

Effect of Additional Loci on Likelihood Ratio Values for Complex Kinship Analysis



Kristen Lewis O'Connor, Ph.D.

National Institute of Standards and Technology

21st International Symposium on Human Identification

San Antonio, Texas

October 12, 2010





Final version of this presentation available at:
<http://www.cstl.nist.gov/strbase/NISTpub.htm>

Questions to Be Addressed

- How does kinship analysis relate to forensic DNA typing?
- Is there value in examining additional loci?
- What has NIST accomplished with kinship analysis?
- Where can one learn more about these topics?

What is our forensic
core competency?

Laying the foundation for a
discussion of kinship analysis

Forensic Core Competency

Direct match

1-to-1

**Standard
STR Typing**

13 core loci

Type of Search

**Standard STR
Typing for
Database Search**

13 core loci

1-to-many

High certainty

**Reference Profiles
(K_1 to K_n)**

Profile 1: 9,13 - 10,11 - ..

Profile 2: 8,11 - 10,12 - ..

Profile 3: 10,11 - 11,12 - ..

Profile 4: 15,16 - 13,13 - ..

Profile 5: 9,10 - 10,10 - ..

Profile 6: 8,14 - 10,13 - ..

Profile 7: 9,12 - 10,11 - ..

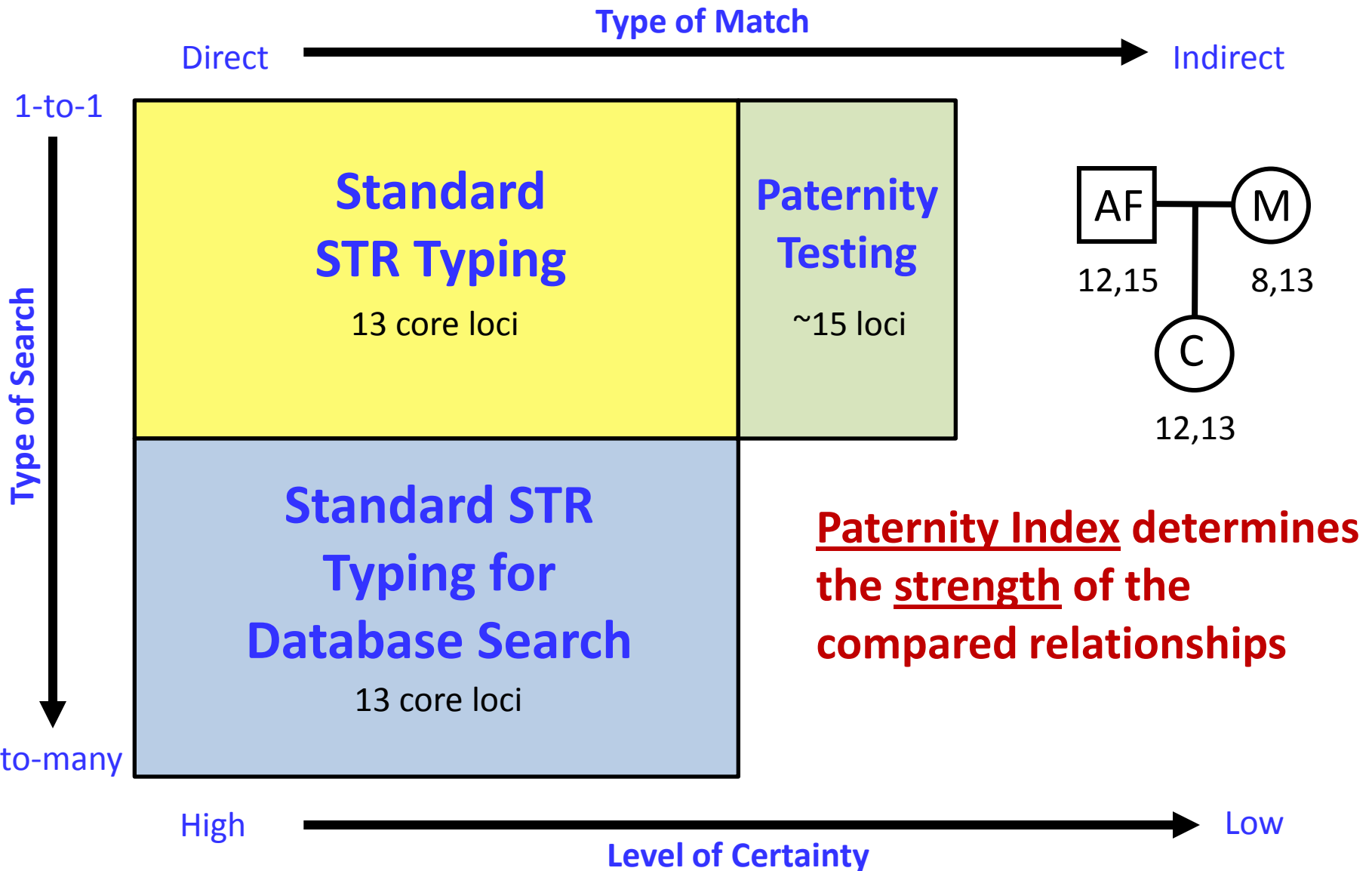
Profile 8: 7,15 - 10,12 - ..

Profile 9: 9,13 - 11,11 - ..

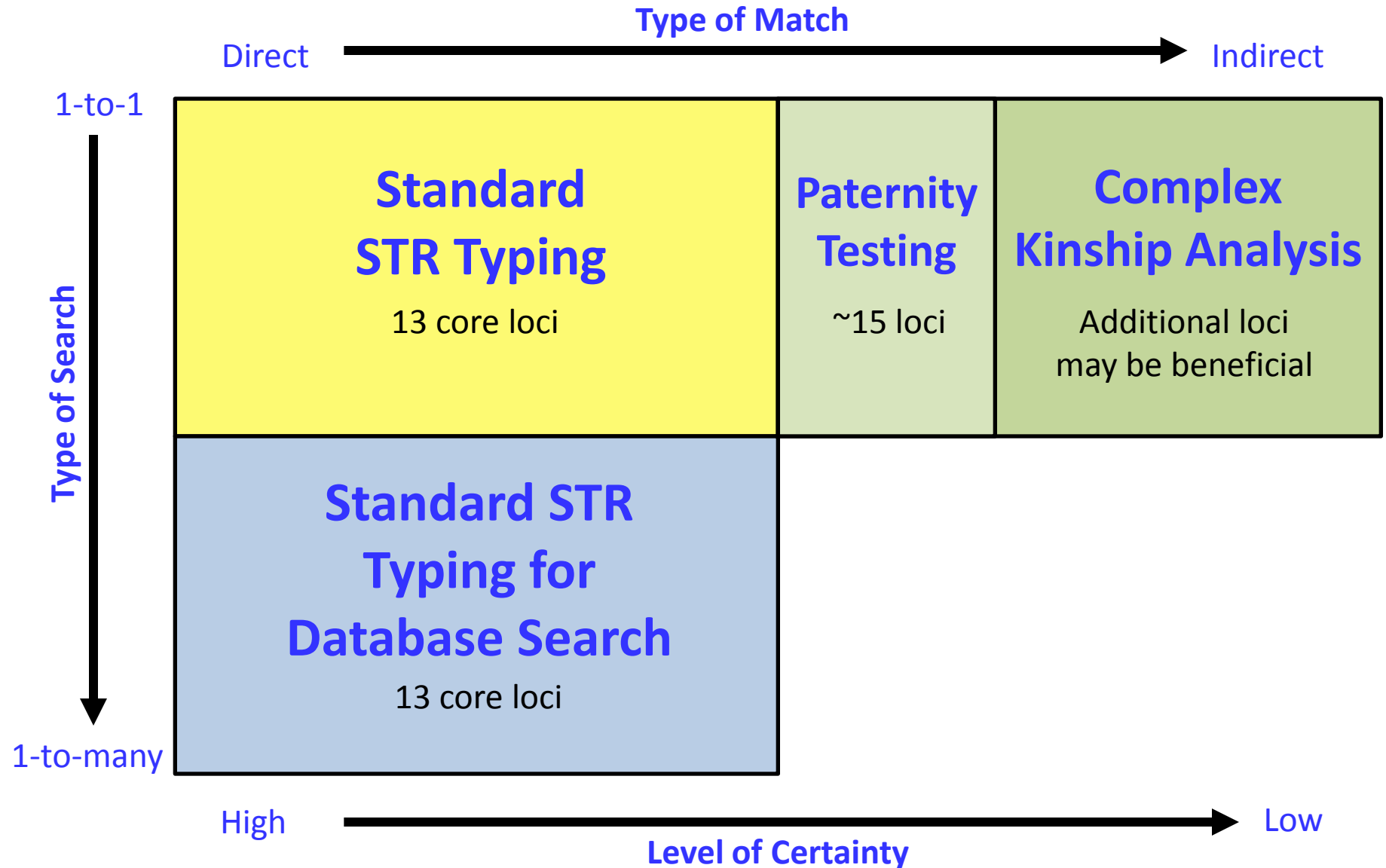
Evidence profile
8,14 - 10,13 - ...

**Random Match Probability
determines the rarity of the profile**

Expanding the Forensic Core Competency



Expanding the Forensic Core Competency



What is kinship analysis?

What is kinship analysis?

Evaluation of relatedness between individuals

Applications

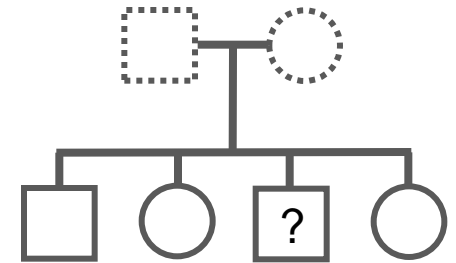
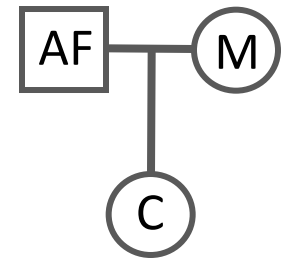
Parentage testing (civil or criminal)

Disaster victim identification

Missing persons identification

Familial searching

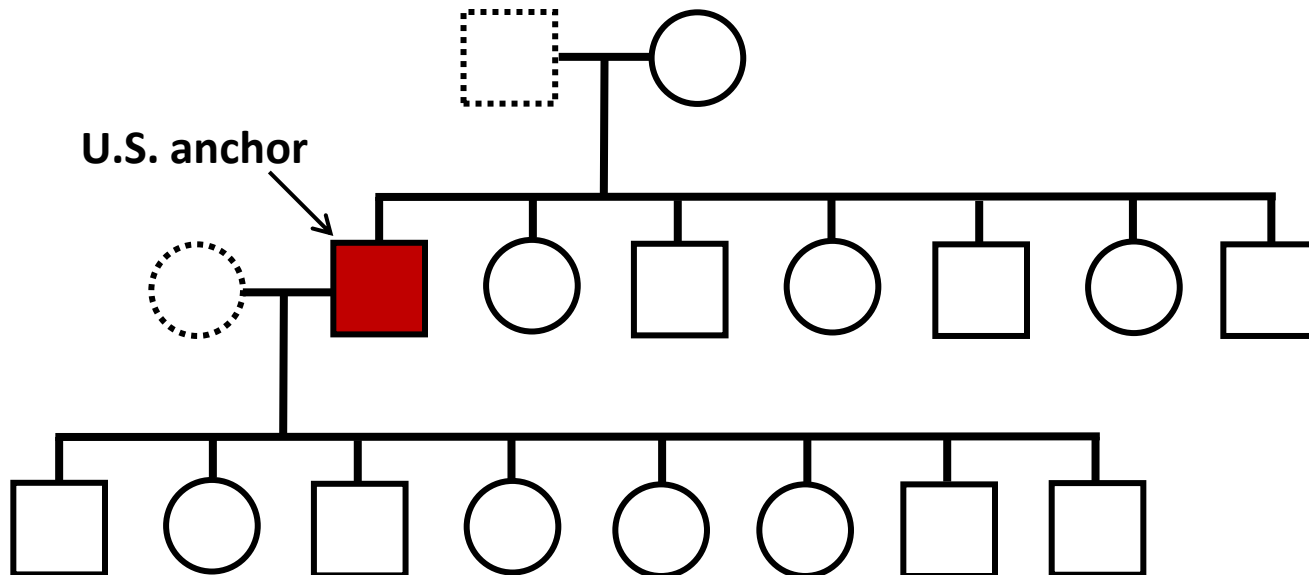
Immigration





Immigration Testing

U.S. Department of Homeland Security



Anchor may sponsor up to 15 relatives (spouse, parents, siblings, children)

79% of refugee claims were fraudulent based on DNA testing or failure to appear for DNA testing (U.S. Dept. of State)

DHS is looking to require DNA to support relationship claims

Why can kinship analysis
be complex?

Direct Matching

Exact match between compared genotypes

**Standard
STR Typing**

Evidence profile → **Suspect profile**
8,14 - 10,13 - ... 8,14 - 10,13 - ...

Direct Matching

Exact match between compared genotypes

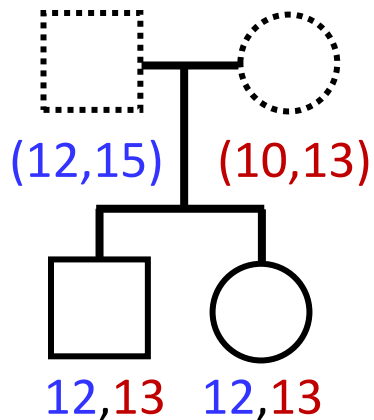
**Kinship
Analysis?**

Kinship profile 1 ↔ **Kinship profile 2**
8,14 - 10,13 - ... 8,14 - 10,13 - ...

Direct Matching

Exact match between compared genotypes

Kinship Analysis



Twins!

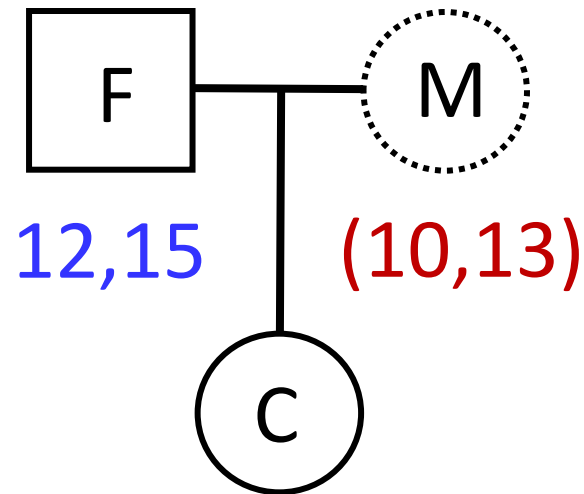
2 alleles shared every locus

Probability of Sharing Alleles from a Common Ancestor

Relationship	0 alleles	1 allele	2 alleles
Identical twin	0	0	1

Indirect Matching

Parent-Offspring



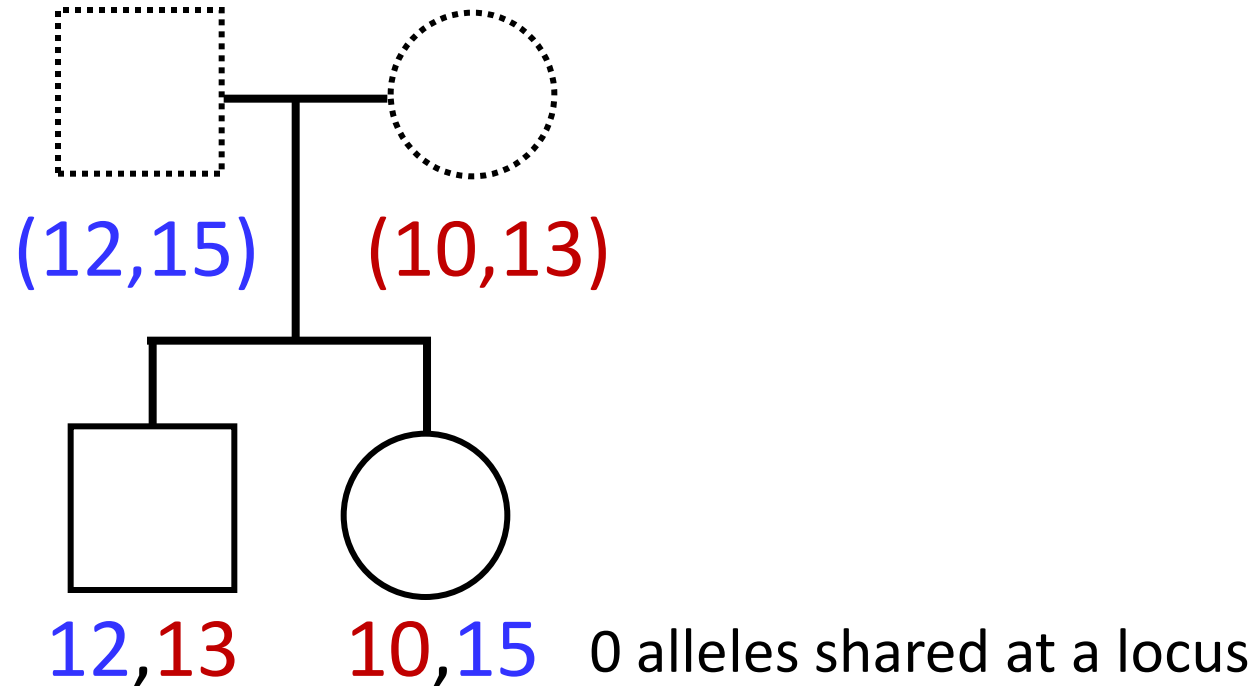
12,13 1 allele shared at every locus

Probability of Sharing Alleles from a Common Ancestor

Relationship	0 alleles	1 allele	2 alleles
Parent-child	0	1	0

Indirect Matching

Full Siblings



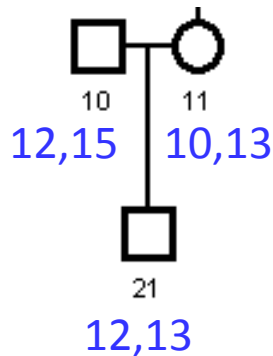
Probability of Sharing Alleles from a Common Ancestor

Relationship	0 alleles	1 allele	2 alleles
Full siblings	1/4	1/2	1/4

What information is required for kinship analysis?

- Alleged relationship
- Genotypes of specific markers
- Method to assess the relationship

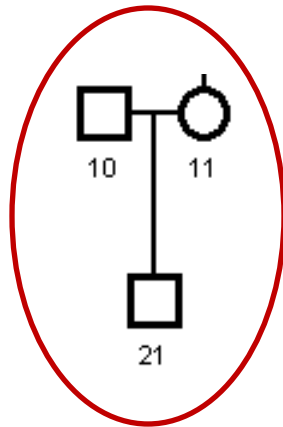
Paternity Index = $\frac{\text{Probability of genotypes if "10" is the true father of "21"}}{\text{Probability of genotypes if an unrelated man is the father of "21"}}$
(Likelihood Ratio)



Paternity trio

What information is required for kinship analysis?

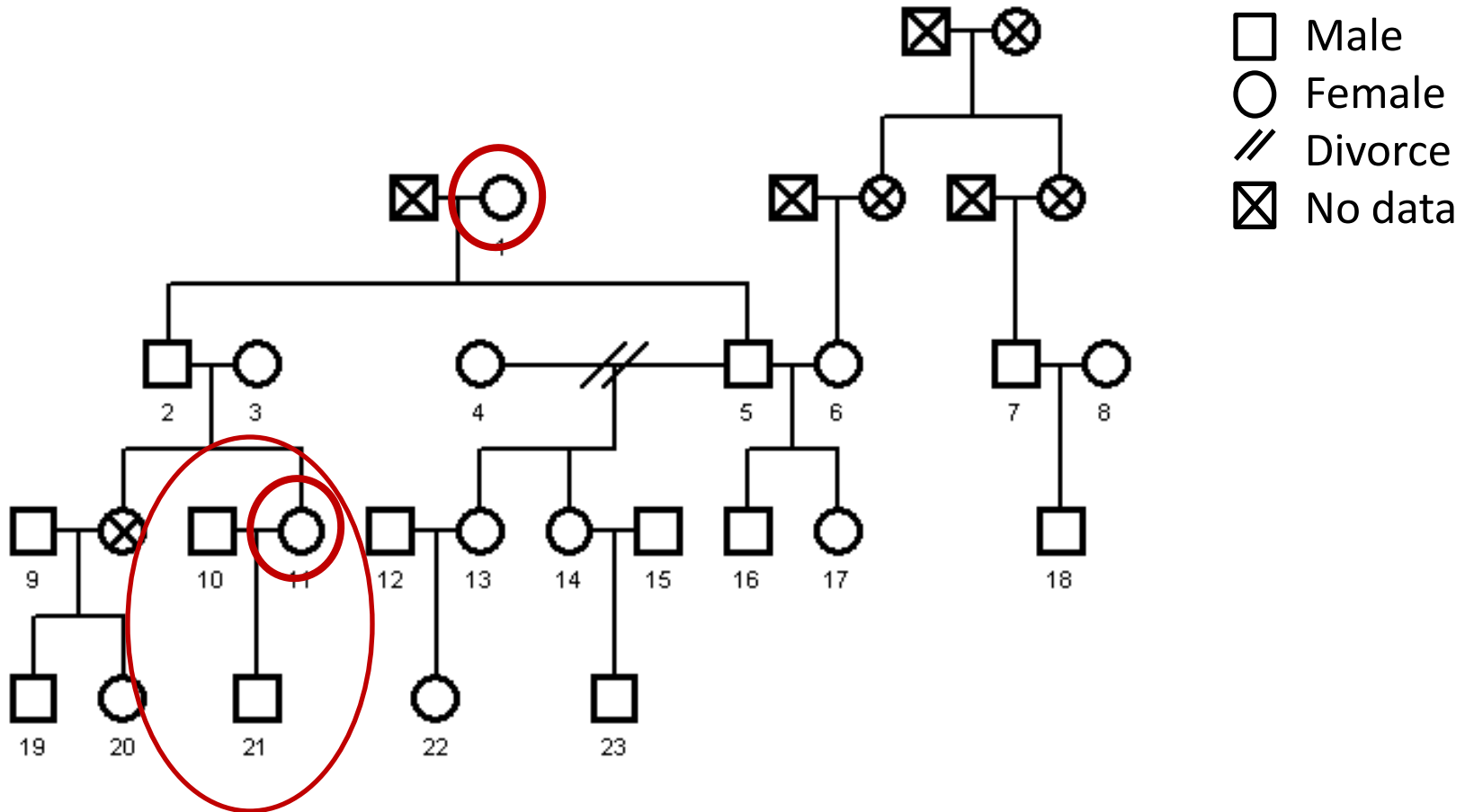
- Alleged relationship
- Genotypes of specific markers
- Method to assess the relationship



Paternity trio

Pedigrees are not always this simple

Complex Pedigree



Why can kinship analysis be complex?

For more distant familial relationships, allele sharing decreases → uncertainty increases

Probability of Sharing Alleles from a Common Ancestor

Relationship	0 alleles	1 allele	2 alleles
Parent-child	0	1	0
Full siblings	1/4	1/2	1/4
Half siblings	1/2	1/2	0
Uncle-nephew	1/2	1/2	0
Grandparent-grandchild	1/2	1/2	0
First cousins	3/4	1/4	0

High
Level of Certainty
↓
Low

Half siblings, uncle-nephew, and grandparent-grandchild are genetically identical

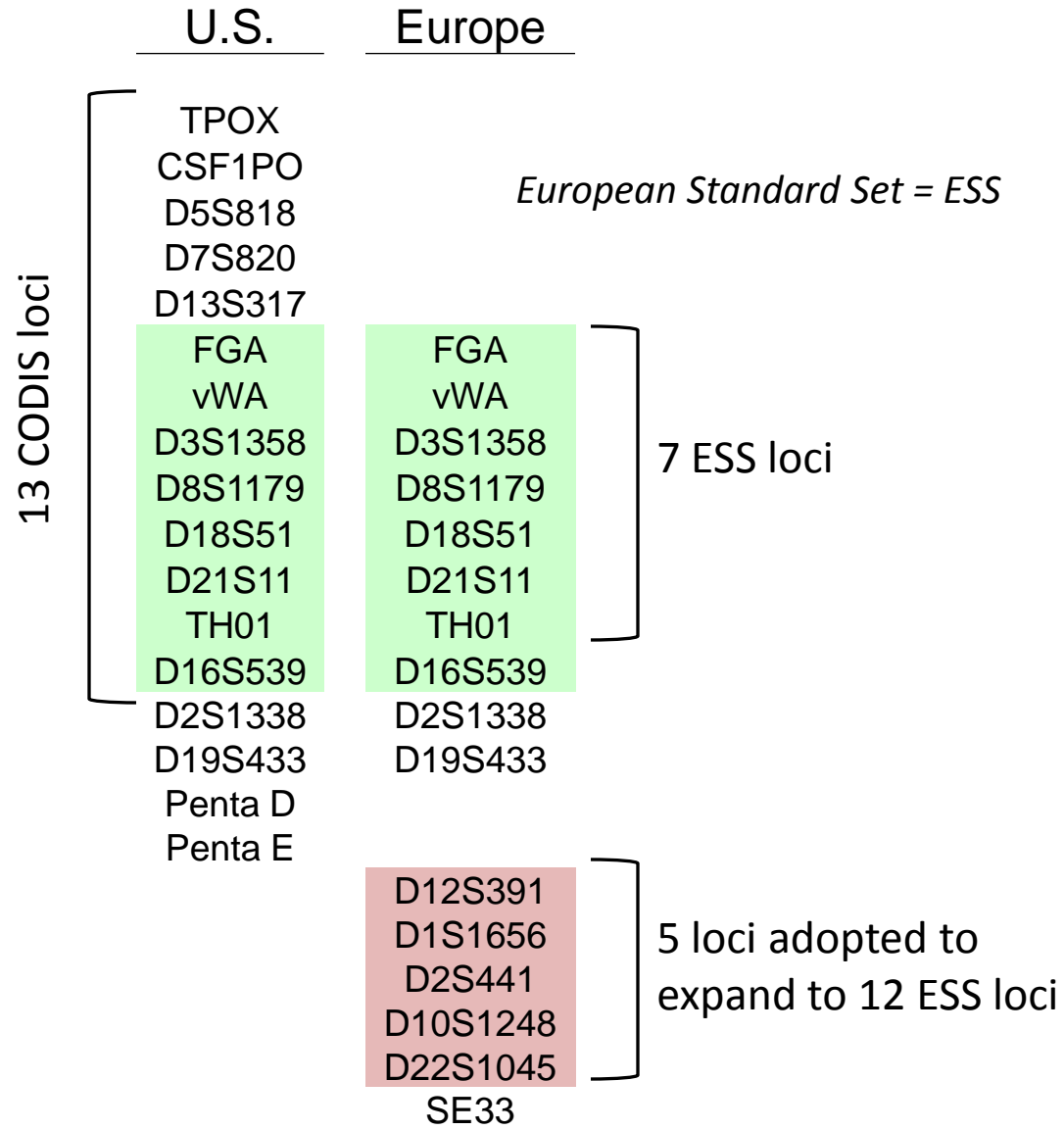
What materials are used for
kinship analysis at NIST?

What markers are being studied for kinship analysis?

- 46 autosomal loci
- 17 Y-chromosomal loci
- 15 X-STRs (AFDIL collaboration)
- Mitochondrial control region

Autosomal STR Markers

46 unique STR loci have been characterized at NIST



NIST 26plex Assay

(non-commercial)

- D1GATA113
- D1S1627
- D1S1677
- D2S1776
- D3S3053
- D3S4529
- D4S2364
- D4S2408
- D5S2500
- D6S474
- D6S1017
- D8S1115
- D9S1122
- D9S2157
- D10S1435
- D11S4463
- D12ATA63
- D14S1434
- D17S974
- D17S1301
- D2S441
- D10S1248
- D22S1045
- D18S853
- D20S482
- D20S1082

See poster #40 for details on additional loci

NIST Sample Set

- NIST U.S. population samples
 - 254 African American, 261 Caucasian, 139 Hispanic
- U.S. father/son samples
 - 178 African American, 198 Caucasian, 190 Hispanic, 198 Asian
- Extended family samples
 - 6 sets of 3–4 generations
 - 165 total samples

NIST Data Analysis Capabilities

Kinship Software

- DNA-VIEW™ v. 29.23 (Charles Brenner)
- KIn CALc v. 4.0 (CA DOJ, Steven Myers)
- GeneMarker® HID v. 1.90 (SoftGenetics)
- FSS DNA Lineage (Forensic Science Service)
- LISA (Future Technologies Inc.)

Population Genetics Software

- Arlequin v. 3.5

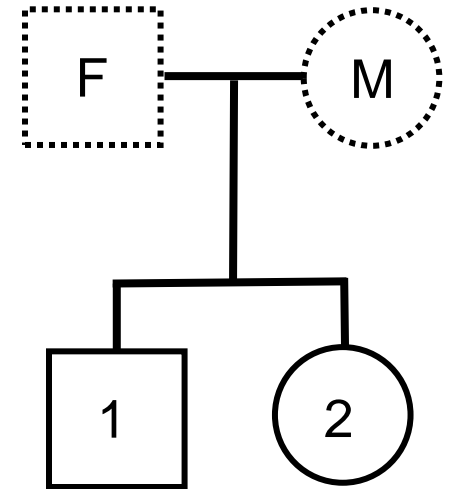
What methods are we using
to assess kinship?

How is kinship assessed?

Likelihood Ratio (LR)

Evaluate genotypes to give weight (strength) to compared relationships

$$LR = \frac{\text{Probability of genotypes if 1,2 are full siblings}}{\text{Probability of genotypes if 1,2 are unrelated}}$$



By the definition of a LR:

LR > 1 supports the numerator (alleged relationship)

LR < 1 supports the denominator (unrelated)

Larger LR values provide more support for the alleged relationship

How is kinship assessed?

Goal: How well does a set of loci perform for kinship analysis?

Method: Evaluate “expected” range of LR_s for different relationship questions

Need: Graphical method to display the LR_s

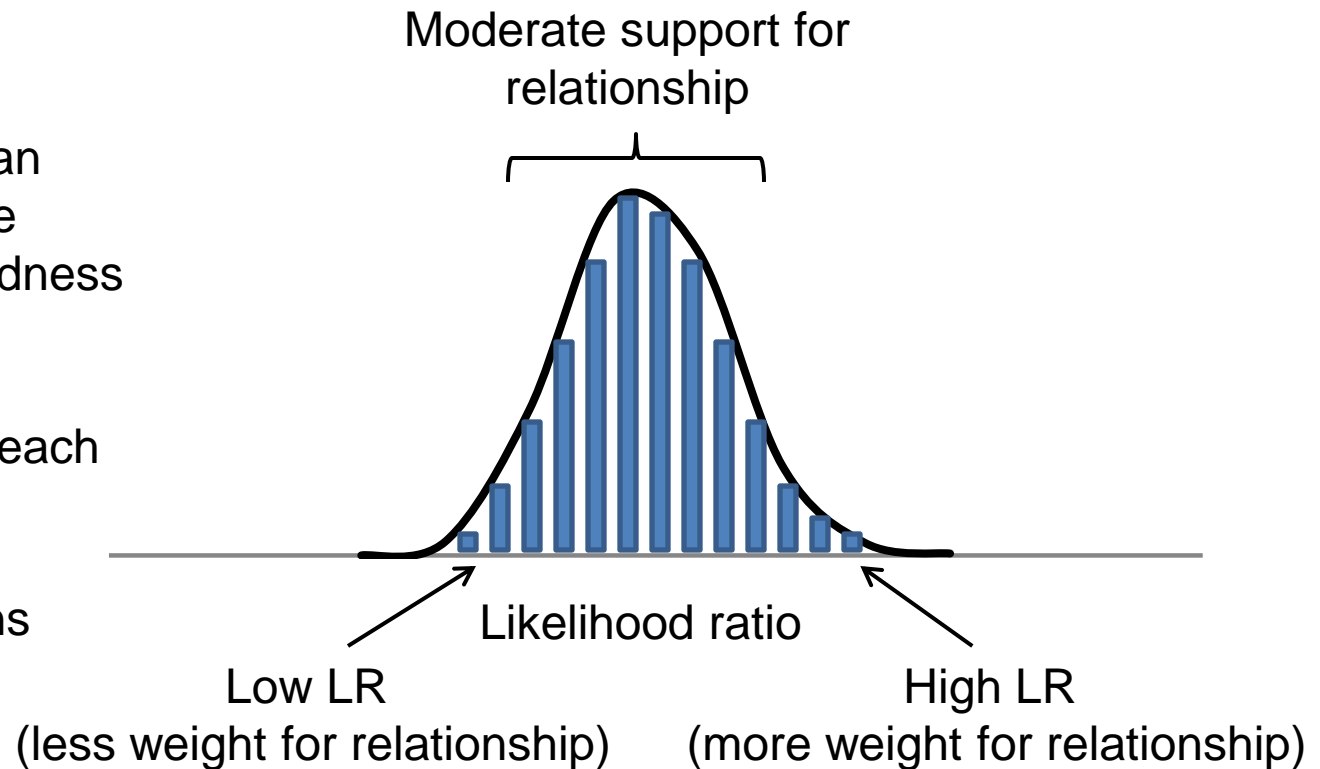
Solution: Distribution of data points (LR values)

What is a likelihood ratio distribution?

Computer simulations can generate many genotype combinations and relatedness scenarios (pedigrees)

Calculate LR values for each pedigree

Generate LR distributions



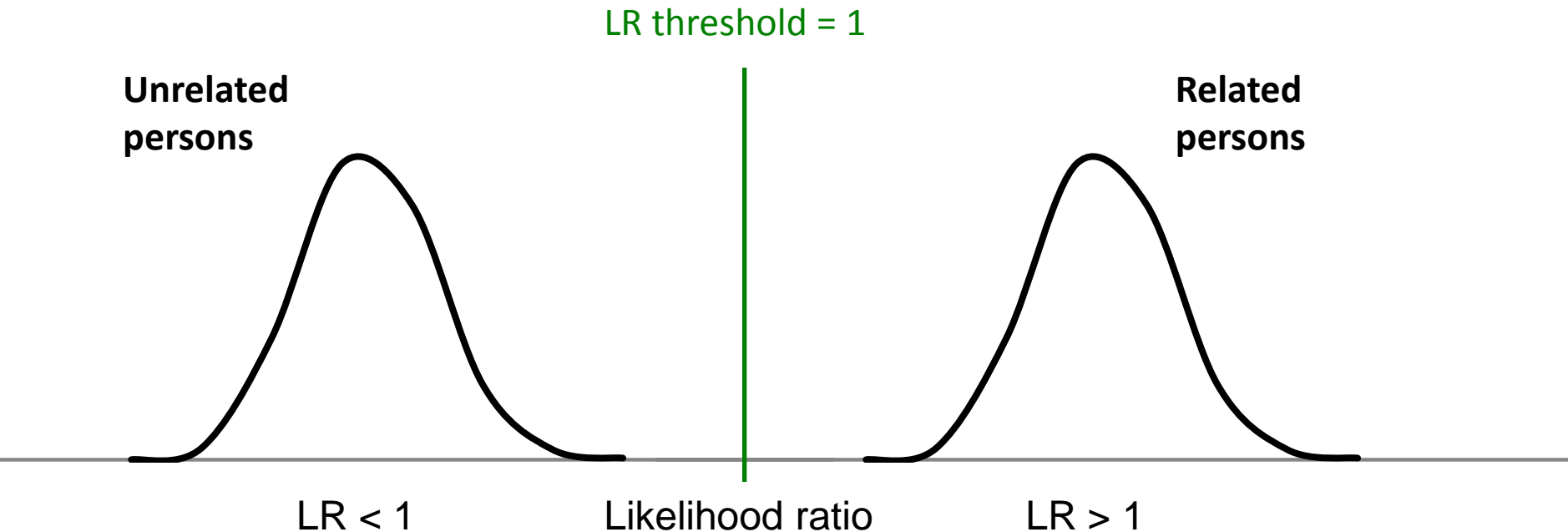
Variables:

Allele frequency

Number of loci

Kinship probabilities (account for shared alleles from common ancestor)

What is a likelihood ratio distribution?



Variables:

Allele frequency

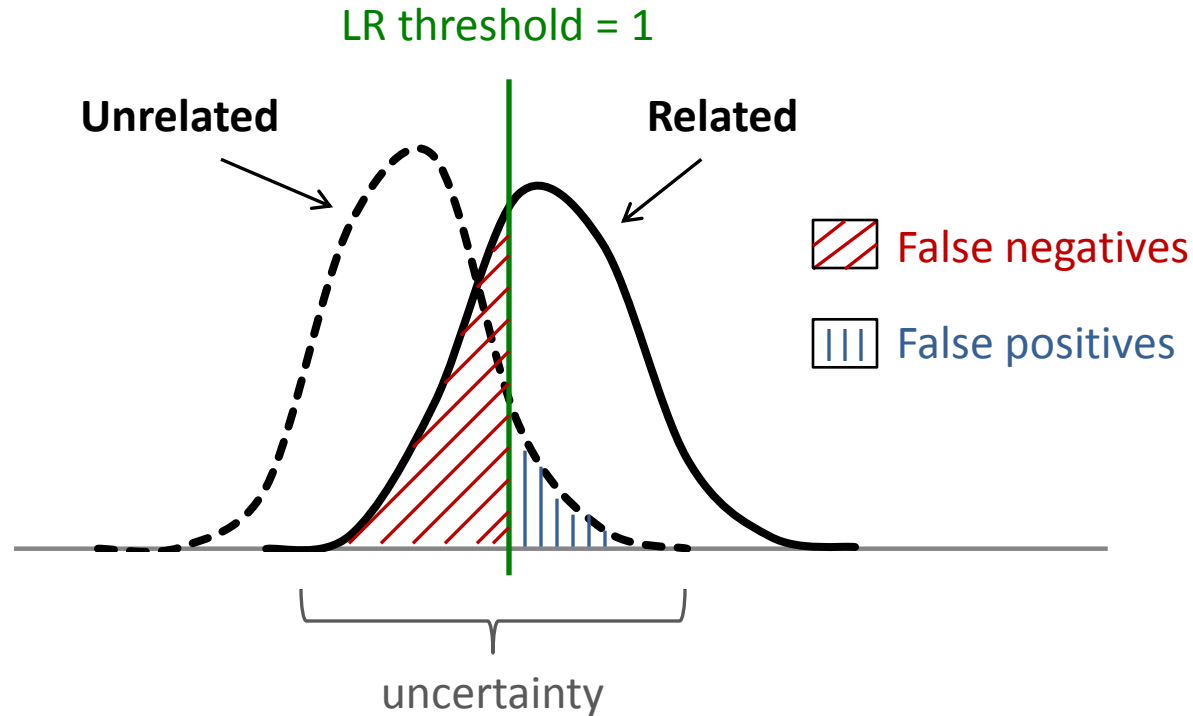
Number of loci

Kinship probabilities (account for shared alleles from common ancestor)

Separate distributions

- High probability of shared alleles from common ancestor (e.g., parent-offspring)
- Discriminating loci genotyped

Overlap of likelihood ratio distributions



Overlapping distributions

- Low probability of shared alleles from common ancestor (e.g., first cousin)
- Less discriminating loci genotyped

Variables:

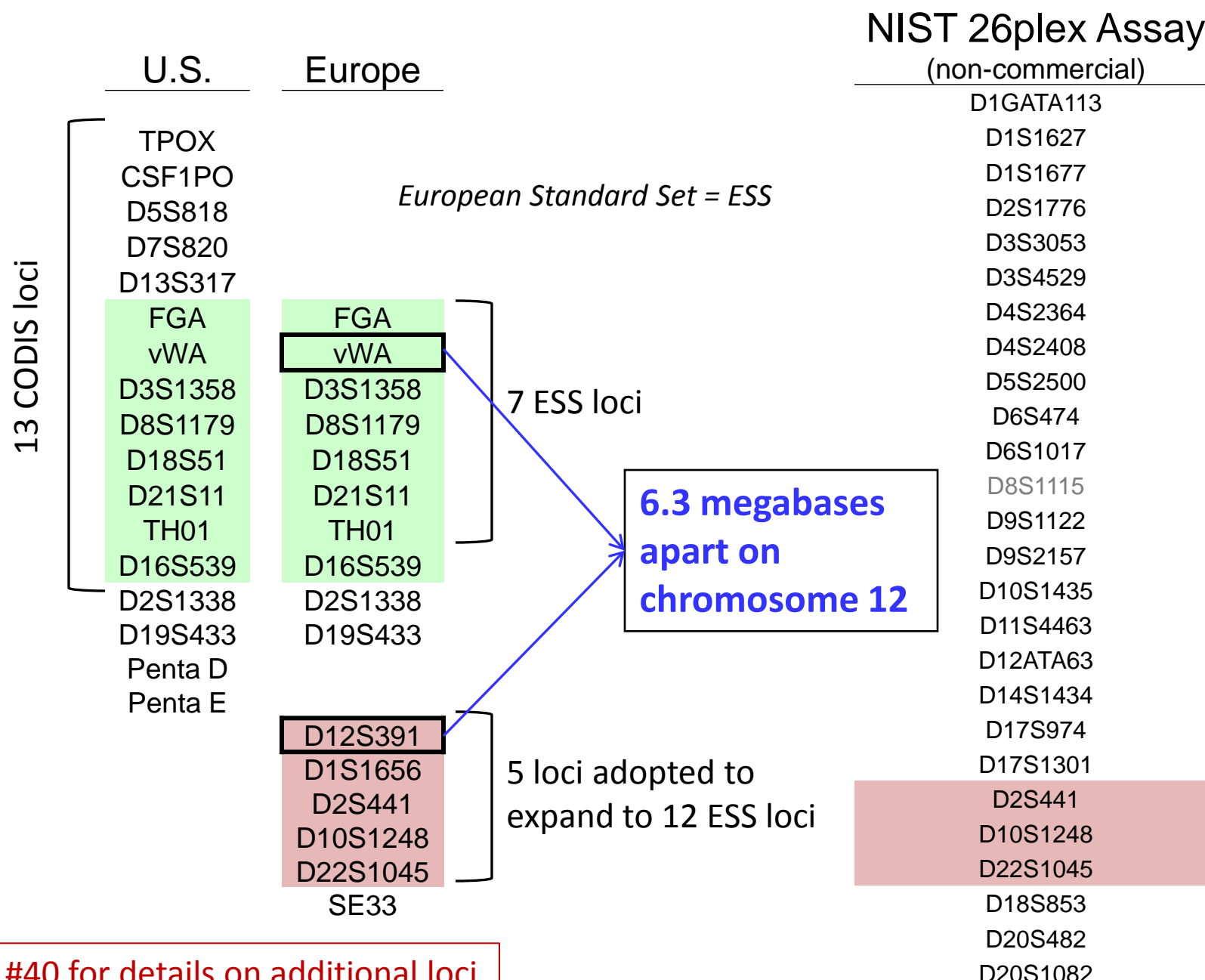
Allele frequency

Number of loci

Kinship probabilities (account for shared alleles from common ancestor)

What kinship questions have we asked
with our dataset?

46 unique STR loci have been characterized at NIST

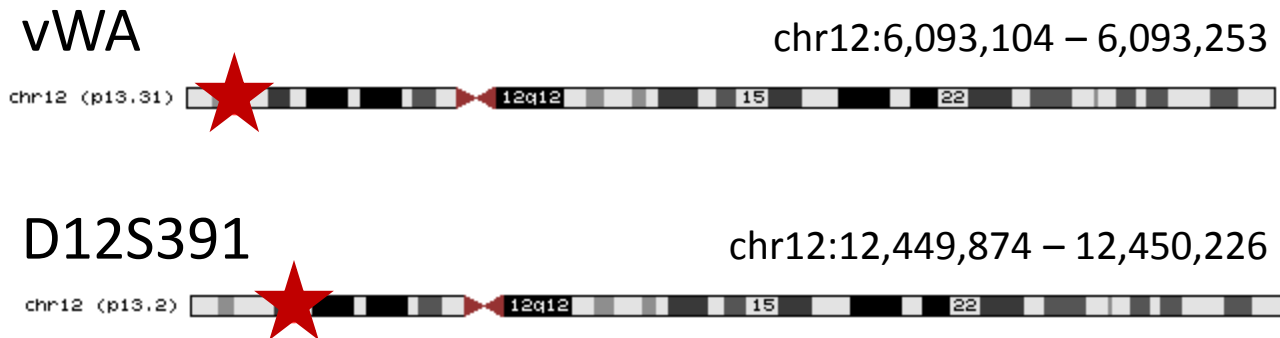


See poster #40 for details on additional loci

Can D12S391 be used with vWA for kinship analysis?

6.3 megabases apart on chromosome 12

UCSC Genome Browser, Feb. 2009 assembly



Are vWA and D12S391 independent? **No!**

Should vWA and D12S391 be multiplied for profile probability calculations in kinship analysis? **No!**

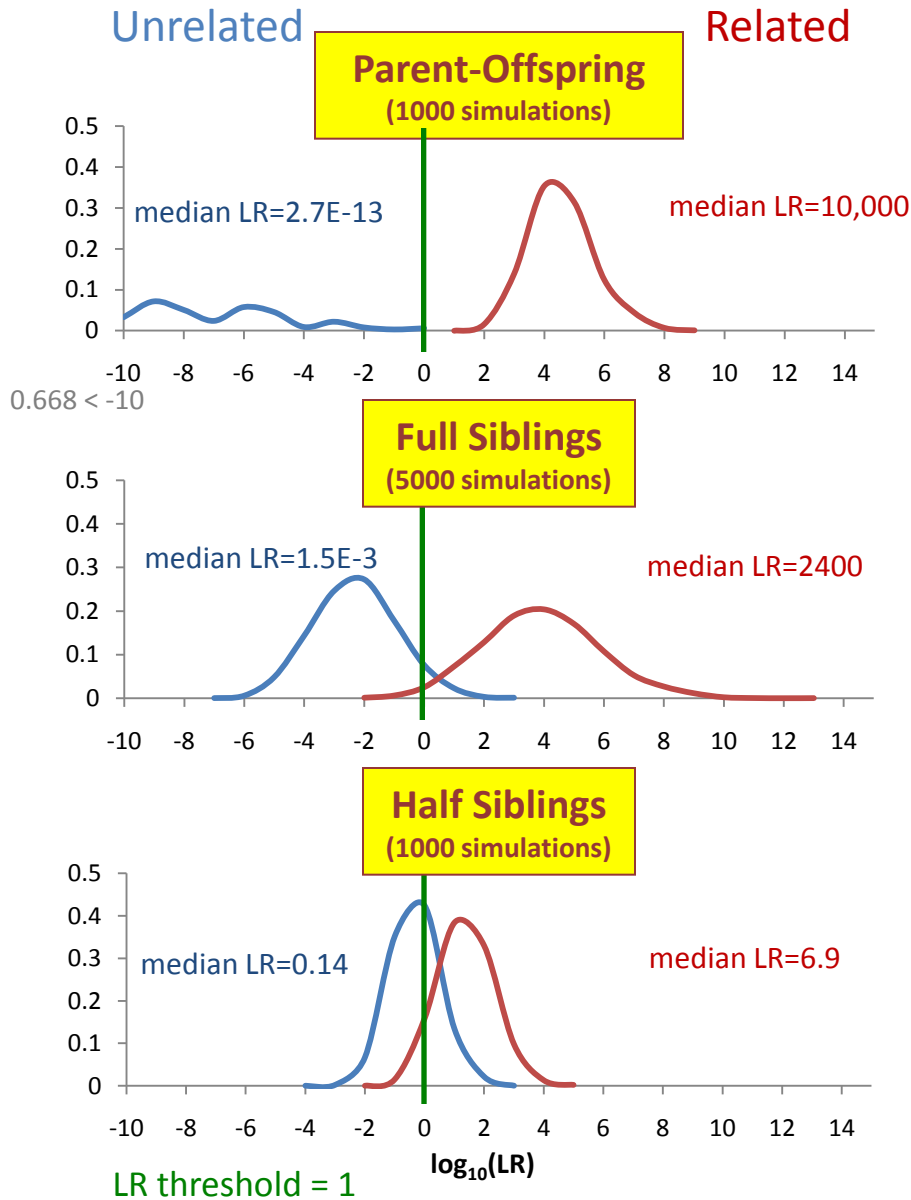
O'Connor KL, et al., Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples, *Forensic Sci. Int. Genet.* (in press)

Budowle B, et al., Population genetic analyses of the NGM STR loci, *Int. J. Legal Med.* (in press)

How do 13 loci perform for
kinship analysis?

How do 13 loci perform for kinship analysis?

Proportion of simulations using U.S. Caucasian allele frequencies



The degree of overlap corresponds with possible values for false positive or false negative results.

Parent-offspring comparisons:
No overlap between unrelated and related LR distributions

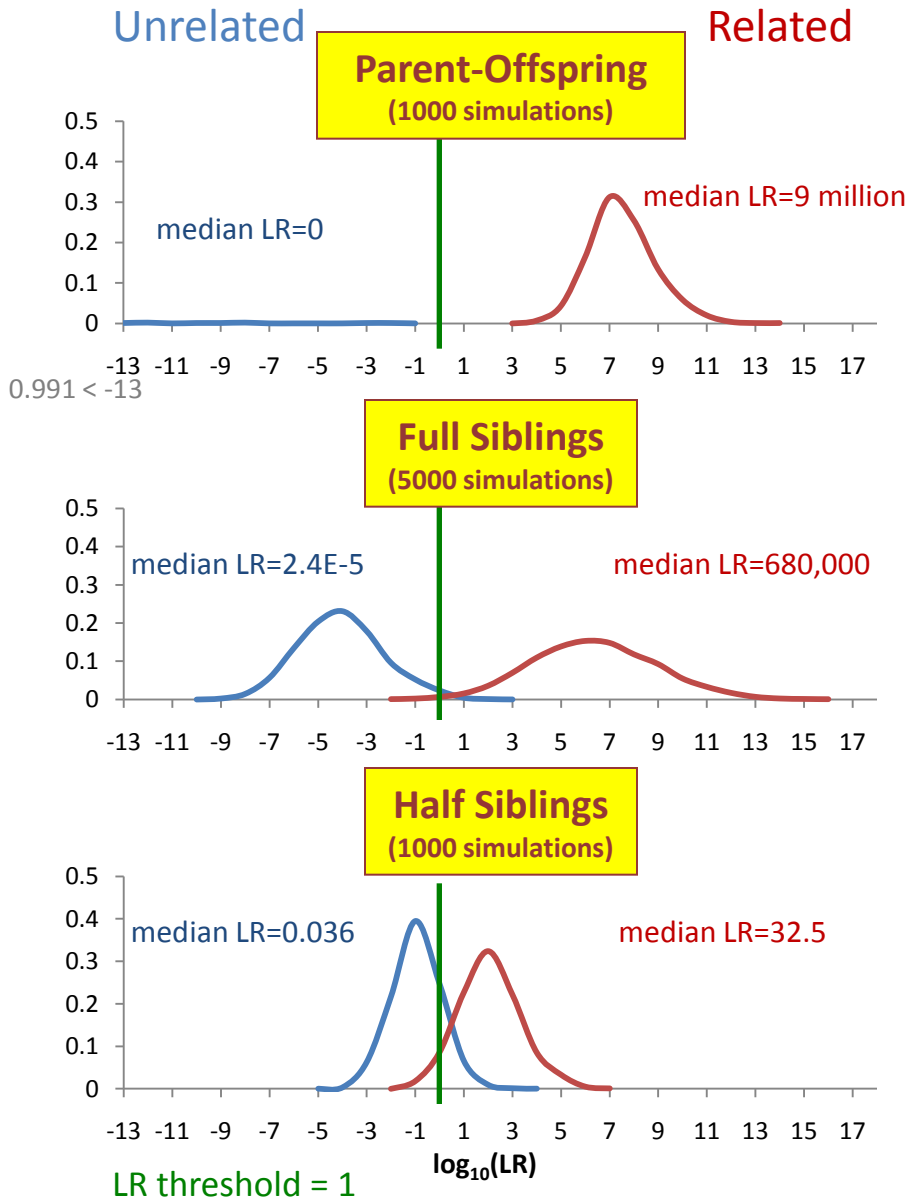
Full sibling comparisons:
False positive rate = 0.027
False negative rate = 0.033

Half sibling comparisons:
False positive rate = 0.155
False negative rate = 0.168

Do additional loci improve the
discrimination of true relatives
vs. unrelated persons?

How do 20 loci perform for kinship analysis?

Proportion of simulations using U.S. Caucasian allele frequencies



Additional loci improve separation of LR distributions for **parent-offspring** and **full siblings**.

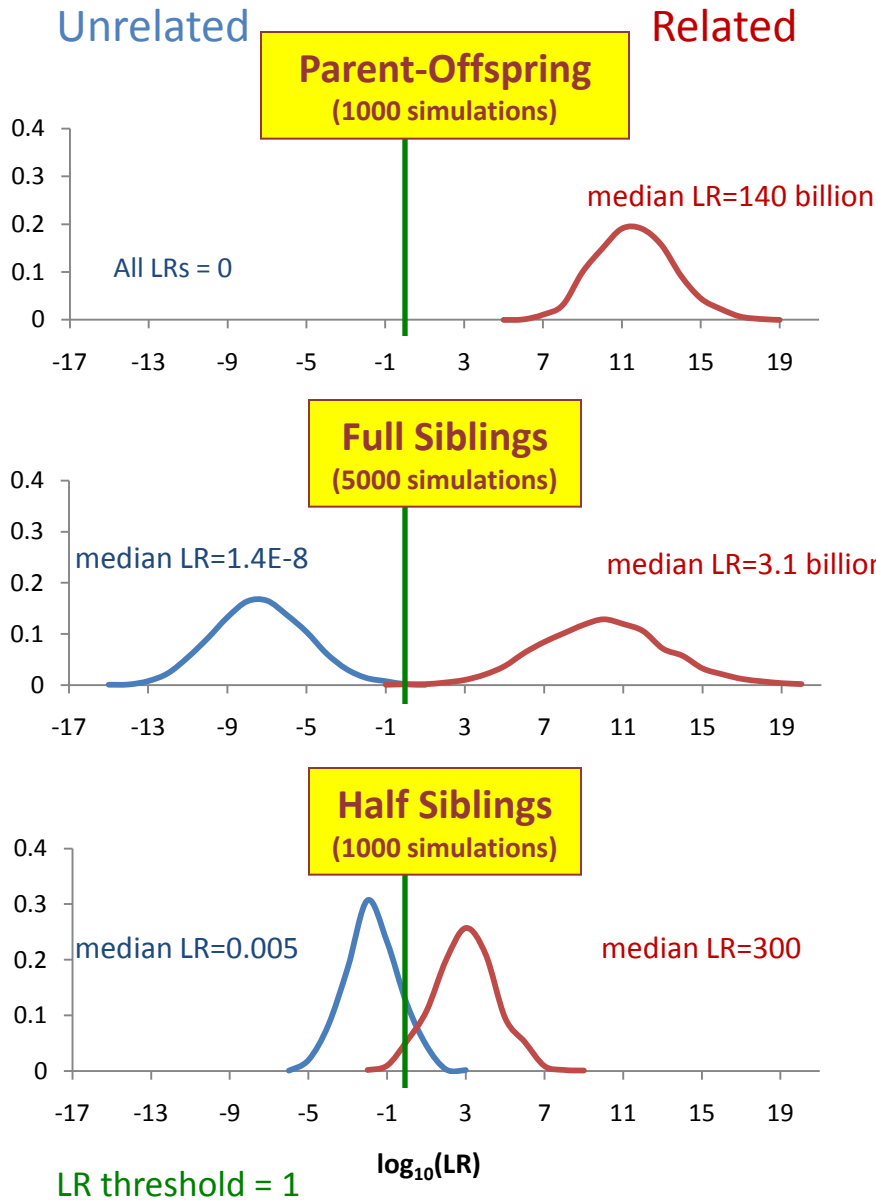
Parent-offspring comparisons:
No overlap between unrelated and related LR distributions

Full sibling comparisons:
False positive rate = 0.006
False negative rate = 0.008

Half sibling comparisons:
False positive rate = 0.075
False negative rate = 0.104

How do 40 loci perform for kinship analysis?

Proportion of simulations using U.S. Caucasian allele frequencies



Additional loci further improve separation of LR distributions for **parent-offspring** and **full siblings**.

Parent-offspring comparisons:
No overlap between unrelated and related LR distributions

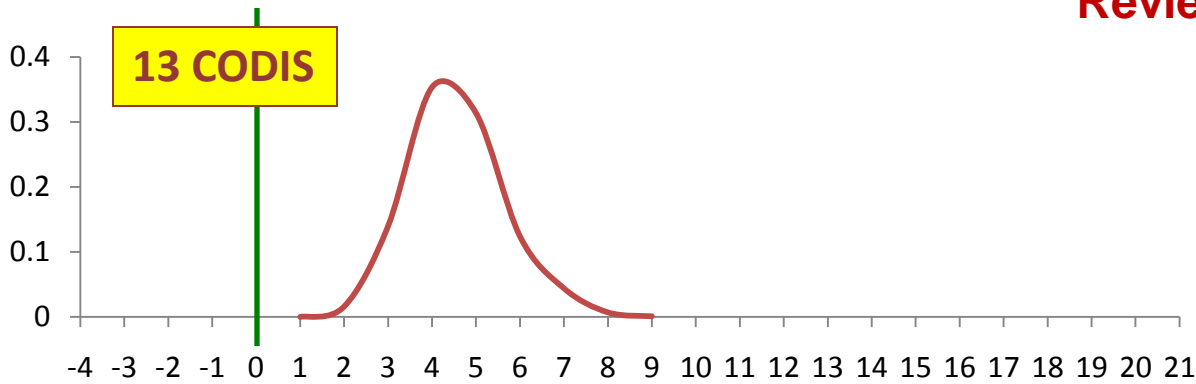
Full sibling comparisons:
False positive rate = 0.0006
False negative rate = 0.0018

Half sibling comparisons:
False positive rate = 0.051
False negative rate = 0.066

Do additional loci improve kinship determination?

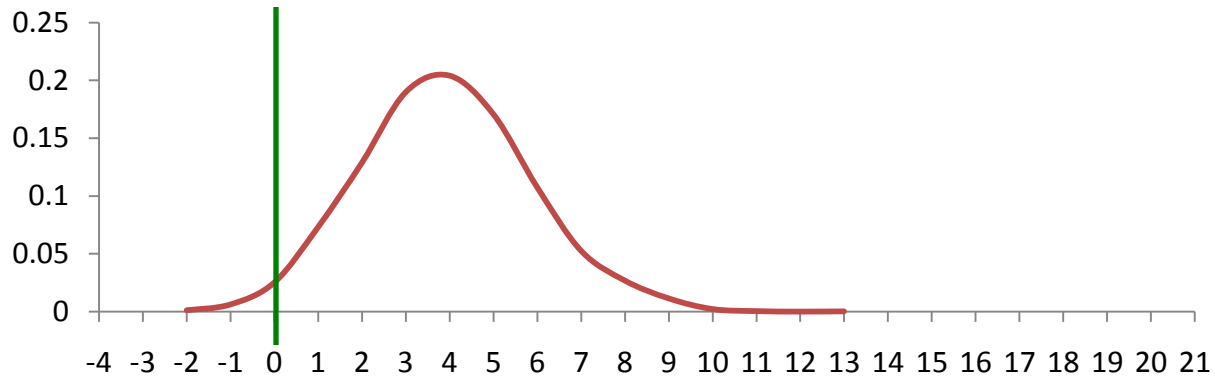
Review previous data

Proportion of simulations using
U.S. Caucasian allele frequencies



Parent-offspring

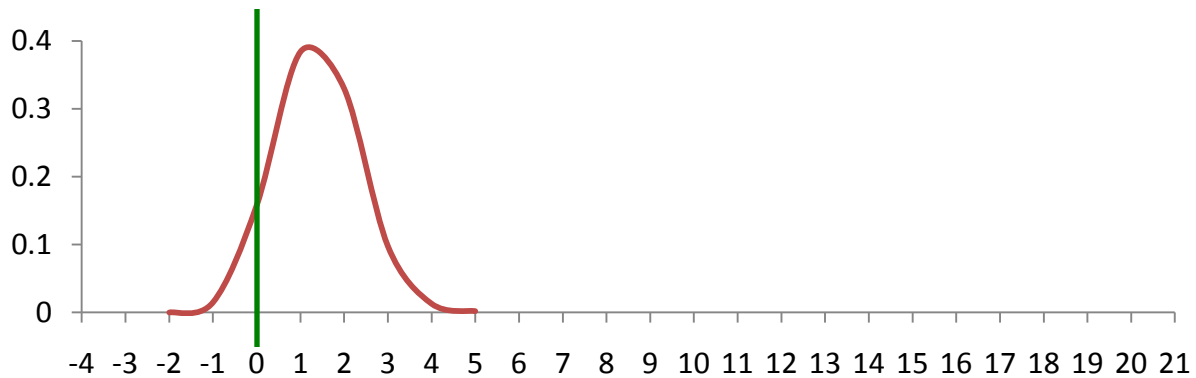
1000 simulations



Full siblings

5000 simulations

(additional simulations
performed for smoother curves)



Half siblings

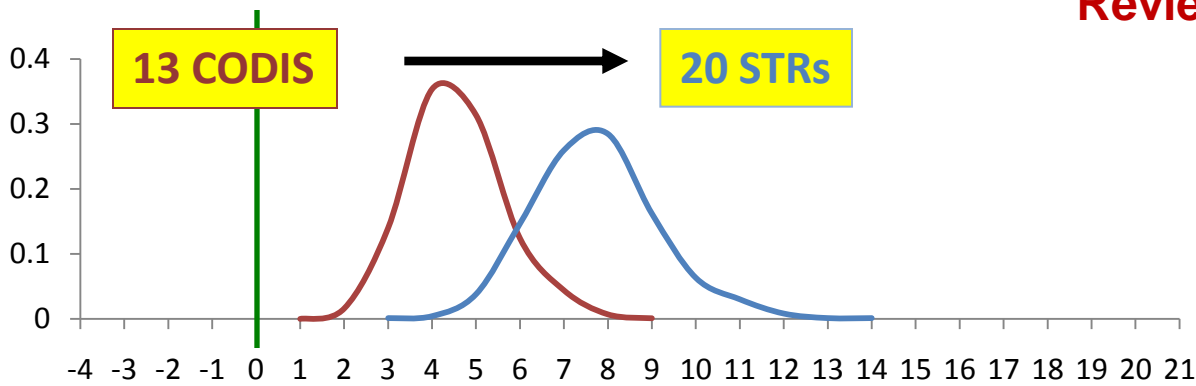
1000 simulations

$\log_{10}(\text{LR})$

Do additional loci improve kinship determination?

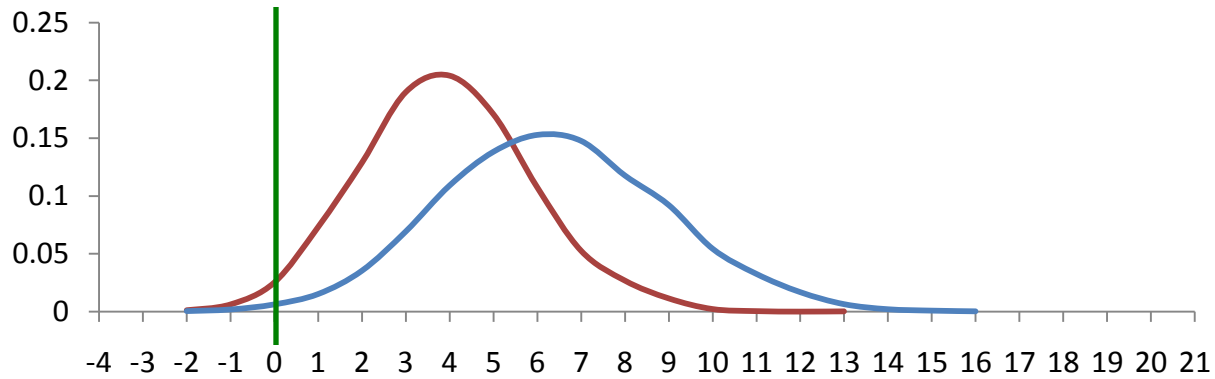
Review previous data

Proportion of simulations using
U.S. Caucasian allele frequencies



Parent-offspring

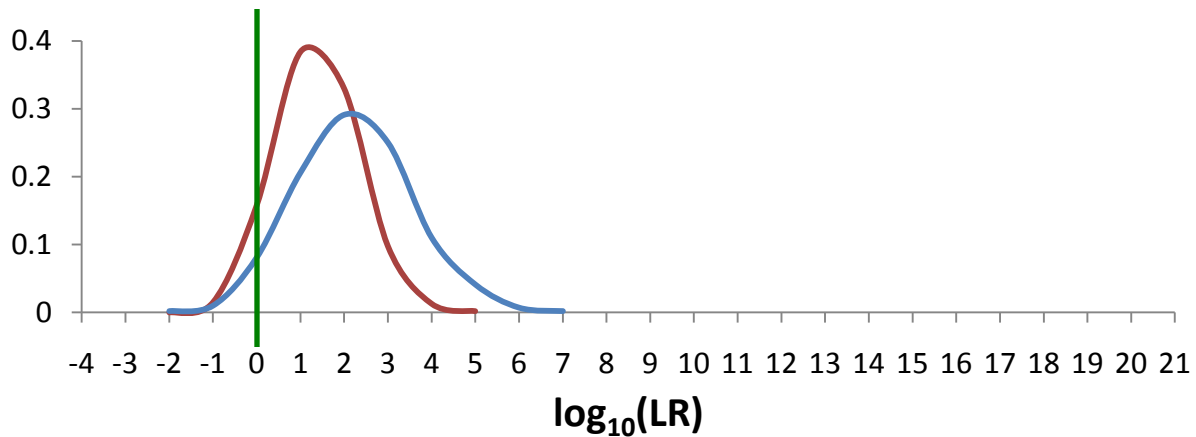
1000 simulations



Full siblings

5000 simulations

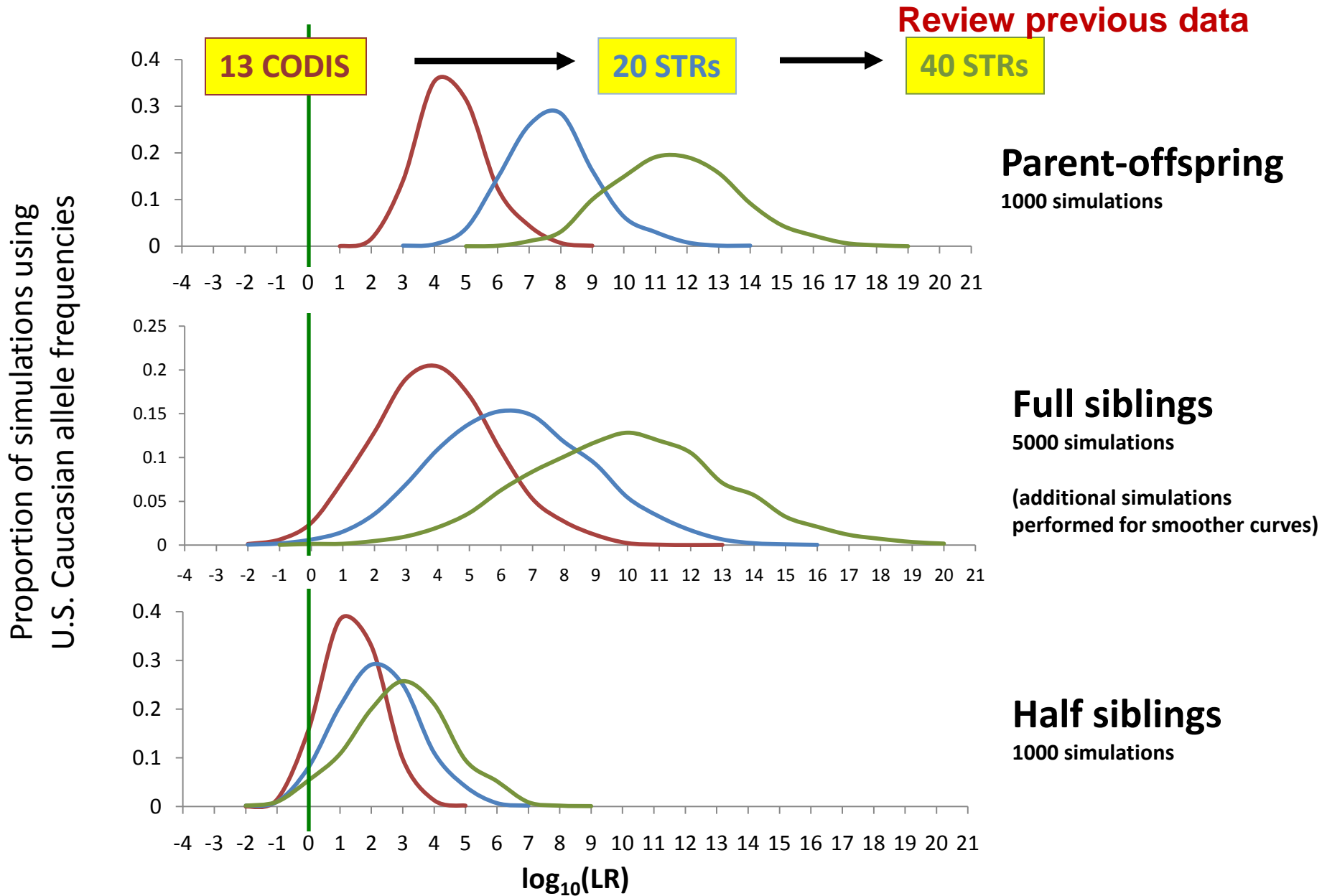
(additional simulations
performed for smoother curves)



Half siblings

1000 simulations

Do additional loci improve kinship determination?



How can uncertainty in kinship determination be reduced?

How can uncertainty in kinship determination be reduced?

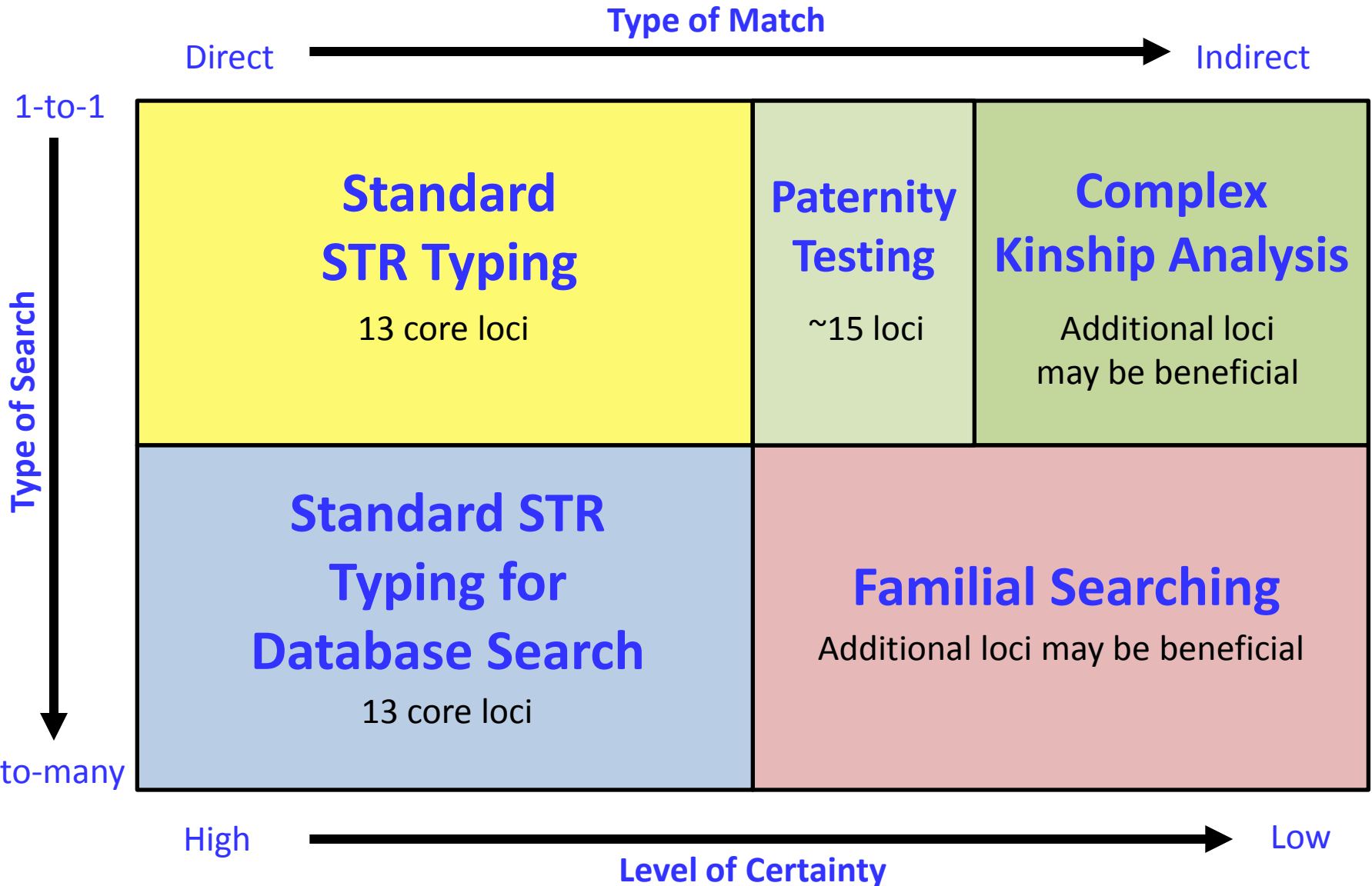
Improve the measurement technique

- Add more family references
- Add more loci *** More chances for mutation**
 - Autosomal STRs* improve identification of **parent-offspring and full siblings**
 - Lineage markers or SNP arrays may improve identification of more distant relatives

Know your limits... simulate... validate!

How does kinship analysis relate to questions that concern you?

Expanding the Forensic Core Competency



Expanding the Forensic Core Competency to Familial Searching

- One-to-many search → many false positives
- Allele sharing method
 - Partial match between evidence profile and database profile
 - Miss true relatives
 - Especially full siblings (1/4 probability of sharing 0 alleles at a locus)
 - Introduce many false positives due to chance allele sharing
- Likelihood ratio approach
 - Kinship probabilities plus allele frequencies account for allele sharing due to familial relationship
- Reduce uncertainty with additional loci (autosomal STRs, Y-STRs)

What is NIST doing to improve
kinship analysis?

What is NIST doing to improve kinship analysis?

Allele frequencies for U.S. population samples

Evaluation of new loci **See Poster #40 tomorrow**

Concordance testing of new multiplexes

Developed a new website to support kinship analysis

Kinship Resource Page on STRBase

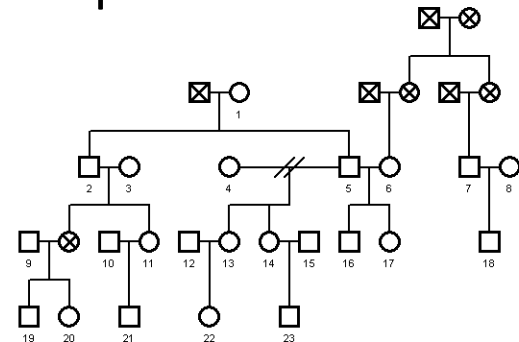
www.cstl.nist.gov/strbase/kinship.htm

NIST Standard Reference Family Data

Aid **validation** of algorithms, software, and loci selection for kinship analysis

- Use genotypes with known inheritance
- Compare LR from algebraic and software calculations
- Test algorithms for mutation, rare alleles, null alleles, incest
- Evaluate use of additional loci to detect relationships

See Poster #35 today



Acknowledgments



Applied Genetics
Group Leader



John
Butler

Workshops
& Textbooks



Becky
Hill

Concordance
& LT-DNA

DNA Biometrics
Project Leader



Peter
Vallone

Rapid PCR
& Biometrics



Kristen
Lewis O'Connor

Kinship Analysis



Erica
Butts

DNA Extraction
Efficiency

Kinship Page on STRBase

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

kristen.oconnor@nist.gov

Funding

NRC – Postdoctoral fellowship support for Kristen O'Connor

FBI – Application of DNA Typing as a Biometric Tool

NIJ – Forensic DNA Standards, Research, and Training