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

Kristen Lewis O'Connor, Ph.D.
Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland, USA

English Speaking Working Group
Stockholm, Sweden

NIST Efforts to Improve Kinship Analysis

- Assist forensic and biometrics communities with kinship testing needs
- Evaluate current and new loci for kinship analysis
- Develop standard datasets and tools to improve validation of kinship analysis methods and software

STR Loci Present in Commercial Kits

13 CODIS loci

U.S.			Europe			
PP16	Identifiler	MiniFiler	ESX/ESI17	NGM	SEfiler	SGM Plus
TPOX	TPOX					
CSF1PO	CSF1PO	CSF1PO				
D5S818	D5S818					
D7S820	D7S820	D7S820				
D13S317	D13S317	D13S317				
FGA	FGA	FGA	FGA	FGA	FGA	FGA
vWA	vWA		vWA	vWA	vWA	vWA
D3S1358	D3S1358		D3S1358	D3S1358	D3S1358	D3S1358
D8S1179	D8S1179		D8S1179	D8S1179	D8S1179	D8S1179
D18S51	D18S51	D18S51	D18S51	D18S51	D18S51	D18S51
D21S11	D21S11	D21S11	D21S11	D21S11	D21S11	D21S11
TH01	TH01		TH01	TH01	TH01	TH01
D16S539	D16S539	D16S539	D16S539	D16S539	D16S539	D16S539
	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338
	D19S433		D19S433	D19S433	D19S433	D19S433
			D12S391	D12S391		
			D1S1656	D1S1656		
			D2S441	D2S441		
			D10S1248	D10S1248		
			D22S1045	D22S1045		
			SE33		SE33	
Penta D						
Penta E						

7 ESS loci
- before
April 2009

Adopted 5
loci to
expand to
12 ESS loci

U.S. is looking to expand the core loci (18-20 total)

to provide more international overlap

23 unique loci

Genetic Data for Analysis of STR Loci

Simulated and **collected** genetic data were used to evaluate the discriminatory power gained by adding loci to the standard U.S. and European forensic panels

- **Simulated** datasets of related and unrelated genotypes

  Genotypes simulated with DNA-VIEW™ v. 29.23 (Charles Brenner)
  - Parent-offspring
  - Full siblings
  - Half siblings/uncle-nephew/grandparent-grandchild
  - Unrelated pairs

- **Collected** pedigree data
  - Six extended pedigrees (3-4 generations, 165 total samples)
  - Number of pairwise comparisons:
    - 1st degree = 310, 2nd degree = 316, 3rd degree = 291, 4th degree = 172

Pairwise comparisons provide a good model to evaluate effect of additional loci. Adding more individuals to the analysis can improve kinship determination.

Nothnagel M, Schmidtke J, Krawczak M. (2010) Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci, *Int. J. Legal Med*. 124(3):205-15.

Forensic STR Loci Examined for Kinship Analysis

Sets of Loci	Number of STRs
CODIS	13
Identifiler®	15
PowerPlex® 16	15
PowerPlex® ESI/ESX 16	14
PowerPlex® ESI/ESX 17	15
Identifiler® + 5 Euro	19
Identifiler® + 5 Euro + SE33	20
NIST 26plex	25
Identifiler® + NIST 26plex	40

Removed vWA for sets with D12S391

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

Forensic STR Loci Examined for Kinship Analysis

Sets of Loci	**Number of STRs**
CODIS	**13**
Identifiler®	15
PowerPlex® 16	15
PowerPlex® ESI/ESX 16	14
PowerPlex® ESI/ESX 17	**15**
Identifiler® + 5 Euro	19
Identifiler® + 5 Euro + SE33	**20**
NIST 26plex	25
Identifiler® + NIST 26plex	40

Removed vWA for sets with D12S391

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

Kinship Statistics

- Allele frequency data from NIST U.S. Caucasian, African American, Hispanic population samples

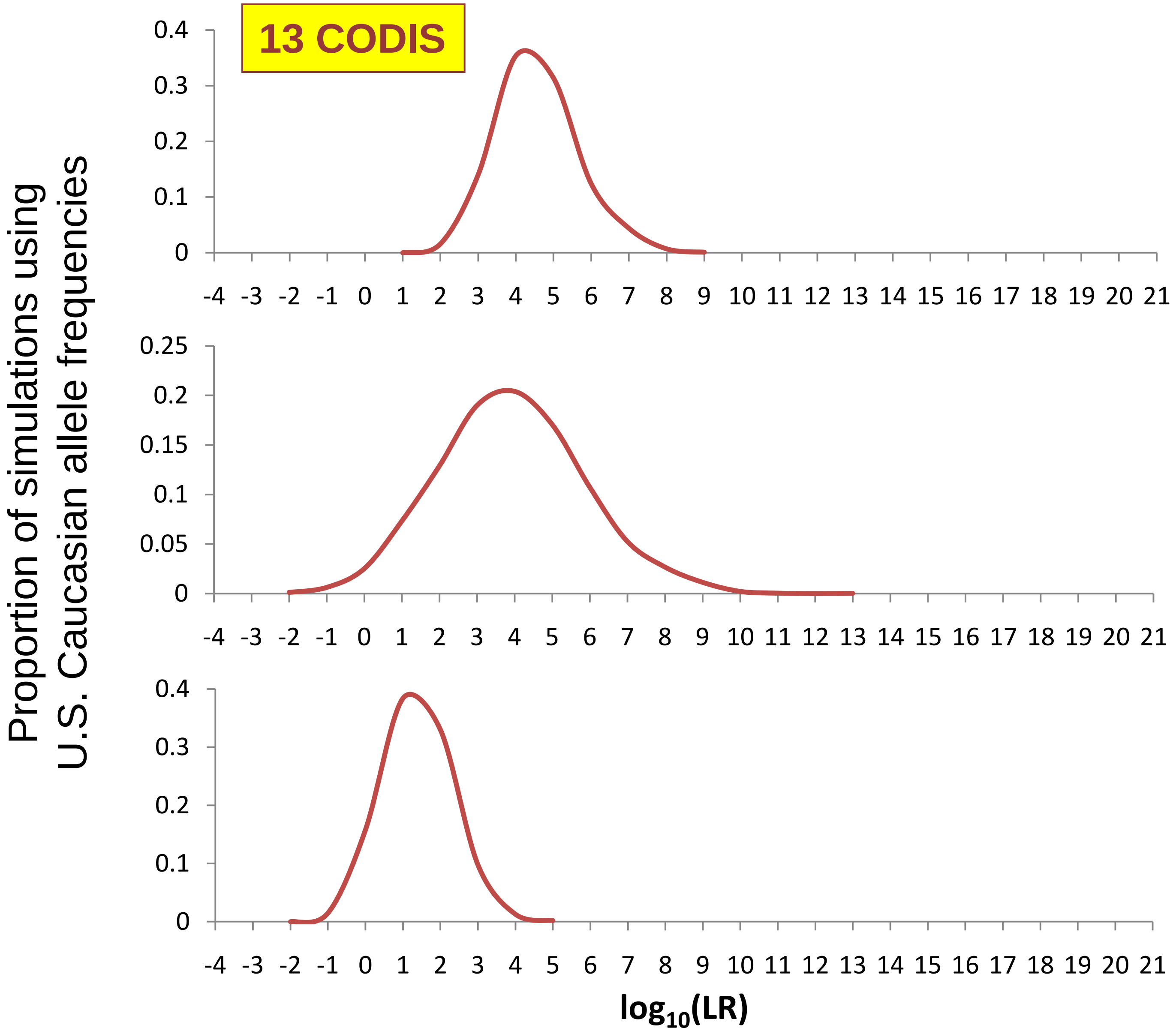
- Likelihood ratio algorithms
  - Simulated data: DNA-VIEW™ v29.23 (Charles Brenner)
  - Real pedigree data: GeneMarker® HID v1.90 (SoftGenetics)
  - Mutation models were applied

Butler, J.M., et al. (2003) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, *J. Forensic Sci*. 48(4):908-911.
Hill, C.R., et al. (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, *J. Forensic Sci*. 53(1):73-80.
Hill, C.R., et al., Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems, *Forensic Sci. Int. Genet.* (in press).

NIST Efforts to Assist Validation of Kinship Analysis

- Simulated data
  - Model a large number of different genotype combinations for related and unrelated individuals
  - Observe expected ranges of likelihood ratio values
  - Evaluate discriminatory power of additional loci

Likelihood Ratio Distributions for True Relationships Using Different Sets of STR Loci



Parent-offspring

1000 simulations

Full siblings

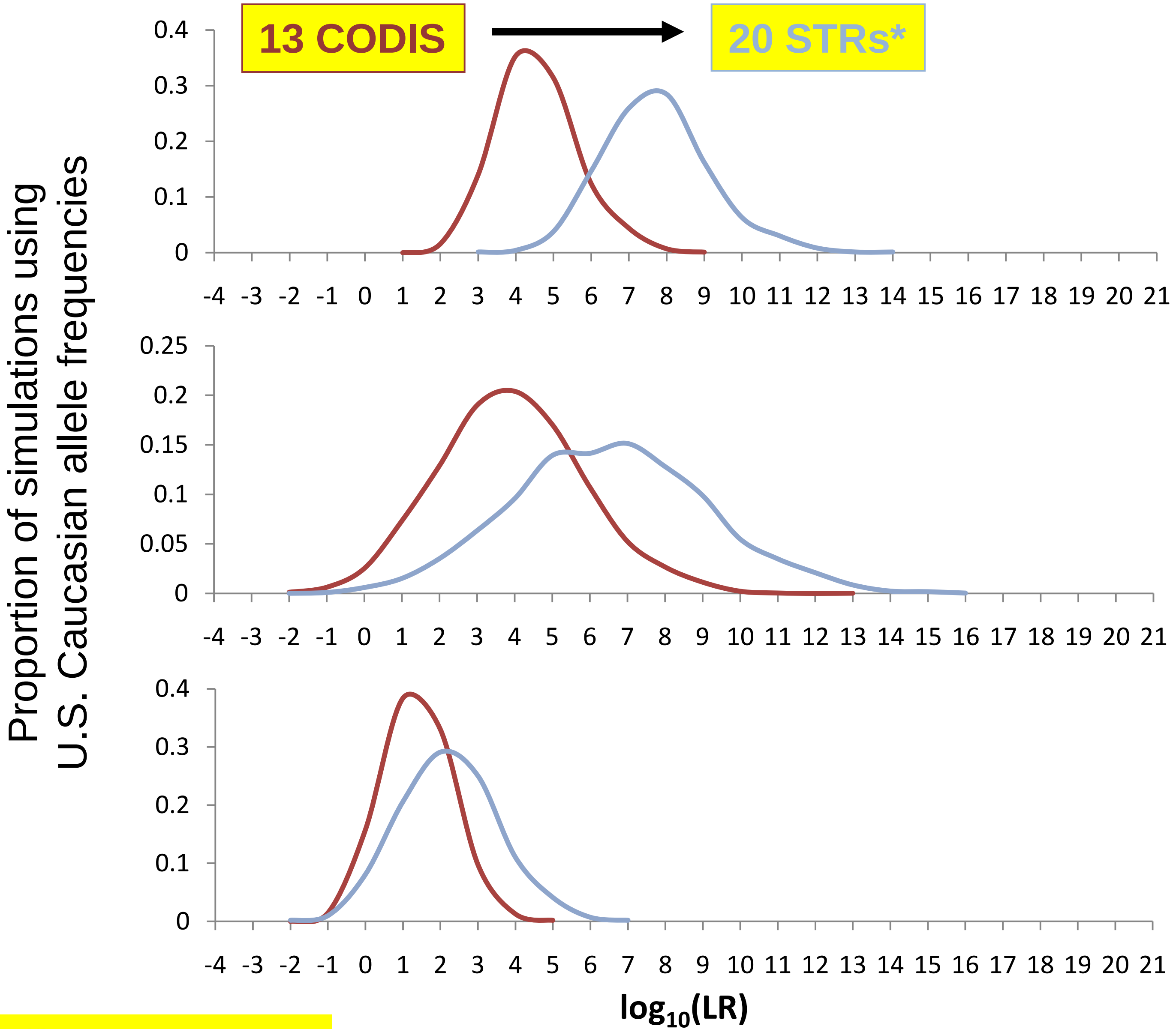
5000 simulations

(additional simulations performed for smoother curves)

Half siblings

1000 simulations

Likelihood Ratio Distributions for True Relationships Using Different Sets of STR Loci



Parent-offspring

1000 simulations

Full siblings

5000 simulations

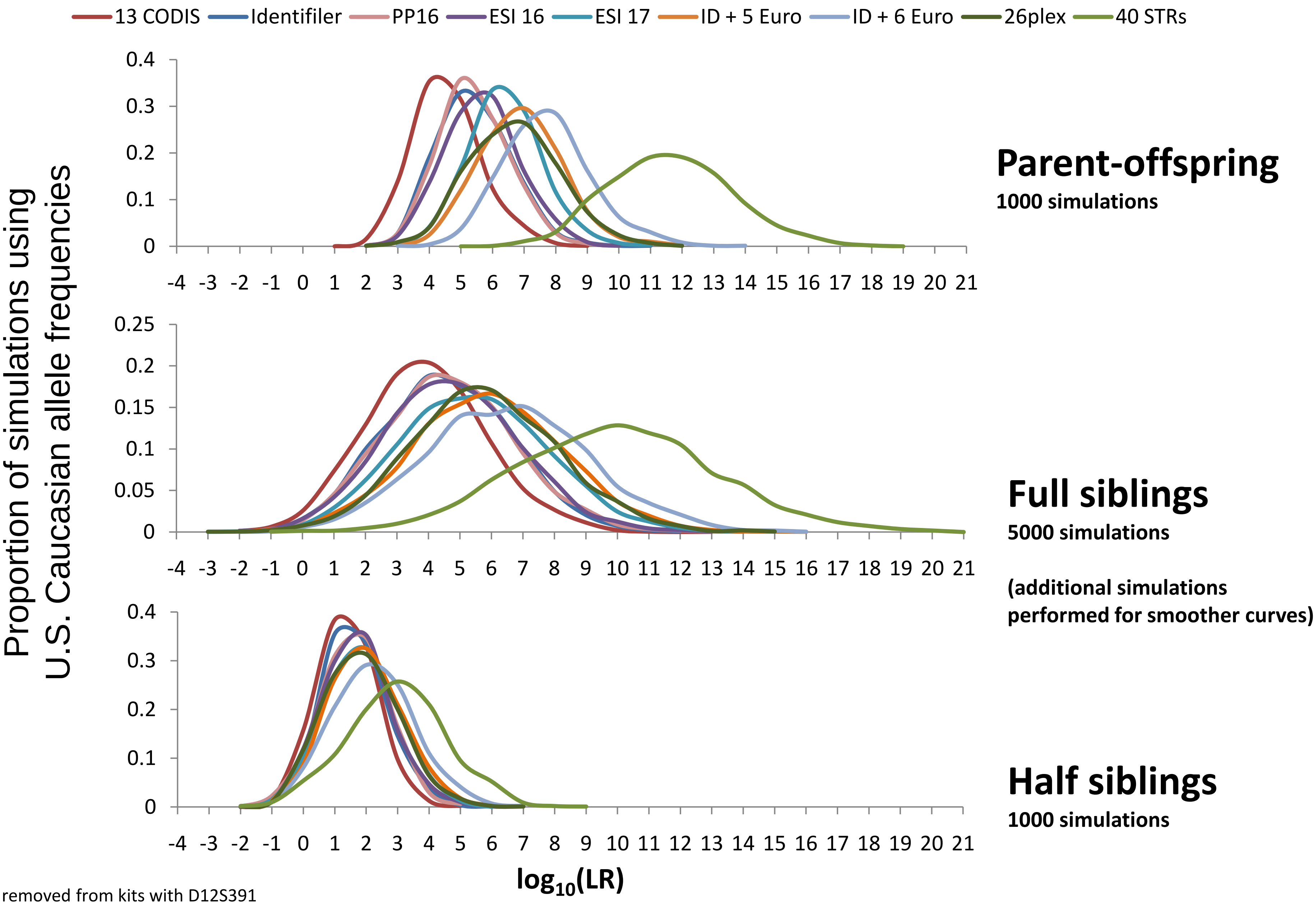
(additional simulations performed for smoother curves)

Half siblings

1000 simulations

*** Identifiler-vWA + 5 Euro + SE33**

Likelihood Ratio Distributions using Different Sets of STR Loci



Parent-offspring

1000 simulations

Full siblings

5000 simulations

(additional simulations performed for smoother curves)

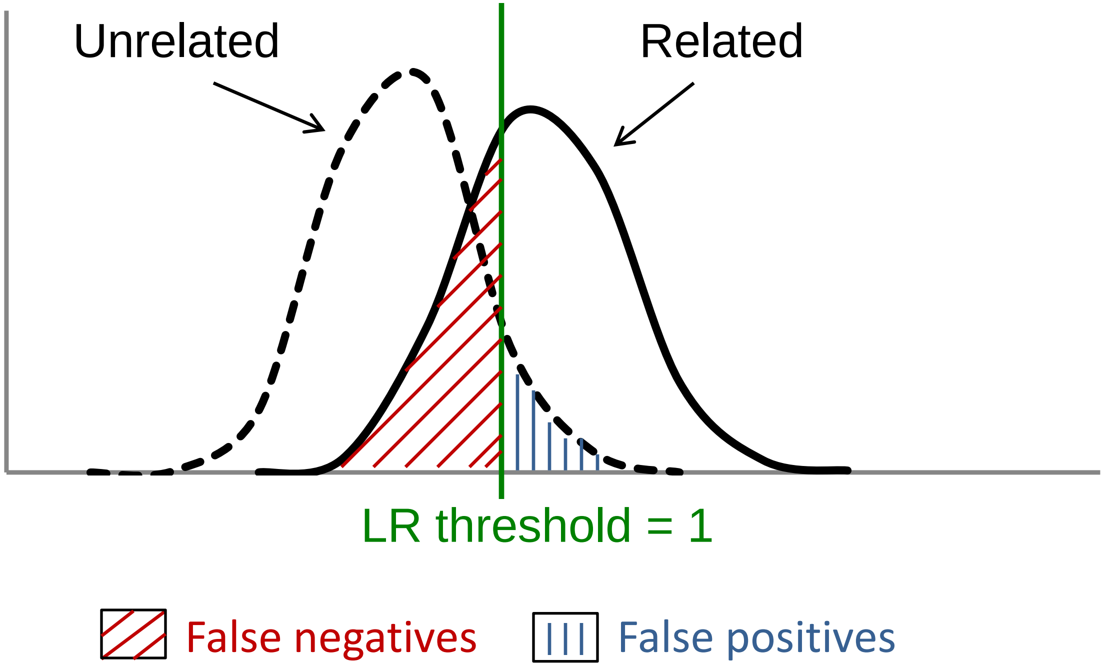
Half siblings

1000 simulations

vWA removed from kits with D12S391

Overlap of Likelihood Ratio Distributions

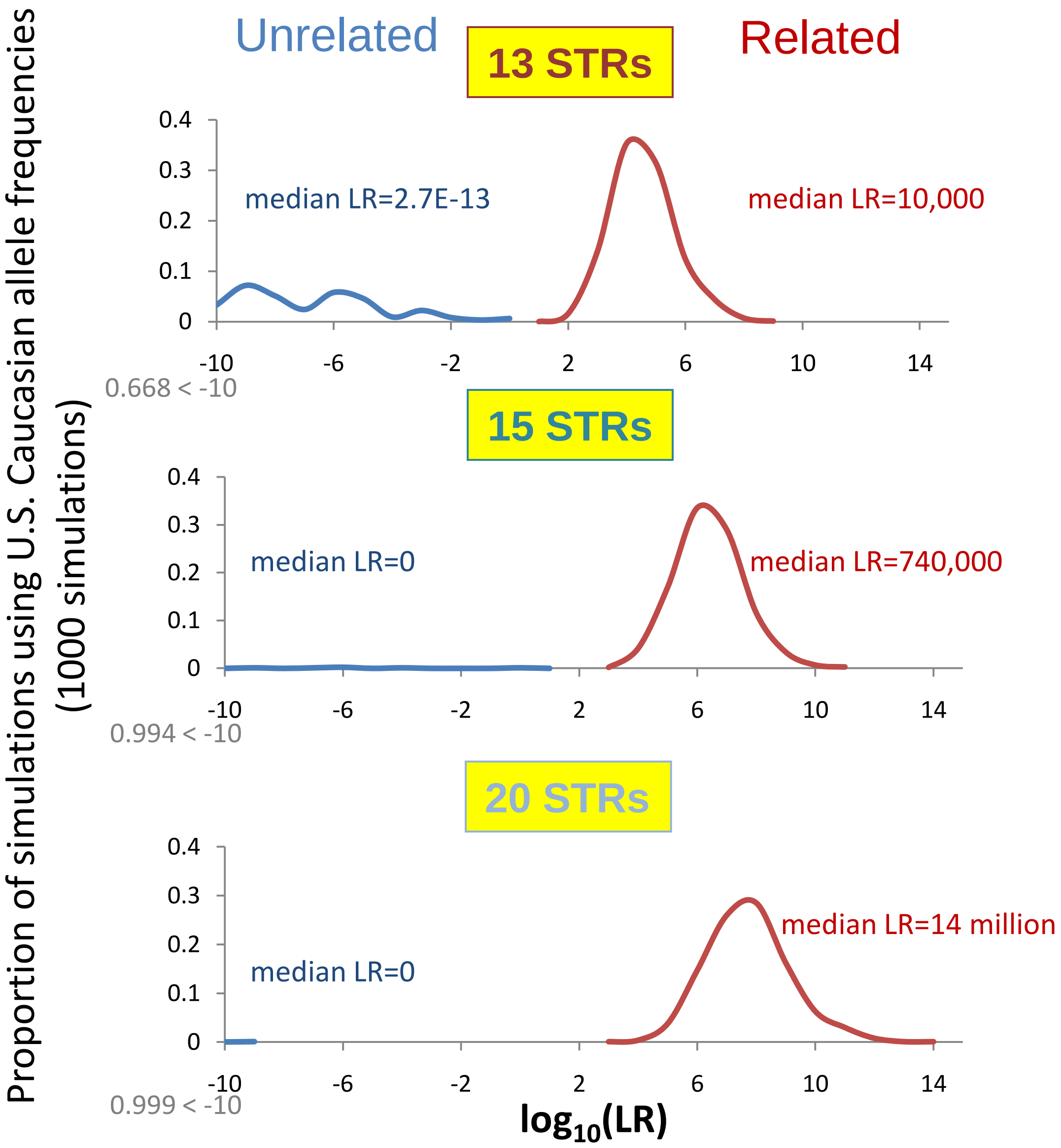
By the definition of a LR:
LR > 1 supports the numerator
LR < 1 supports the denominator



Probabilistic nature of relationship inference

- Incorrect relationships may be suggested
  - True relatives may appear unrelated
  - Unrelated individuals may appear related
- When distributions are pulled apart, the area of overlap is reduced
  - Goal is to reduce the area of overlap to zero
  - Can this be done with additional STR markers?

LR Distributions for Parent-Offspring with Different Numbers of STR Loci



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

For Pairwise Comparisons

No overlap for parent-offspring simulations with 13 STR loci

Additional loci separate distributions even more

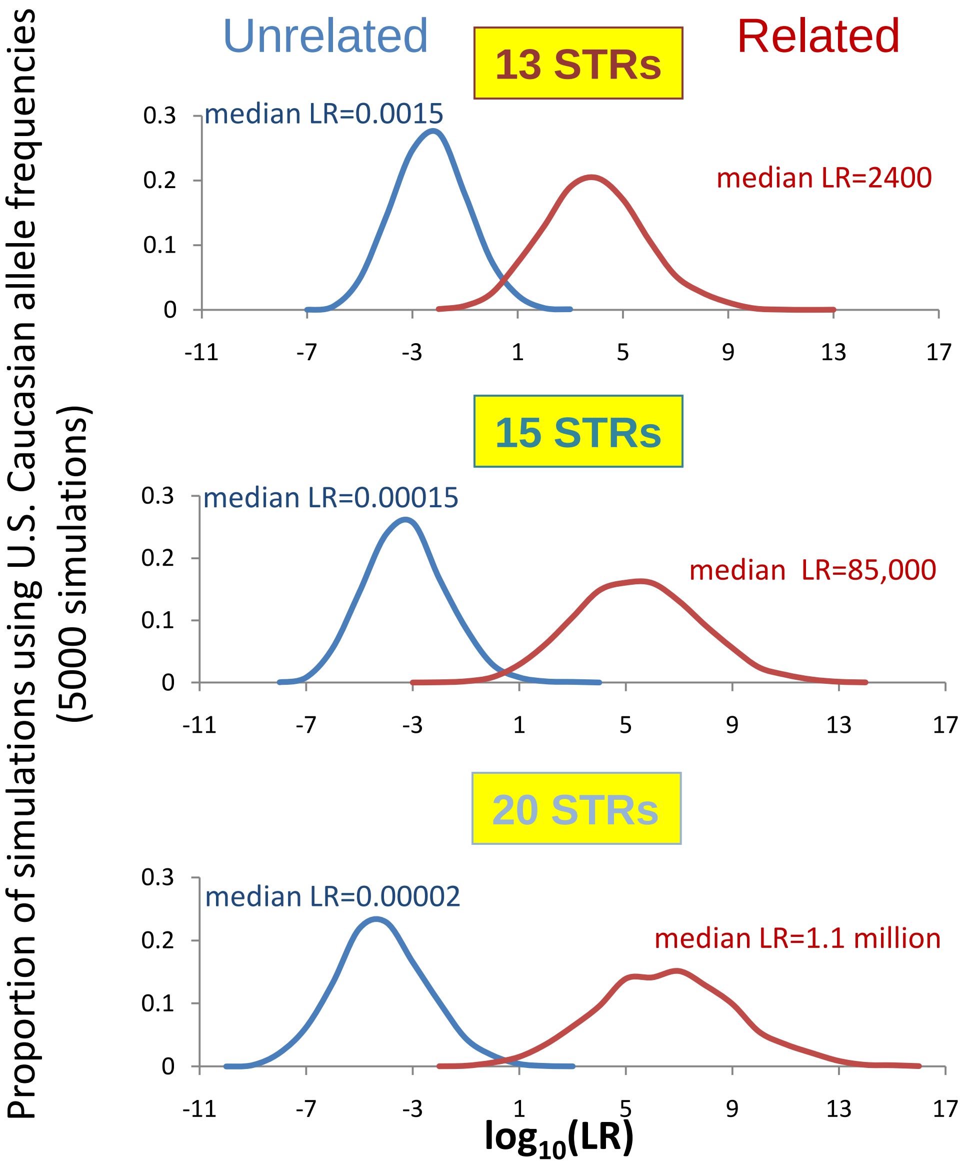
After Large Database Search

False positives/negatives are likely for parent-offspring with 13 STR loci

Fortuitous matches would be reduced with 20 STR loci

CODIS **ESI/ESX 17-vWA** **Identifiler-vWA + 5 Euro + SE33**

LR Distributions for Full Siblings with Different Numbers of STR Loci



Additional STR loci reduce overlap

Range of related and unrelated distributions is wider with 15 loci vs. 13 loci

20 loci increase LR values for true FS and decrease LR values for unrelated pairs

If LR threshold is 1, false positive rate =

- 0.03 with 13 loci
- 0.01 with 15 loci
- <0.01 with 20 loci

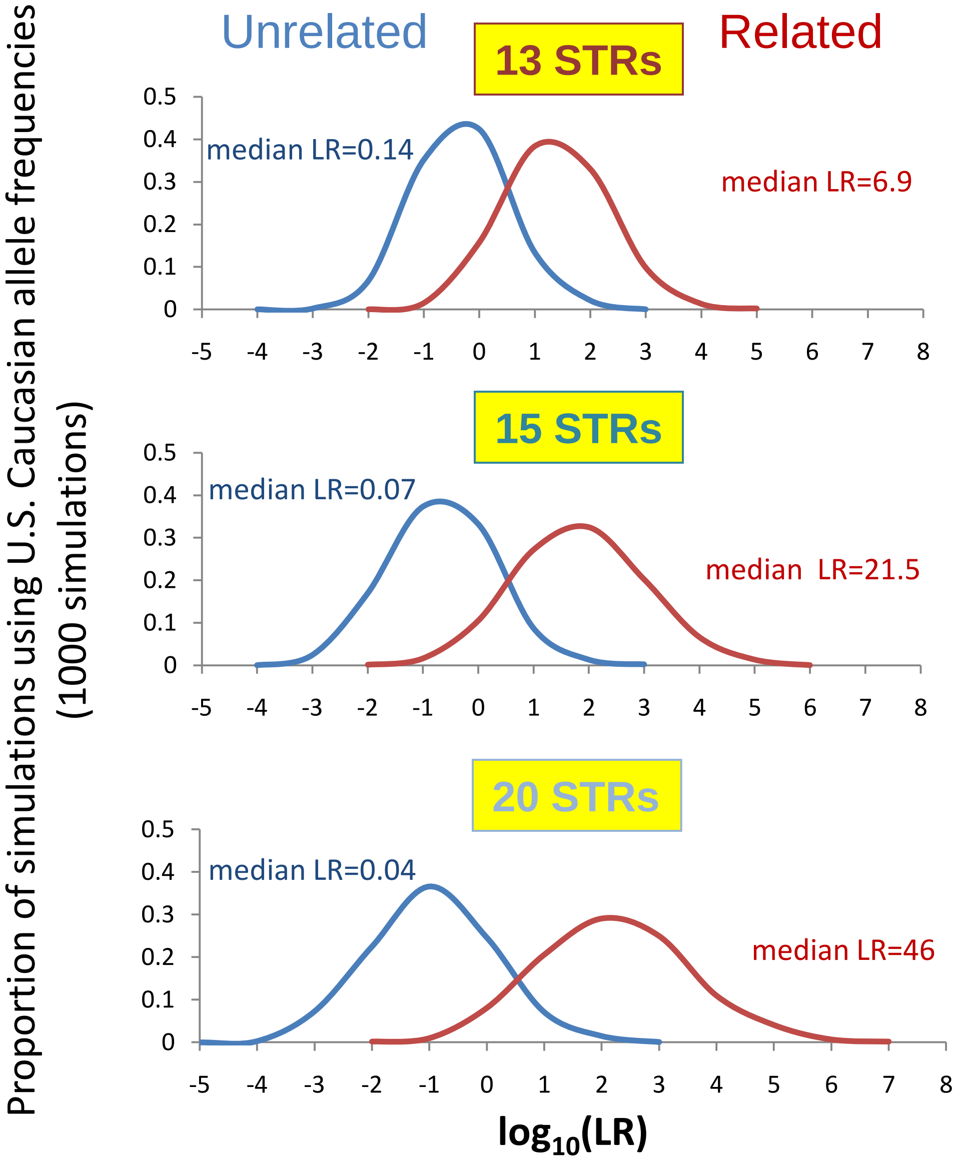
If LR threshold is 1, false negative rate =

- 0.03 with 13 loci
- 0.01 with 15 loci
- <0.01 with 20 loci

Fortuitous kinship matches from database search would be reduced with 20 loci

CODIS ESI/ESX 17-vWA Identifiler-vWA + 5 Euro + SE33

LR Distributions for Half Siblings with Different Numbers of STR Loci



Amount of overlap is not substantially reduced with additional loci

Large overlapping range of LR values for related and unrelated pairs

Difficult to define a LR threshold for kinship determination

If LR threshold is 1, false positive rate =

0.16 with 13 loci
0.10 with 15 loci
0.09 with 20 loci

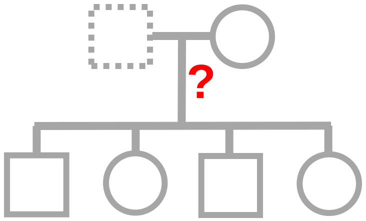
If LR threshold is 1, false negative rate =

0.17 with 13 loci
0.12 with 15 loci
0.09 with 20 loci

CODIS **ESI/ESX 17-vWA** **Identifiler-vWA + 5 Euro + SE33**

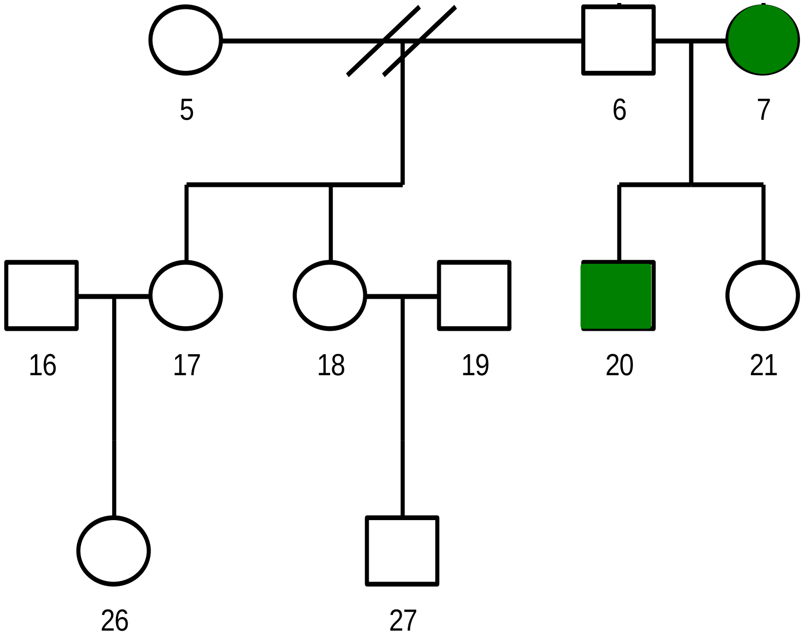
Reference Family Data to Assist Validation of Kinship Analysis

- NIST Standard Reference Family Data
  - Currently being developed to aid validation of kinship analysis algorithms, software, and loci selection
- Genetic data from known pedigrees
  - Use genotypes with known inheritance
  - Make direct comparisons between likelihood ratio values from algebraic and software calculations
  - Test algorithms for mutation, rare alleles, incest
  - Illustrate benefit/limitations of different loci to detect relationships



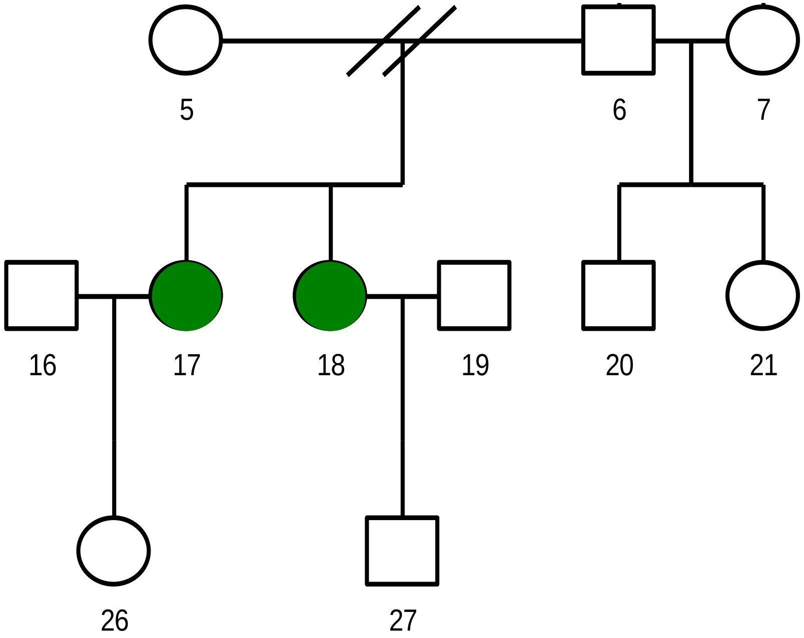
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
Parent-offspring (7 vs 20)	3.2E+04	7.4E+07	8.5E+08	8.0E+09



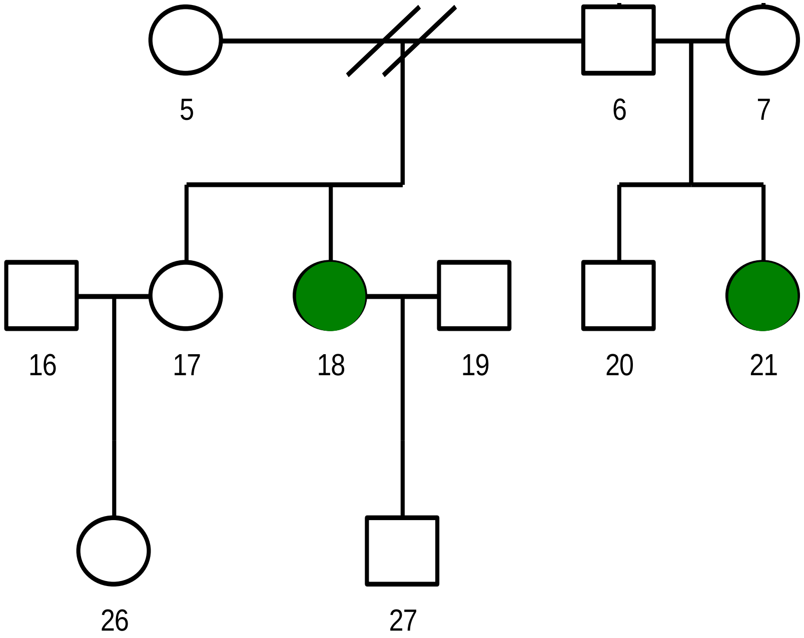
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
Full siblings (17 vs 18)	132	268	67	7143



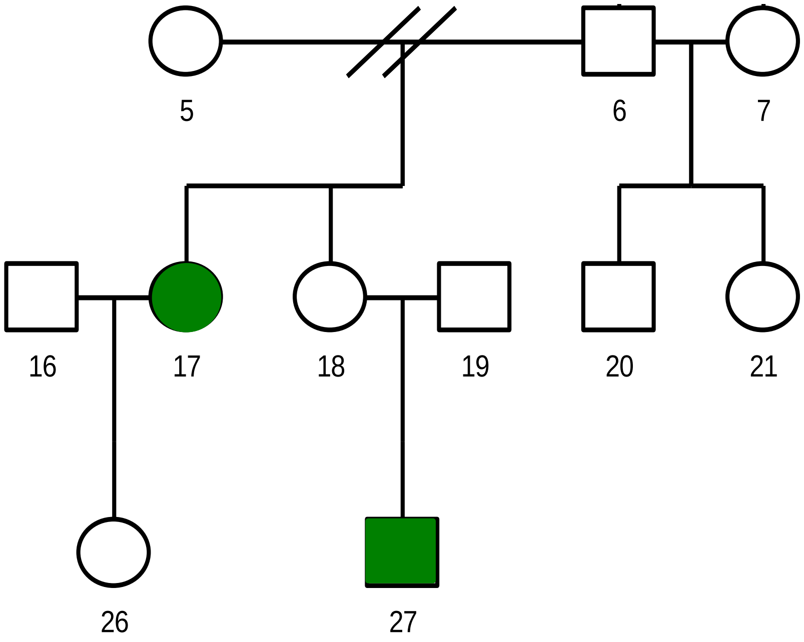
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	Likelihood Ratio Values using Different Sets of Loci			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
Half siblings (18 vs 21)	0.9	1.5	5.1	2.0



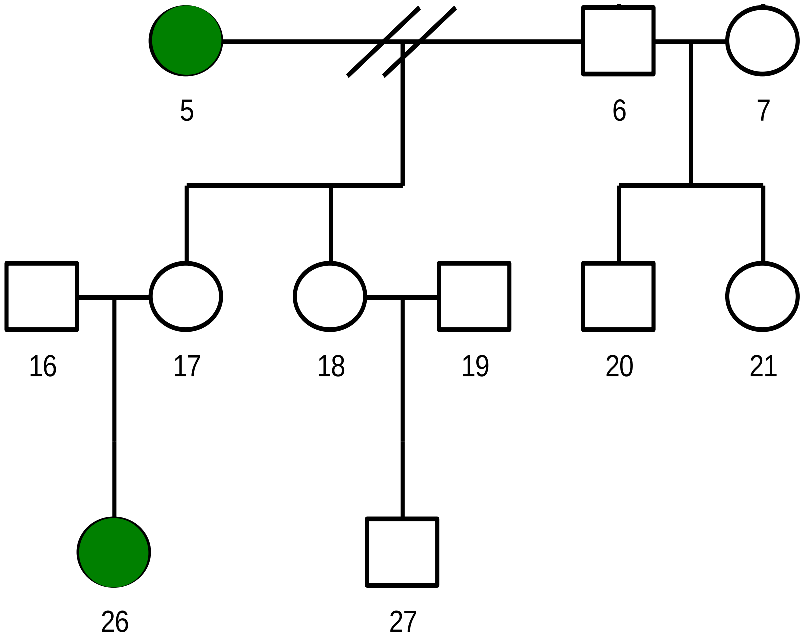
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
Aunt-niece (17 vs 27)	1.1	0.9	0.4	1.9



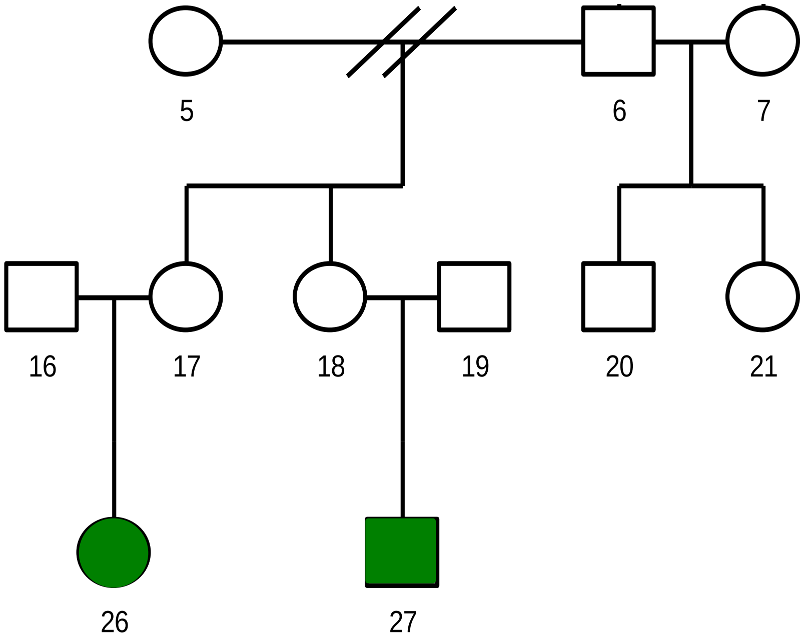
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
Grandmother-granddaughter (5 vs 26)	10.1	0.1	0.2	4.4



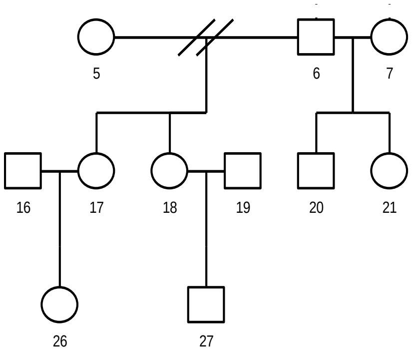
Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs **(CODIS)**	**14 STRs** **(ESI/ESX 16-vWA)**	**15 STRs** **(ESI/ESX 17-vWA)**	**20 STRs** **(Identifiler-vWA + 5 Euro + SE33)**
First cousins (26 vs 27)	1.7	1.1	0.9	2.0



Specific Example of LR Values for Different Relationships using Collected Pedigree Data

Relationship Tested	**Likelihood Ratio Values using Different Sets of Loci**			
	13 STRs (CODIS)	**14 STRs** (ESI/ESX 16-vWA)	**15 STRs** (ESI/ESX 17-vWA)	**20 STRs** (Identifiler-vWA + 5 Euro + SE33)
Parent-offspring (7 vs 20)	3.2E+04	7.4E+07	8.5E+08	8.0E+09
Full siblings (17 vs 18)	132	268	67	7143
Half siblings (18 vs 21)	0.9	1.5	5.1	2.0
Aunt-niece (17 vs 27)	1.1	0.9	0.4	1.9
Grandmother-granddaughter (5 vs 26)	10.1	0.1	0.2	4.4
First cousins (26 vs 27)	1.7	1.1	0.9	2.0



Additional loci improve LR values for 1st degree relatives (parent-offspring, full siblings)

LR values are not greatly improved for more distant relatives for which the probability of allele sharing is reduced

Summary of 20 STR Loci

(Identifiler-vWA + 5 Euro + SE33)

- Robust set of loci for close relatives
  - Parent-offspring and full siblings
  - Single multiplex PCR would require 6-dye chemistry and possible primer redesign
    - Depending on manufacturer and primer design, SE33 may not fit

- SE33 is a powerful locus for determining first degree relatives
  - Large amount of allelic variation
  - High mutation rate

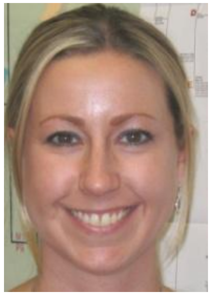
- More distant relationships remain difficult to identify with 20 STR loci
  - Half siblings, uncle-nephew, grandparent-grandchild, cousins, etc.
  - Nearly 10% false positive and false negative rates with LR threshold = 1
  - Lineage markers, more individuals, or non-genetic information (use Bayesian statistics) can increase confidence in a kinship test

NIST
Applied
Genetics

Acknowledgments

Applied Genetics
Group Leader

John
Butler



Becky
Hill

DNA Biometrics
Project Leader

Peter
Vallone

Kristen
Lewis
O'Connor

Erica
Butts

http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm
kristen.oconnor@nist.gov

<u>Funding</u>
NRC – Postdoctoral fellowship support for Kristen O'Connor
FBI – Application of DNA Typing as a Biometric Tool
NIJ – Forensic DNA Standards, Research, and Training