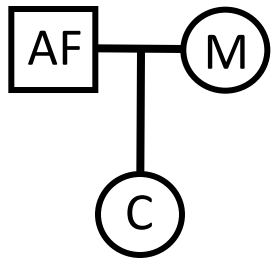CRI > 1 supports the numerator (claimed relationship)
CRI < 1 supports the denominator (alternative relationship)

Larger CRI values provide more support for the claimed relationship

# Likelihood Ratio (LR)

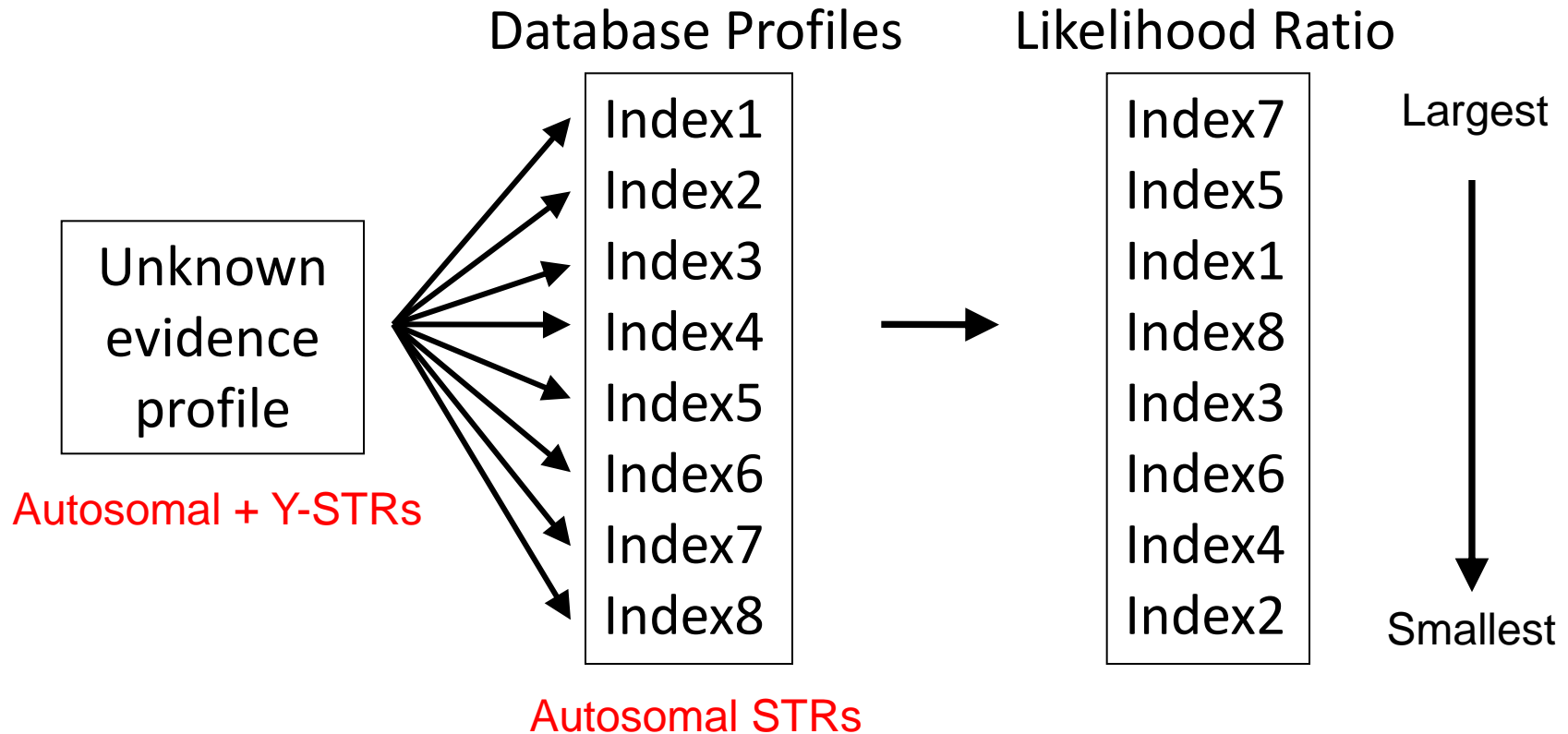**Hypothesis 1 =** Paternity Trio, **Hypothesis 2 =** Unrelated

Paternity trio



**LR = 168,468,800**

| Locus | Probability (Hypothesis 1) | Probability (Hypothesis 2) | Likelihood Ratio |
|---|---|---|---|
| D8S1179 | 0.001545163 | 0.000574194 | 2.691012 |
| D21S11 | 0.0003079 | 0.000171693 | 1.793322 |
| D7S820 | 0.00078148 | 0.000138664 | 5.635774 |
| CSF1PO | 0.003673636 | 0.000798261 | 4.602047 |
| D3S1358 | 0.002522579 | 0.001086988 | 2.320706 |
| TH01 | 0.001420379 | 0.00032926 | 4.313852 |
| D13S317 | 0.000454644 | 4.37E-05 | 10.39317 |
| D16S539 | 9.47E-05 | 2.80E-05 | 3.38817 |
| D2S1338 | 4.87E-05 | 1.15E-05 | 4.250356 |
| D19S433 | 0.004076747 | 0.000661891 | 6.159245 |
| VWA | 0.000131184 | 5.26E-05 | 2.492709 |
| TPOX | 0.008606737 | 0.005087928 | 1.691599 |
| D18S51 | 0.000328927 | 9.07E-05 | 3.625514 |
| D5S818 | 0.002742154 | 0.000772507 | 3.549682 |
| FGA | 0.000532767 | 0.000198233 | 2.687581 |
| **Total** | **2.27E-47** | **1.35E-55** | **168,468,800** |

It is **168 million times** more likely that we observe these DNA profiles if the Alleged Father is the true father than if an unrelated man is the father of the child.

# Familial Search Schematic

Database Profiles       Likelihood Ratio

| | |
|---|---|
| **Unknown evidence profile** | Index1<br>Index2<br>Index3<br>Index4<br>Index5<br>Index6<br>Index7<br>Index8 |

Largest

| |
|---|
| Index7<br>Index5<br>Index1<br>Index8<br>Index3<br>Index6<br>Index4<br>Index2 |

Smallest

Autosomal + Y-STRs

Autosomal STRs

Based on your validation studies, select <u>X</u> profiles with highest LR values for subsequent genotyping (Y-STRs or mtDNA) to filter out unrelated individuals

# Challenge of Identifying True Relatives in a Database

| Rank | Index # | Likelihood Ratio (LR) | |
|------|---------|-----------------------|---|
| 1 | Index5243 | 7048 | — Unrelated |
| 2 | Index1438 | 5503 | — Unrelated |
| **3** | **Index48998** | **45** | — True relative |
| 4 | Index45677 | 3 | — Unrelated |
| 5 | Index39732 | 0 | |
| 6 | Index134 | 0 | |
| 7 | Index7701 | 0 | |
| . | . | . | |
| . | . | . | |
| 412 | Index22093 | 0 | |
| 413 | Index208 | 0 | |

- Unrelated individuals may have higher LRs due to chance allele sharing
  - Included in subsequent investigation → "False positive"
  - More false positives as database size increases
- True relatives will not always have the highest LR
  - Potentially not included in subsequent investigation →"False negative"

# British Serial Killer: "Saturday Night Strangler"

- In 1973, three 16-year olds were raped, strangled, and left in the wooded area near a busy road in South Wales

- Unsolved for nearly 30 years… then low-copy number testing developed suspect profile

- 500 nominal suspects were investigated
    - 353 persons of interest were typed but no matches to perpetrator
    - Suspect #200 was Joe Kappen (no DNA sample since he died 12 years earlier)

- In 2001, Dr. Jonathan Whitaker (FSS) manually searched the national database looking for an allele sharing pattern typical of parent-child
    - Identified Paul Kappen who was 7 years old at the time of the murders

- Deduced Joe Kappen's profile from those of ex-wife, daughter, and son

- Exhumed Joe Kappen's body and obtained DNA from teeth/bone that matched the perpetrator's profile

Pauline Floyd and Geraldine Hughes

Sandra Newton

Joe Kappen

# Familial Searching Policies

- United Kingdom

- New Zealand

- California

- Colorado

- Virginia

Visit this website to view each laboratory's policies:

http://www.denverda.org/dna/Familial_DNA_Database_Searches.htm

# Worldwide Perspective



**http://www.councilforresponsiblegenetics.org/dnadata/**

# National DNA Databases



Operational DNA Database
Planned DNA Database

Download Full Report
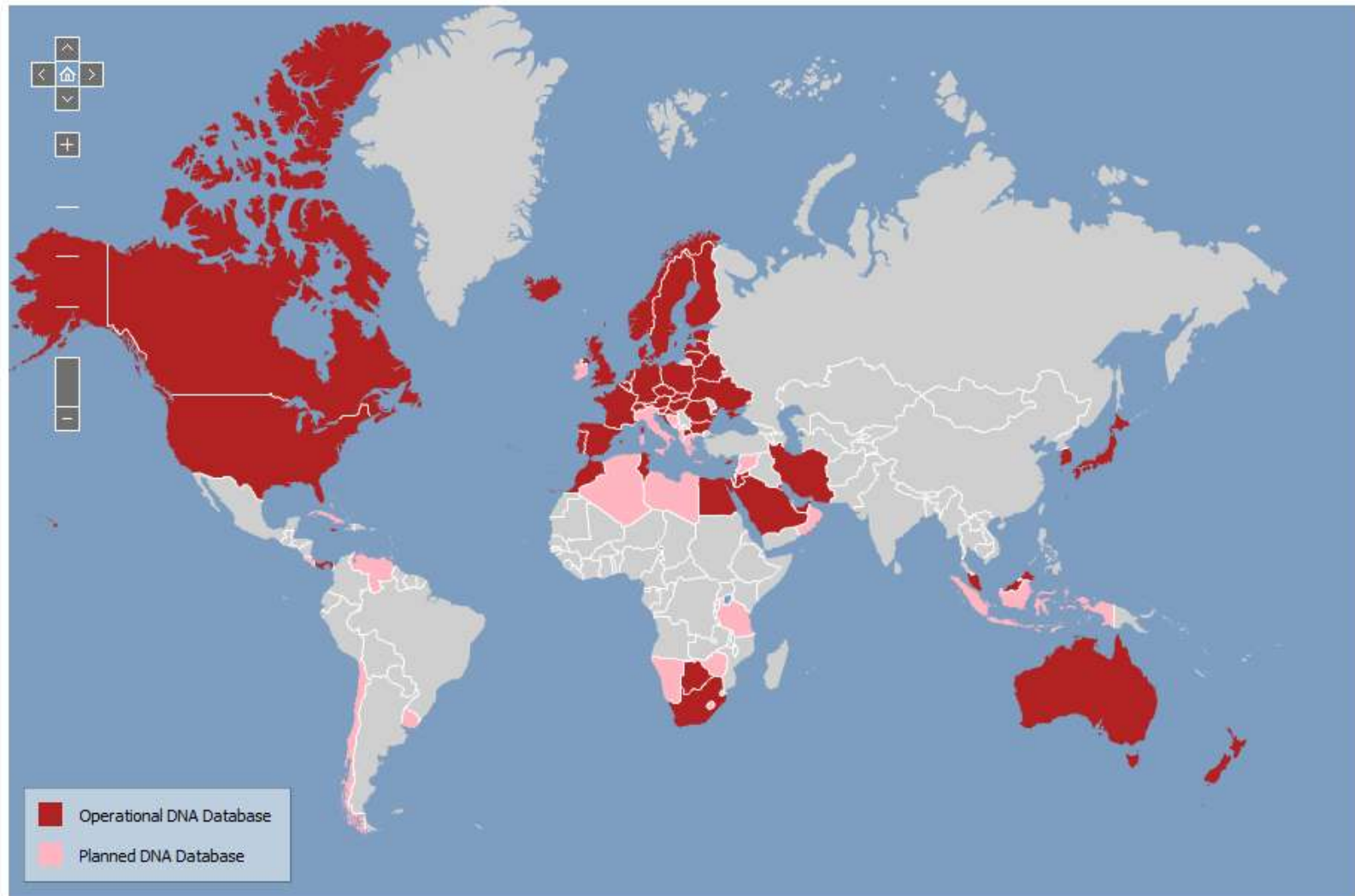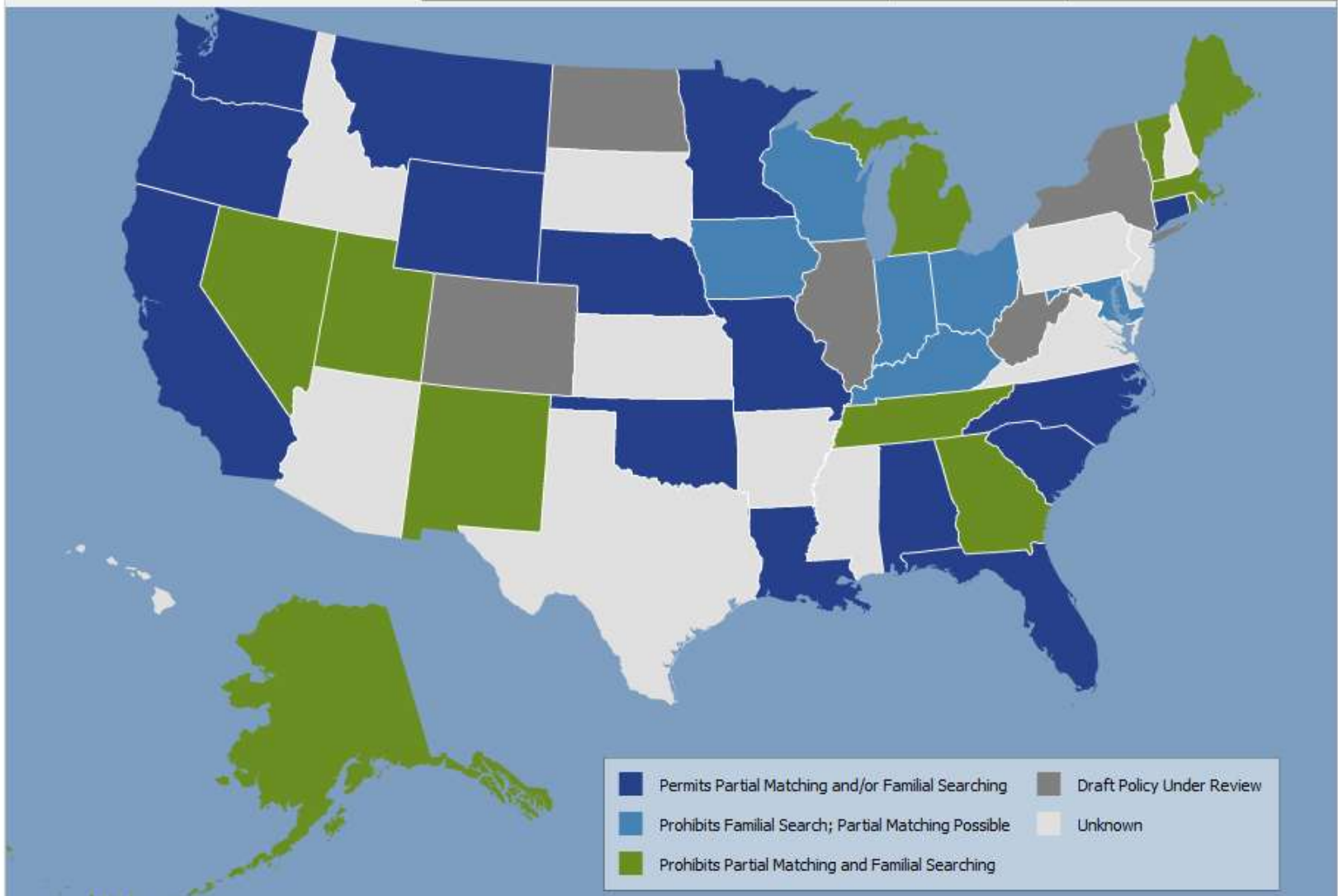
(2011)

# U.K. Perspective on Familial Searching

- Statistics from 2004 to January 2011
  - 179 cases submitted; 36 successes/81 cases completed (44.4% success rate)

- Metadata (age, locality, ethnicity) contained in database record
  - Increase success of familial searching by filtering the ranked list on non-genetic information

Statistics from: Chris Maguire (formerly at FSS)

# United States



**State Rules on Partial/Familial Searching** | **States Collecting DNA Samples from Arrestees** | **National Report**

Legend:
- Permits Partial Matching and/or Familial Searching
- Prohibits Familial Search; Partial Matching Possible
- Prohibits Partial Matching and Familial Searching
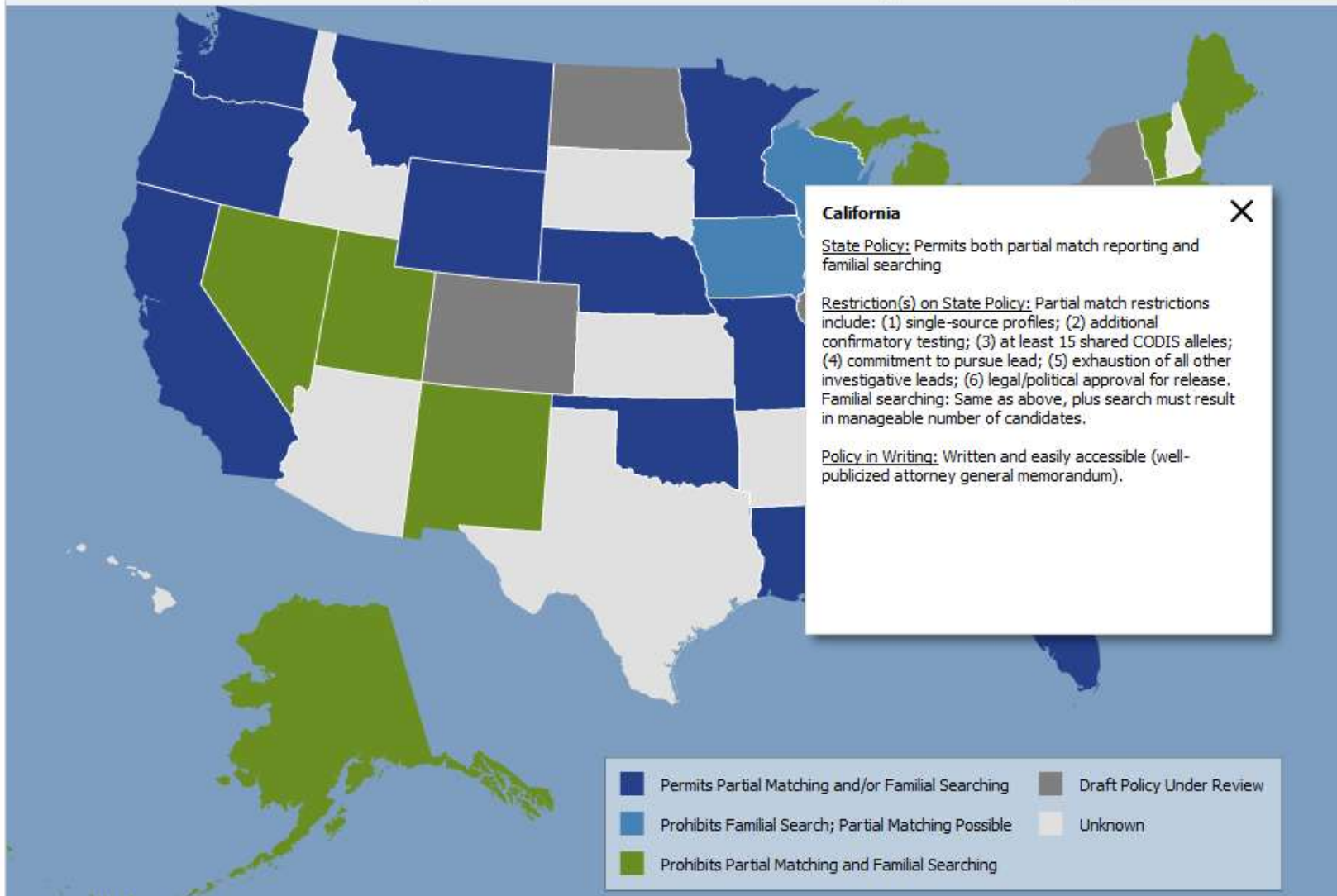- Draft Policy Under Review
- Unknown

# United States

# Virginia

# Texas



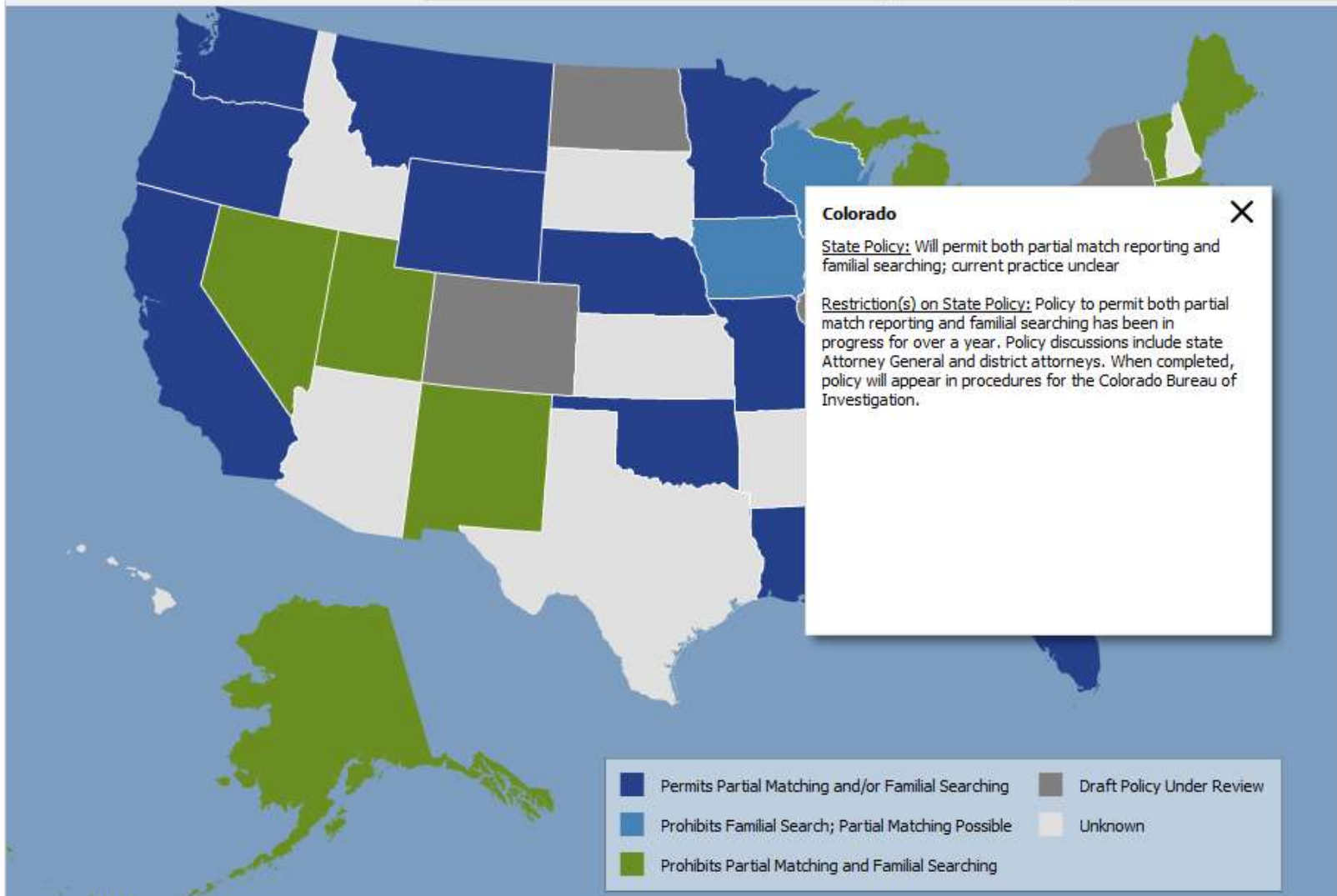State Rules on Partial/Familial Searching | States Collecting DNA Samples from Arrestees | National Report

**Texas**          ×

State Policy: unknown

Restriction(s) on State Policy: n/a

Performs familial
searches but details
are not available

Permits Partial Matching and/or Familial Searching — Draft Policy Under Review

Prohibits Familial Search; Partial Matching Possible — Unknown

Prohibits Partial Matching and Familial Searching
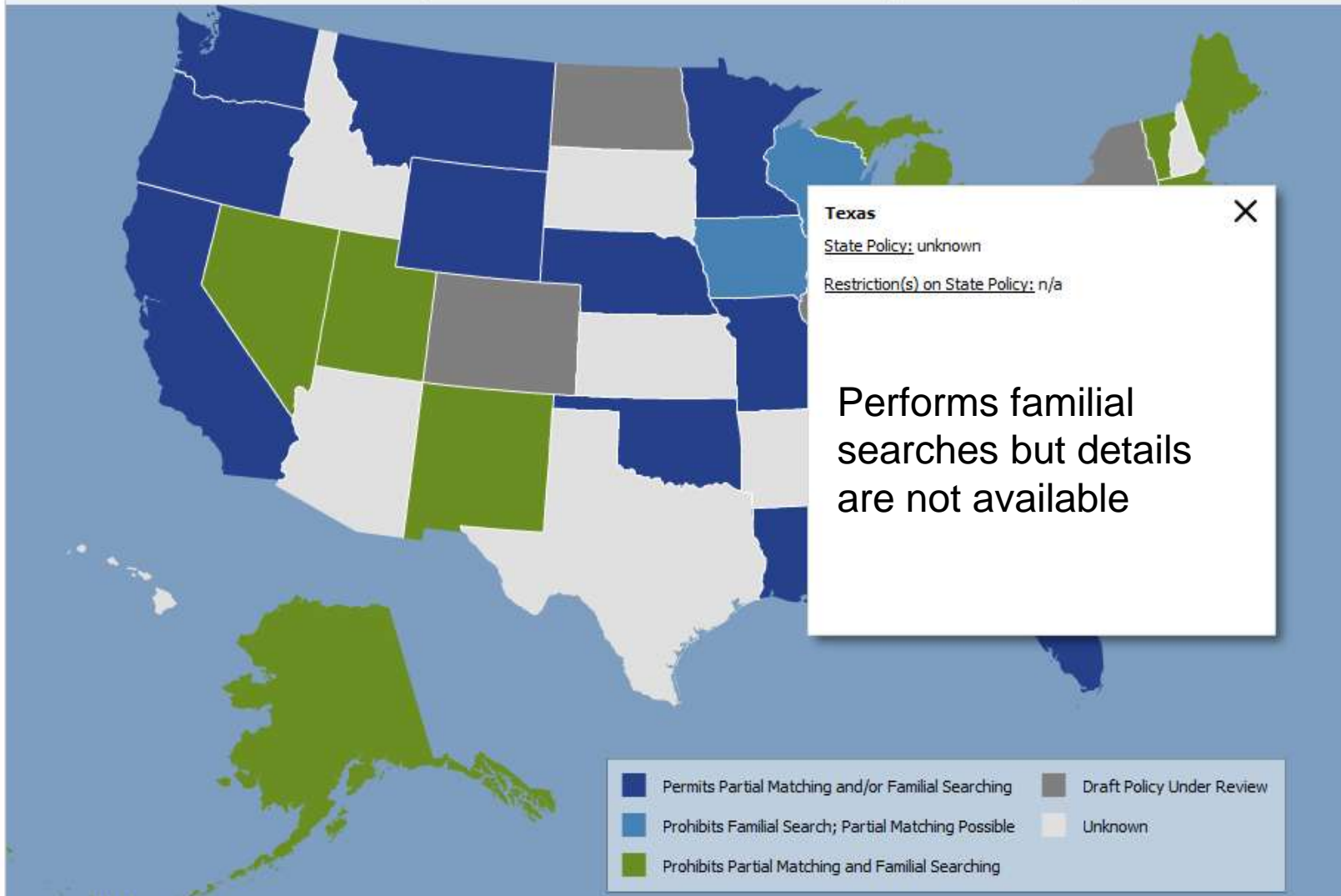
# U.S. Perspective on Familial Searching

- Colorado: all forensic unknowns, 10 identifications, 1 conviction (as of June 2011)

- California: 13 searches, 2 arrests (as of March 2011)

- Virginia: validation completed (March 2011)

- Texas

# Issues to Consider for Your Familial Searching Validation

# DNA Markers

- Evidence profile should have autosomal AND Y-STRs (or mtDNA) typed prior to search

- Set policy requirements:
  - Minimum number of autosomal STRs (less than 13 is not recommended)
  - Complete profiles (no missing alleles, no mixtures unless major profile can be deduced)

Myers et al., Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach. FSI Genetics 2011 Nov;5(5):493-500.

# Known Test Profiles

- Test families
  - Recommend 100 sets of genotypes with known relatedness
  - Father and two full siblings: select one sibling profile as "the evidence"
  - No need to add relatives' profiles to database – calculate pairwise LR
  - Autosomal STR profiles for 100 family quartets will be available on the NIST STRBase Kinship site

http://www.cstl.nist.gov/strbase/kinship.htm

Myers et al., Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach. FSI Genetics 2011 Nov;5(5):493-500.

# More Validation Considerations

- Mutation
  - If not accounting for mutations, a true parent-offspring relationship could be excluded (rare)
  - If accounting for possible mutations, the number of false positives will increase

- Population structure
  - Optional "theta correction" acts to decrease LR values
  - Need to define appropriate value

# What Allele Frequency Database to Use?

- Perform search and calculate LRs using separate population allele frequencies
- End up with multiple ranked lists separated by population
- Which LR value do you use?

Possible Options
1. Take minimum LR among the population lists
   - Presupposes that an individual's alleles tend to be more common in their own group than in other groups
2. Take highest LR values above define threshold, regardless of population
   - Remove any duplicate profiles that are present in more than one population list

# Incorporating Y-STR Information into Initial Database Search

If the evidence has been types for Y-STRs, you can factor the weight of a paternal relative match, assuming one exists and is identified through subsequent Y-STR testing of ranked database profiles.

$$\text{''Odds''} = LR_{\text{autosomal STR}} * LR_{\text{Y-STR}} * 1/N$$

Calculated for each evidence profile vs. database profile

Based on the frequency of the Y-STR profile of the evidence

Based on the database size **at time of search**

## California's Approach

"… incorporating the database size appear[s] sufficient to convey the value of the comparison as an investigative lead while tempering the magnitude of results obtainable when making a large number of comparisons to unrelated individuals."

Myers et al., Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach. FSI Genetics 2011 Nov;5(5):493-500.

# Statistical Thresholds

- Consider those in line with recommendations by the SWGDAM Ad Hoc Committee on Partial Matches

- For an offender to be investigated further, "odds" for either the parent-child relationship or the full-sibling relationship must be:
  - greater than or equal to 1 for at least one population *and*
  - no less than 0.1 for the remaining two populations

- At least one population group had attained a posterior probability of relatedness:
  - greater than or equal to 50% *and*
  - neither of the other two had posterior probabilities lower than 9.1%.

- **"Thresholds are appropriate given the intended purpose is to develop a lead for further investigation, not to directly identify an individual"**

<span style="color:red">**California's Approach**</span>

Myers et al., Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach. FSI Genetics 2011 Nov;5(5):493-500.

# Thresholds for Further Investigation

- Balance between false positives and false negatives
  - Increasing the LR threshold makes familial searching less efficient but reduces the number of false positive leads

- Balance between finding true relative (if one exists in database) and cost of additional genotyping and investigation
  - May be constrained by the number of samples to take forward for additional testing
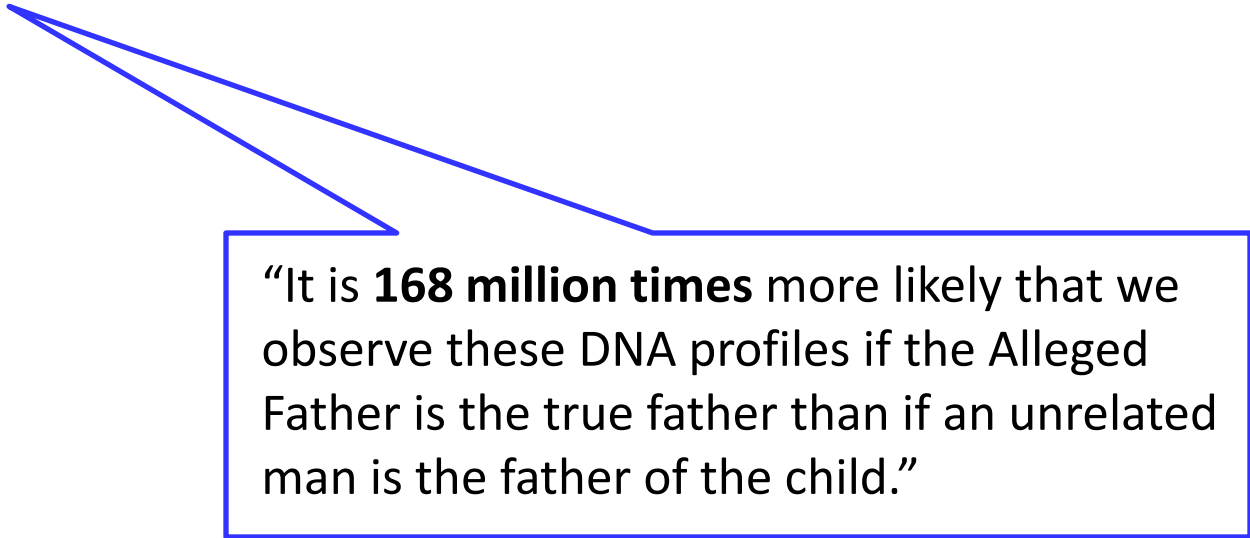
# General Comments

Understanding what your statistical results mean

# Statement Describing Strength of the "Familial Match"

## Likelihood Ratio (LR)

Expresses the probability of obtaining the DNA profiles under two mutually exclusive hypotheses

"It is **168 million times** more likely that we observe these DNA profiles if the Alleged Father is the true father than if an unrelated man is the father of the child."

# Estimating Relationships: Bayesian Approach

- What is the **probability of a relationship** given the observed genotypes?

- Often, this is what we want to know

- Different from likelihood ratios, where we calculate the **probabilities of the observed genotypes** given different hypothesized relationships

# Statement Describing Strength of the "Familial Match"

Combine the DNA information with the non-genetic information

$$\text{Posterior Odds} = \text{Likelihood Ratio} \times \text{Prior Odds}$$

$$\text{``Odds''} = LR_{\text{autosomal STR}} * LR_{\text{Y-STR}} * 1/N$$

Posterior odds provides a numerical weight to the opinion of the relationship. This is often expressed as a posterior probability:

*Probability of Relationship = PO/(PO+1) × 100*
*Probability of Relationship = (CRI × Pr /[CRI × Pr + (1-Pr)]) × 100*
where PO = Posterior Odds, Pr = Prior Probability, and CRI = Combined Relationship Index

"The probability that these two individuals are related as parent-offspring is 98%."

# Guidance on Privacy Concerns

- DOJ, Office of Justice Programs, Global Privacy and Information Quality Working Group

- Finalizing an issue paper, "An Introduction to Familial DNA Searching: Issues for Consideration" (anticipated publication by December 2011)

- Also working to develop a model privacy policy for states and justice agencies that are implementing familial searching capabilities

http://www.it.ojp.gov/biometricsprivacy

# Resources

- http://www.denverda.org/dna/Familial_DNA_Database_Searches.htm
- http://www.councilforresponsiblegenetics.org/dnadata/
- http://www.fbi.gov/about-us/lab/codis/familial-searching
- http://www.it.ojp.gov/biometricsprivacy
- Myers et al., Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach, FSI Genetics 2011 Nov;5(5):493-500.
- NIST STRBase Kinship site: http://www.cstl.nist.gov/strbase/kinship.htm

**Contact Info:** Kristen O'Connor
oconnor_kristen@bah.com