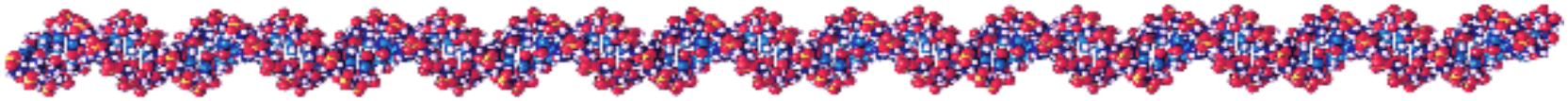
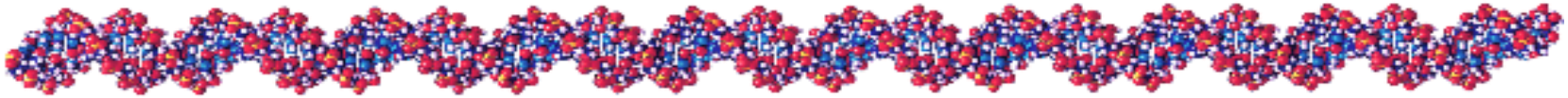


The Role of DNA in Kinship Analysis



Kristen E. Lewis, Ph.D.

National Institute of Standards and Technology
Biochemical Science Division
Applied Genetics Group – DNA Biometrics Project

Lockheed Martin BEACON Lecture
April 14, 2010

Outline

- Forensic DNA analysis / DNA biometrics
- Complex kinship testing
- Kinship analysis software

Forensic DNA Analysis

DNA Biometrics

Steps in Forensic DNA Analysis

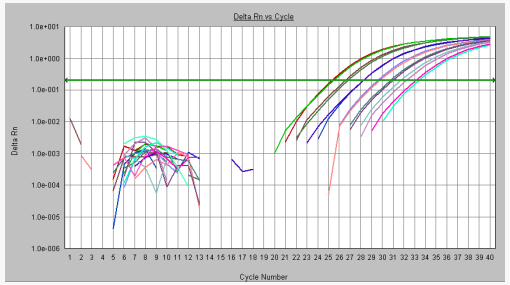
Usually 1-2 day process (a minimum of ~8 hours)

Blood Stain Buccal swab

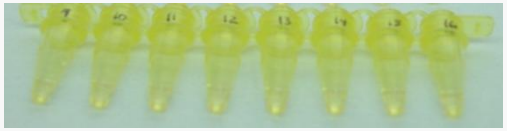
Sample Collection & Storage

Biology

DNA Extraction

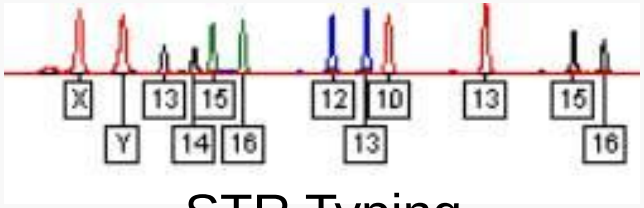


DNA Quantitation



Multiplex PCR Amplification

DNA separation and sizing



STR Typing

Interpretation of Results

Technology

Genetics

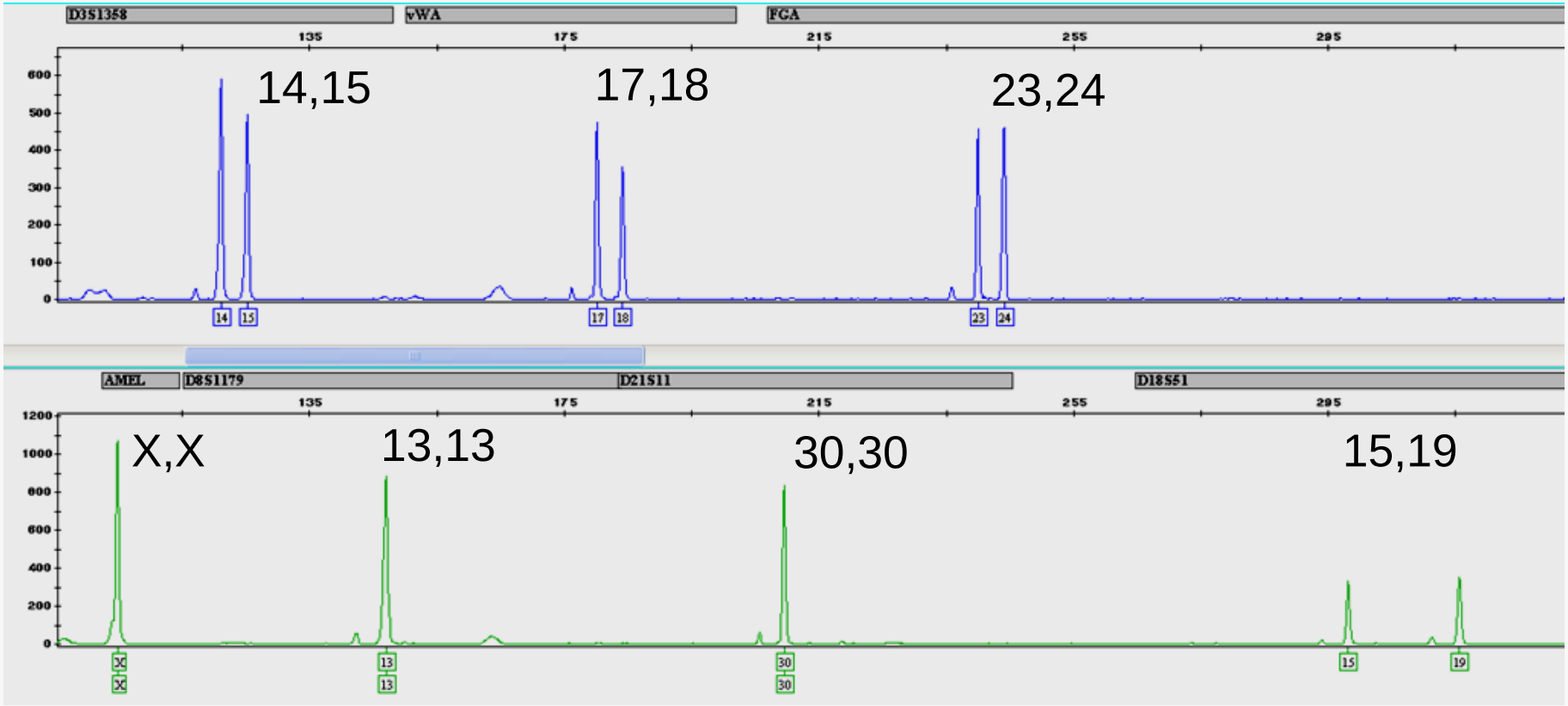
Statistics Calculated

DNA database search
Paternity test
Reference sample

Applied Use of Information

Forensic DNA Typing

Example of 6 markers + sex-typing marker



A common set of 13 Short Tandem Repeat (STR) markers (“CODIS loci”) are genotyped to search state and national databases

DNA Typing as a Biometric

Advantages

- High level of accuracy
  (Gold Standard)
- Solid foundation of forensic DNA testing
  (pop stats, molecular biology, protocols, training, court acceptance, education)
- **Kinship analysis (unique to DNA)**
- Potential use for:
  - Phenotype (traits; eye/hair color)
  - Ancestry

Challenges

- Expensive
- **Time consuming**
- Sample collection (invasive, stability issues)
- Technical expertise required for analysis
- Low level template, mixtures, PCR inhibition
- Policy/Privacy/Ethical issues

Interest in Rapid DNA Typing

- DoD (field testing, rapid intelligence, mass fatalities)
- DHS (kinship determination, border security, immigration)
- DoJ (law enforcement, initial information)
- Industry (security, authentication)

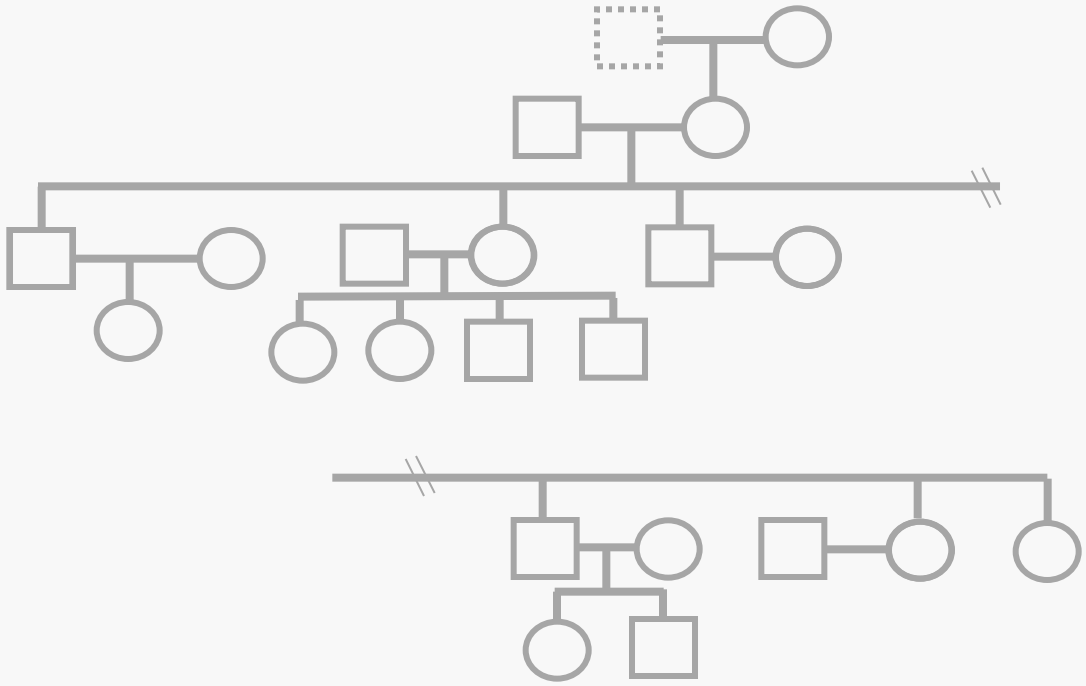
Rapid DNA Typing Systems under Development

- Systems are currently under development and are not yet commercially available

- Network Biosystems (Woburn, MA) http://www.netbio.com
- MicroLab Diagnostics and Lockheed Martin (Charlottesville,VA) http://www.microlabdiagnostics.com
- Microchip Biotechnologies, Inc. (Pleasanton, CA) http://www.microchipbiotech.com
- Forensic Science Service (UK) http://www.forensic.gov.uk/

Use of DNA as a Biometric Tool American Academy of Forensic Science, Feb 22, 2010, Seattle, WA
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

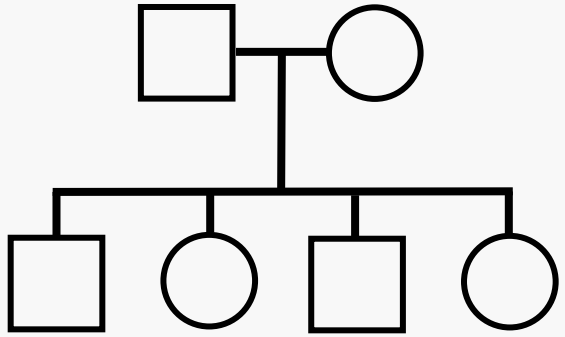
Complex Kinship Analysis



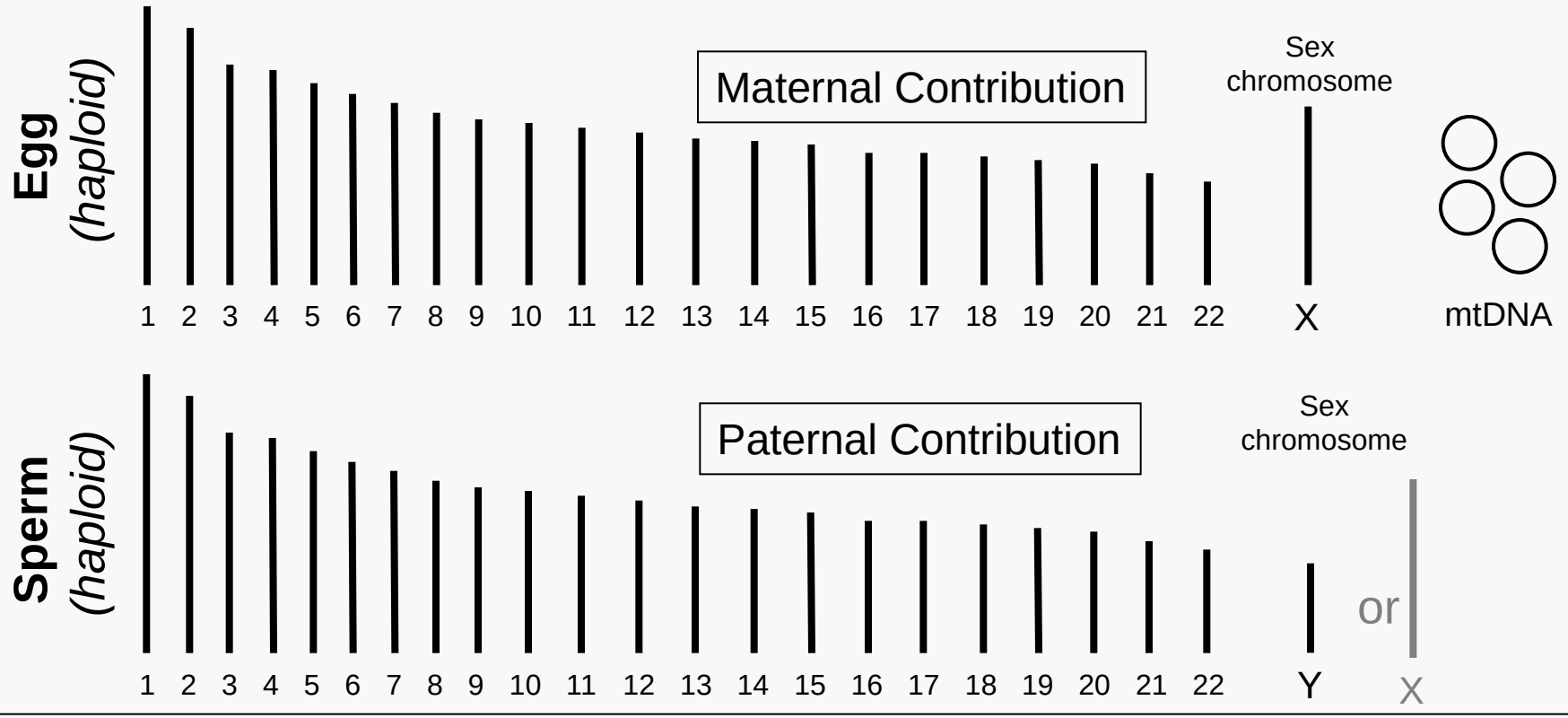
Using DNA to Detect Genetic Relationships

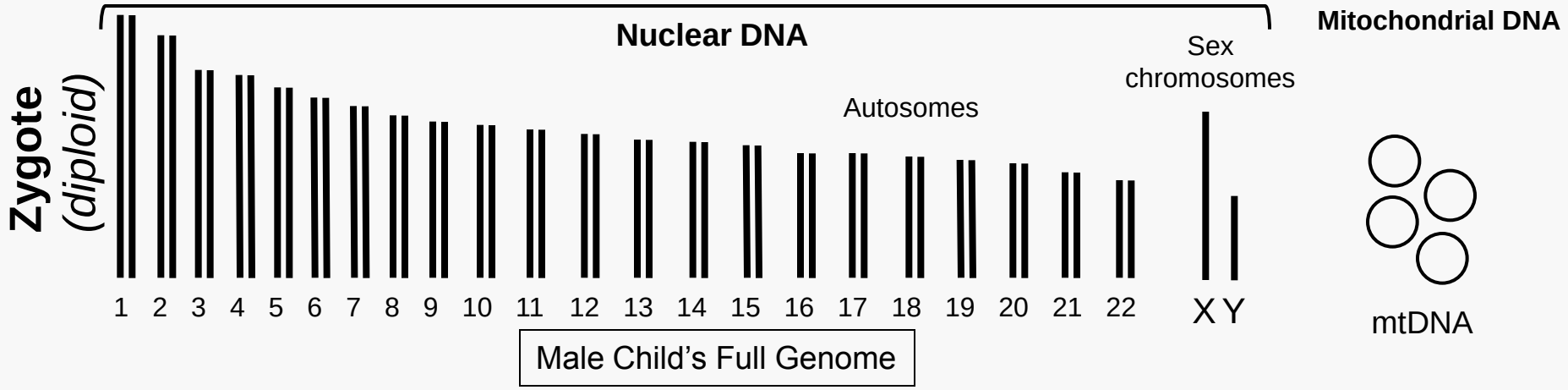
- DNA profiles can be used to evaluate the probability of a specific kinship relationship

- Various situations
  - Paternity (civil and criminal)
  - Familial searching
  - Mass disasters
  - Unidentified human remains
  - Inheritance
  - Immigration
  - Military intelligence



Human Genome and Inheritance

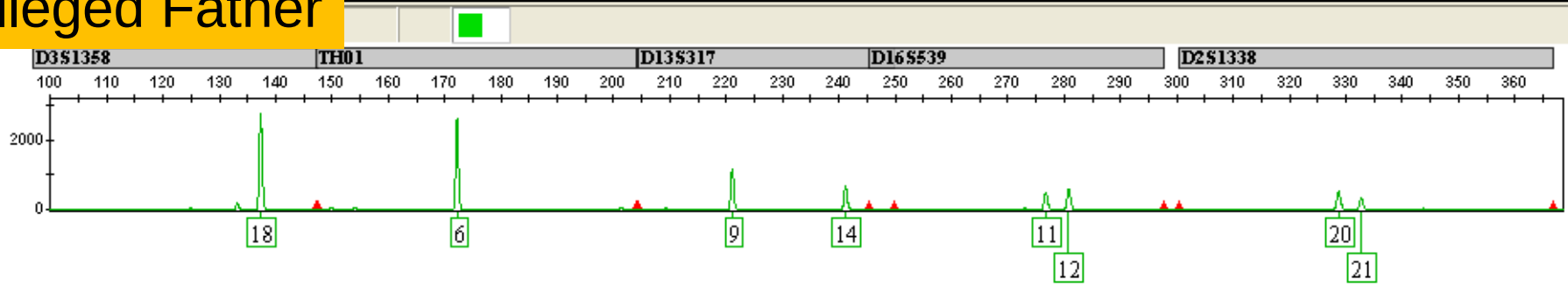




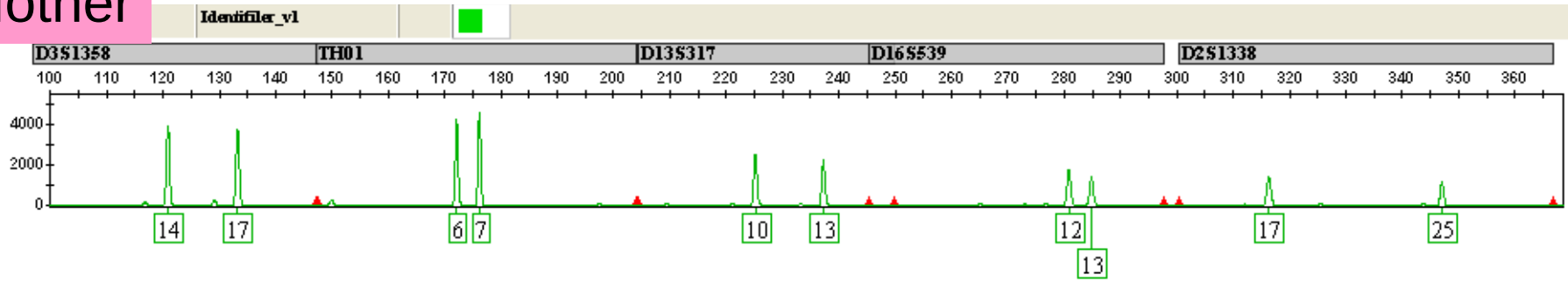
John M. Butler (2009) *Fundamentals of Forensic DNA Typing*, Figure 2.7

Autosomal Paternity Example

Alleged Father

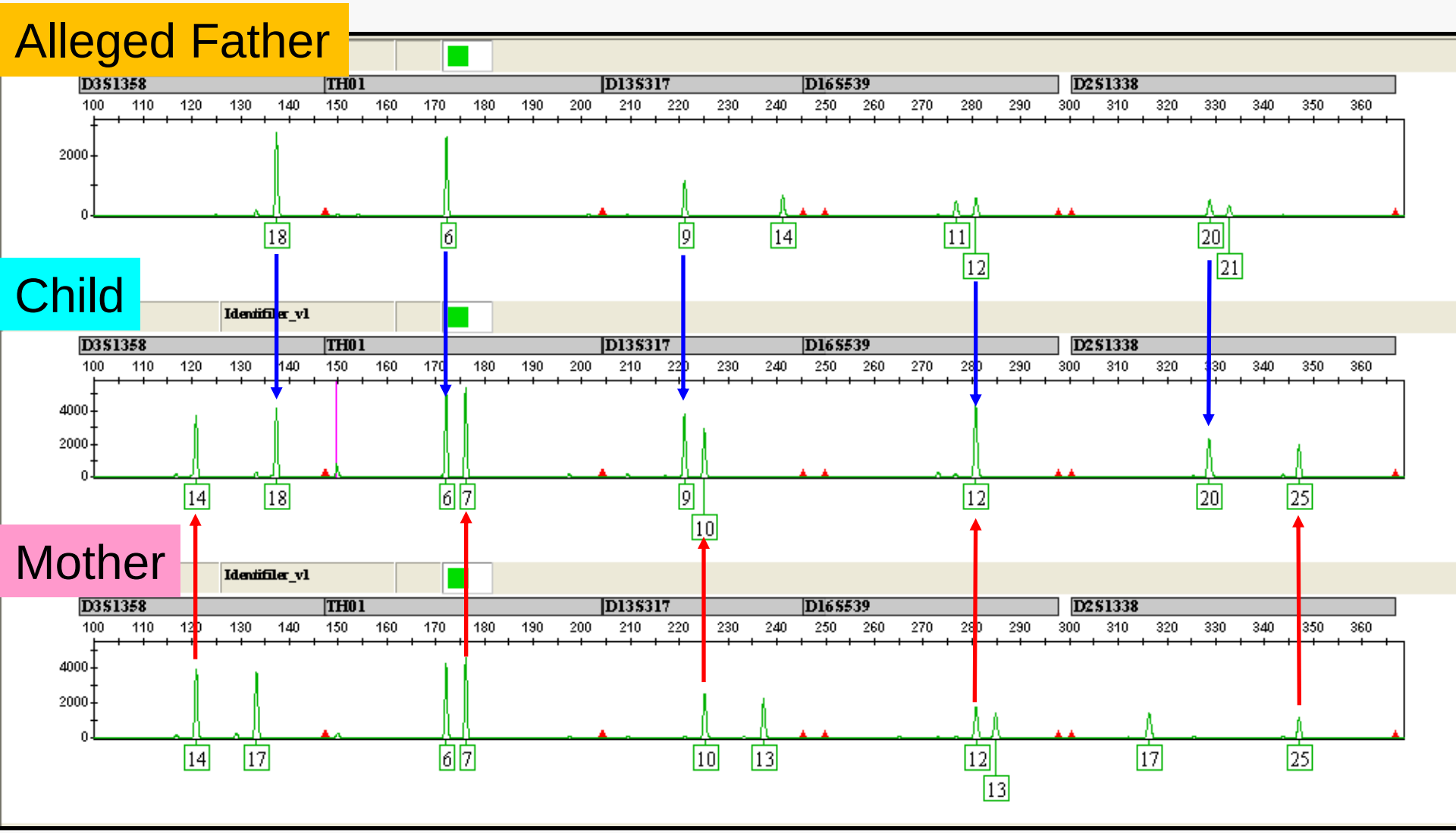


Mother



Focusing on 5 markers…

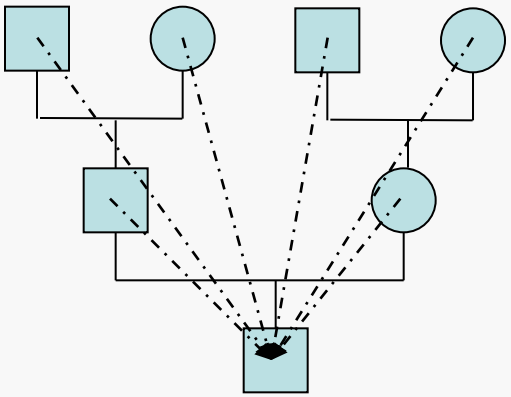
Autosomal Paternity Example



Focusing on 5 STR markers…

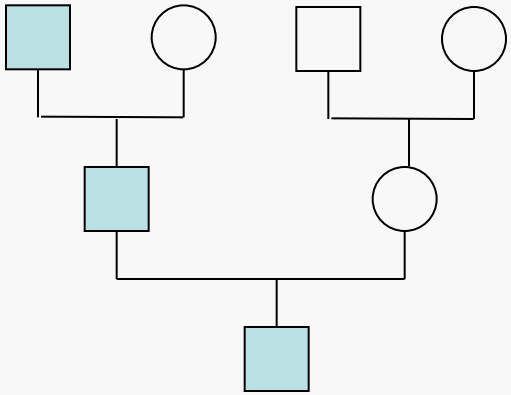
Different Inheritance Patterns

CODIS STR Loci

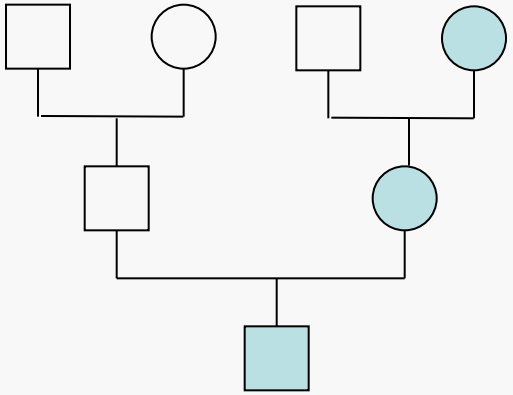


Autosomal
(passed on in part, from all ancestors)

Lineage Markers

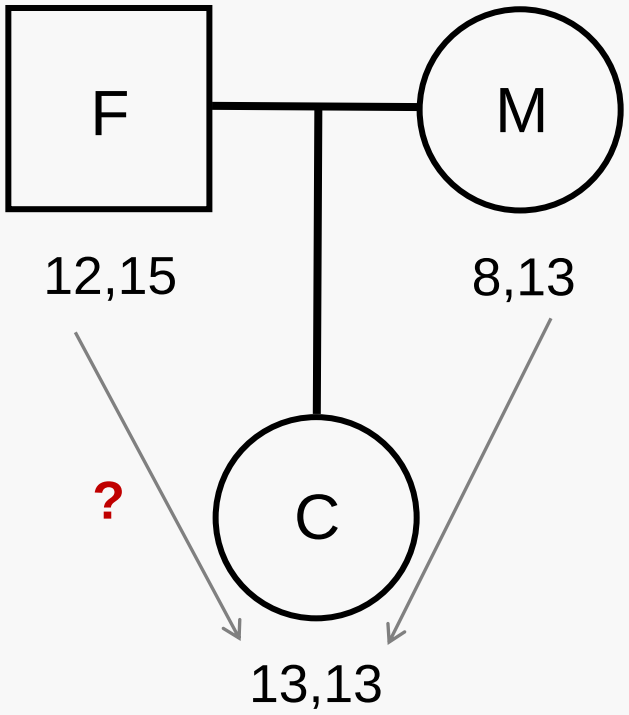


Y-Chromosome
(passed on complete, but only by sons)

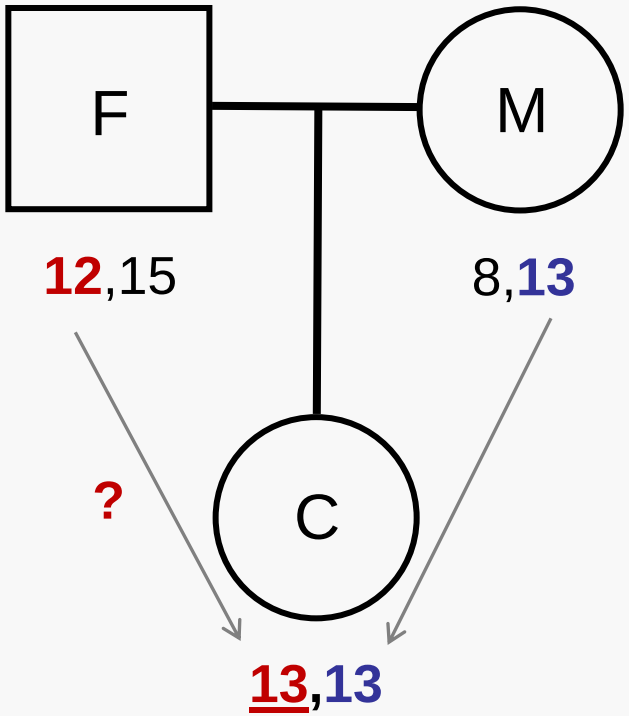


Mitochondrial
(passed on complete, but only by daughters)

Mutation of STR Markers



Mutation of STR Markers



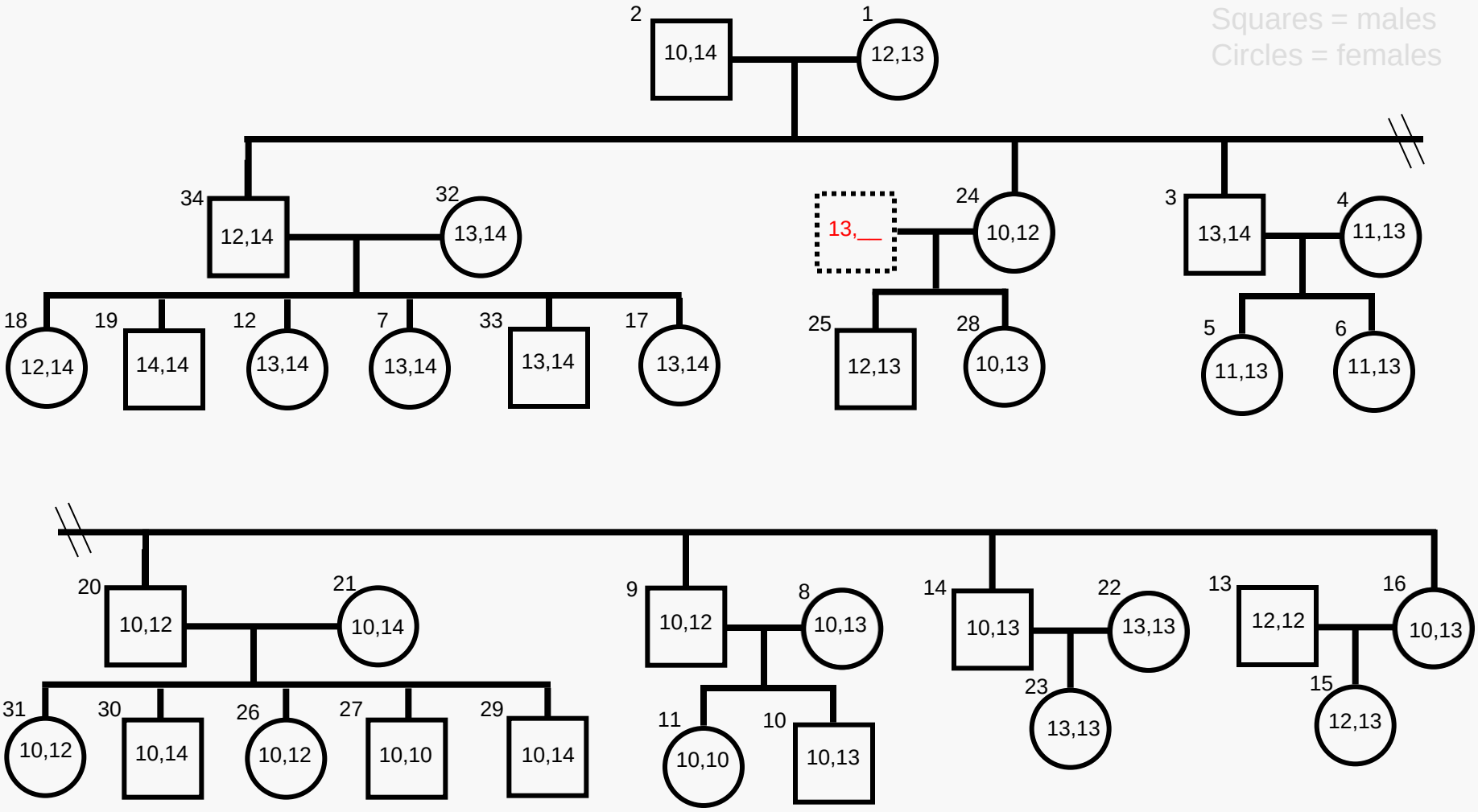
- Typically observe 1-step mutations (99%)
- Gain or loss of repeat unit possible
- Average STR mutation frequency = 0.001
- Must factor in mutation for kinship analyses

Extended Family Samples

- 6 sets of family samples (3 - 4 generations)
- N = 165 (total samples)
- Loci examined
  - 43 autosomal loci (13 CODIS + 2, 25 NIST loci, PowerPlex ESI 17, NGM loci)
  - 17 Y-chromosomal loci (Yfiler loci)
  - 15 X-STRs (AFDIL collaboration)
  - Mitochondrial control region (to be typed)

- These samples are being used to illustrate the value (or limitations) of current and additional marker systems

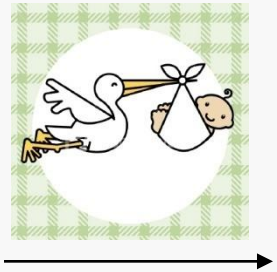
Extended Family Samples



Family Pedigree with STR Marker **D8S1179**

Examine Basic Allele Sharing

	Marker	Father		Mother	
1	CSF1PO	10	10	10	12
2	TPOX	8	8	8	10
3	TH01	6	6	6	9
4	vWA	17	18	17	20
5	D16S539	11	13	8	9
6	D7S820	9	9	8	12
7	D13S317	11	14	8	12
8	D5S818	12	13	11	13
9	FGA	21	22	21	25
10	D8S1179	12	14	13	14
11	D18S51	14	16	14	17
12	D21S11	28	30	31	32.2
13	D3S1358	16	17	17	17
14	D2S1338	22	23	23	25
15	D19S433	12	14	14	14



	Marker	Father		Child		Mother	
1	CSF1PO	10	10	10	10	10	12
2	TPOX	8	8	8	8	8	10
3	TH01	6	6	6	6	6	9
4	vWA	17	18	17	17	17	20
5	D16S539	11	13	9	13	8	9
6	D7S820	9	9	8	9	8	12
7	D13S317	11	14	8	14	8	12
8	D5S818	12	13	11	13	11	13
9	FGA	21	22	21	25	21	25
10	D8S1179	12	14	14	14	13	14
11	D18S51	14	16	14	17	14	17
12	D21S11	28	30	28	31	31	32.2
13	D3S1358	16	17	17	17	17	17
14	D2S1338	22	23	23	23	23	25
15	D19S433	12	14	12	14	14	14

The child inherits an allele from each parent

Siblings and Avuncular

	Marker	Brother		Sister	
1	CSF1PO	10	10	10	12
2	TPOX	8	8	8	10
3	TH01	6	6	6	9
4	vWA	17	17	17	20
5	D16S539	9	13	9	11
6	D7S820	8	9	9	12
7	D13S317	8	14	11	12
8	D5S818	11	13	11	12
9	FGA	21	25	21	25
10	D8S1179	14	14	13	14
11	D18S51	14	17	14	14
12	D21S11	28	31	30	32
13	D3S1358	17	17	16	17
14	D2S1338	23	23	23	25
15	D19S433	12	14	14	14

	Marker	Uncle		Nephew	
1	CSF1PO	10	10	11	12
2	TPOX	8	8	8	10
3	TH01	6	6	6	9.3
4	vWA	17	18	16	17
5	D16S539	11	13	11	13
6	D7S820	9	9	9	11
7	D13S317	11	14	11	12
8	D5S818	12	13	12	12
9	FGA	21	22	20	24
10	D8S1179	12	14	10	13
11	D18S51	14	16	13	13
12	D21S11	28	30	27	31
13	D3S1358	16	17	16	18
14	D2S1338	22	23	18	22
15	D19S433	12	14	14	14

Evaluating Relatedness

- Examining or counting shared alleles is qualitative

- Likelihood ratio can be used to help quantify the information

$$LR = \frac{\text{Probability of data under one hypothesis}}{\text{Probability of data under alternative hypothesis}}$$

Likelihood Ratio Equations

Motherless Paternity Case

Genotype Combinations at One Marker		**Likelihood Ratio**
Child's Genotype	Alleged Father's Genotype	
AA	AA	$1/p_A$
AA	BB	0
AA	AB	$1/2p_A$
AA	BC	0
AB	AB	$(p_A + p_B)/4p_Ap_B$
AB	AC	$1/4p_A$
AB	CD	0

Multiply LR across all loci

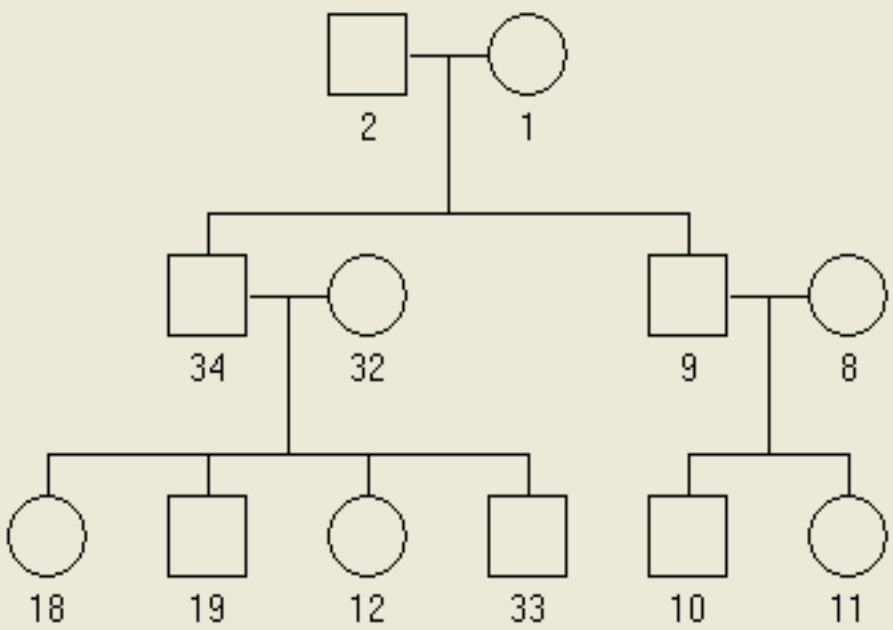
$$LR = \frac{\text{Probability of data if the alleged father is the true father}}{\text{Probability of data if an unrelated man is the true father}}$$

Allele Sharing Probabilities

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

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

Complex Kinship Testing



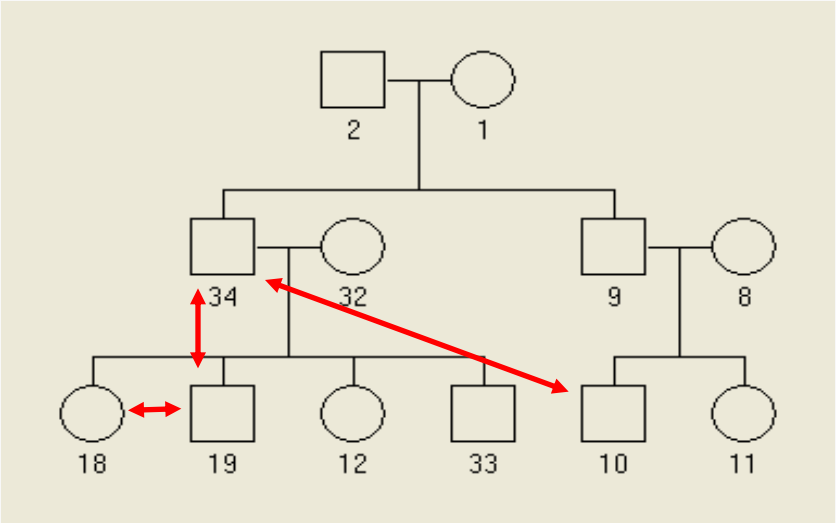
The statistical power for complex kinship testing significantly decreases compared to one-to-one matching

Requirements:

- Genotypes of individuals being tested
- Allele frequencies for the loci involved in the testing
- **A Hypothesis!**

- Basic statistical equations are known
- Difficult to identify distant relationships
- Discriminatory power comes from multiple family members and the use of informative markers

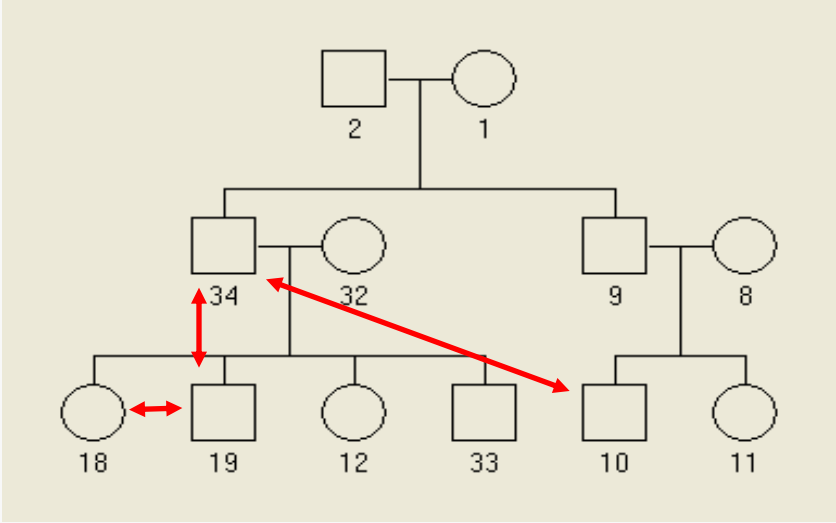
Likelihood Ratios with 15 Loci



Comparison	LR for 34 & 19	LR for 18 & 19	LR for 34 & 10
Parent-Child	1.28E+06		
Full Siblings		2.76E+07	
Half Siblings			
Cousins			
Uncle-Nephew			6.65E-01
Grandparent-Grandchild			
	Parent/ Child	Full Sibs	Uncle/ Nephew

LR calculations were performed with GeneMarker® HIDv1.90

Likelihood Ratios with 15 Loci



Comparison	LR for 34 & 19	LR for 18 & 19	LR for 34 & 10
Parent-Child	1.28E+06	9.08E+05	0.00E+00
Full Siblings	3.22E+04	2.76E+07	6.07E-03
Half Siblings	7.38E+03	4.89E+04	6.65E-01
Cousins	1.95E+02	8.96E+02	1.52E+00
Uncle-Nephew	7.38E+03	4.89E+04	6.65E-01
Grandparent-Grandchild	7.38E+03	4.89E+04	6.65E-01
	Parent/ Child	Full Sibs	Uncle/ Nephew

LR calculations were performed with GeneMarker® HIDv1.90

Benefit of Additional Loci

Likelihood Ratios with 40 Loci

	15	40	15	40	15	40
Comparison	**LR for 34 & 19**	**LR for 34 & 19**	**LR for 18 & 19**	**LR for 18 & 19**	**LR for 34 & 10**	**LR for 34 & 10**
Parent-Child	1.28E+06	6.68E+16	9.08E+05	0.00E+00	0.00E+00	0.00E+00
Full Siblings	3.22E+04	5.73E+12	2.76E+07	1.57E+19	6.07E-03	3.30E+03
Half Siblings	7.38E+03	8.63E+11	4.89E+04	4.99E+12	6.65E-01	8.98E+05
Cousins	1.95E+02	1.32E+08	8.96E+02	1.05E+09	1.52E+00	2.17E+04
Uncle-Nephew	7.38E+03	8.63E+11	4.89E+04	4.99E+12	**6.65E-01**	**8.98E+05**
Grandparent-Grandchild	7.38E+03	8.63E+11	4.89E+04	4.99E+12	6.65E-01	8.98E+05

Parent/ Child

Full Sibs

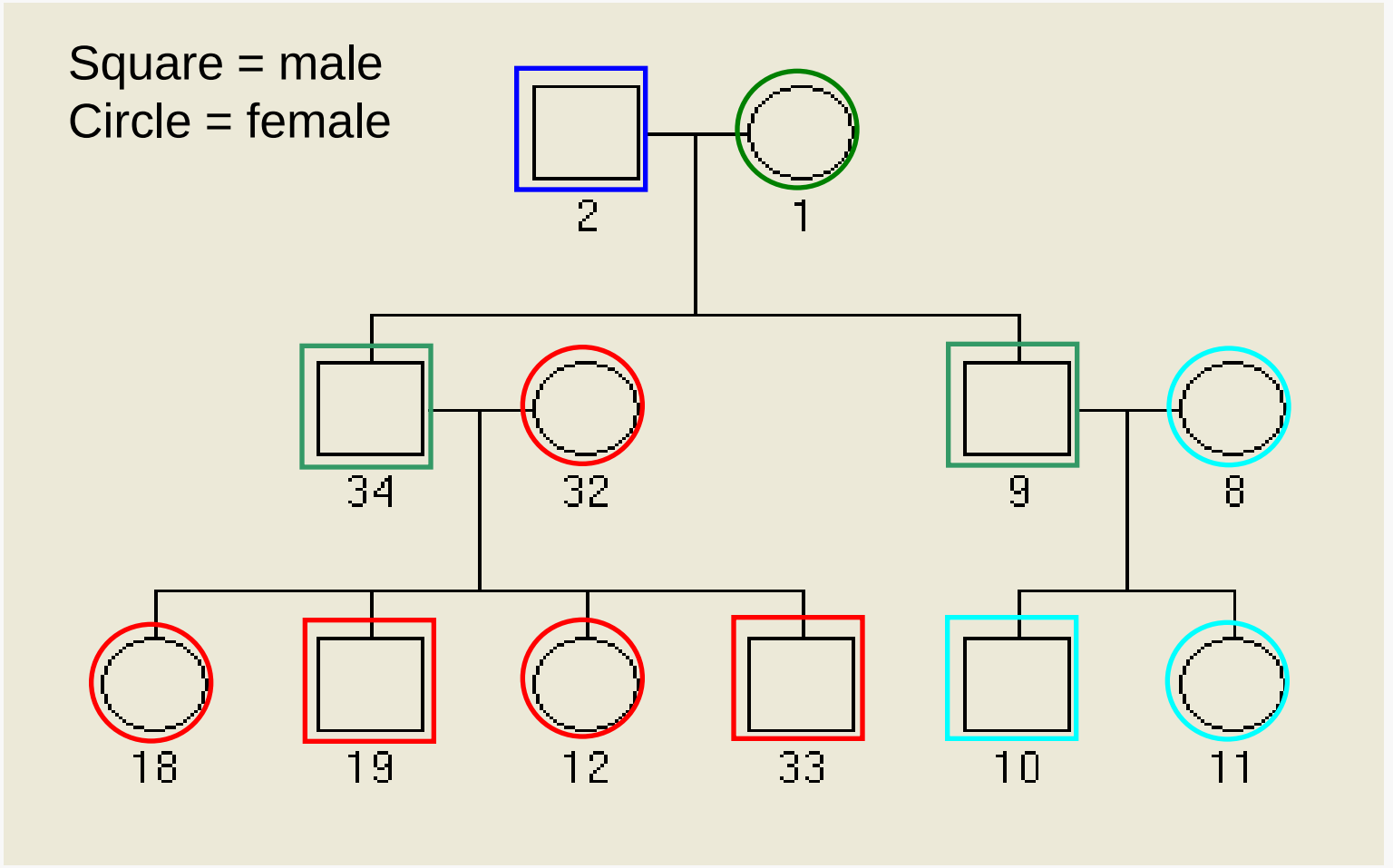
Uncle/ Nephew

LR calculations were performed with GeneMarker® HIDv1.90

15 STR loci typed with commercial Identifiler kit
25 STR loci typed with an in-house NIST assay

Hill, C.R., Butler, J.M., Vallone, P.M. (2009) A 26plex autosomal STR assay to aid human identity testing. J. Forensic Sci. 54(5): 1008-1015.

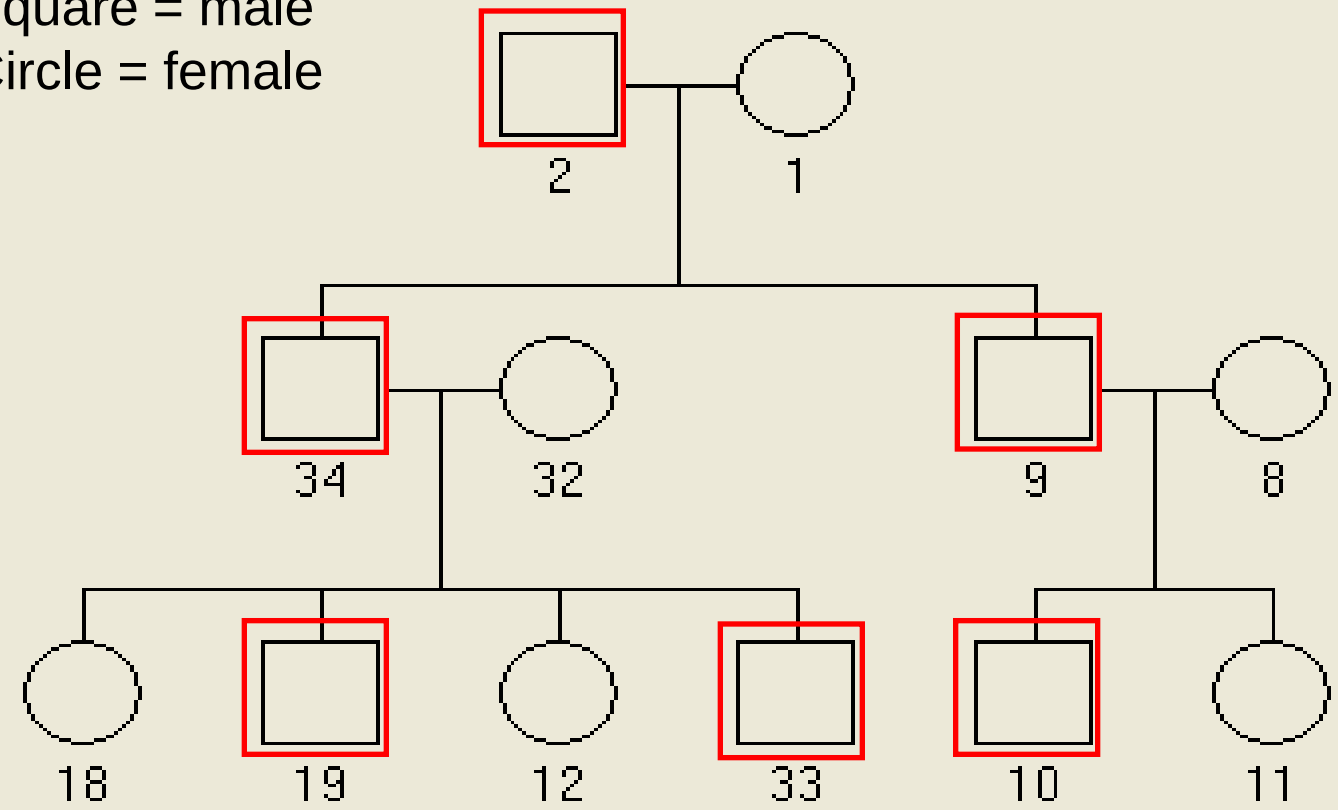
What would Mitochondrial sequencing tell you?



Mitochondrial DNA is maternally inherited

What would a Y-STR test tell you?

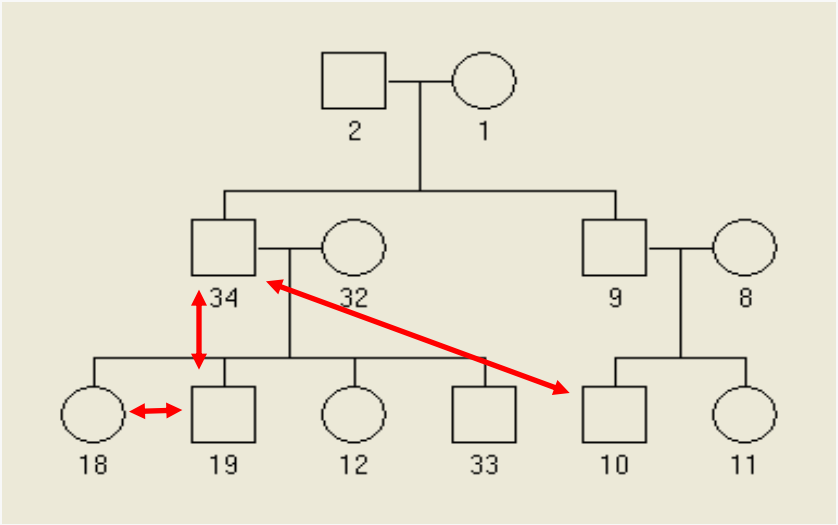
Square = male
Circle = female



Y STR	Allele
DYS456	17
DYS389I	13
DYS390	24
DYS389II	29
DYS458	18
DYS19	14
DYS385	11
DYS385	15
DYS393	13
DYS391	11
DYS439	13
DYS635	23
DYS392	13
Y_GATA_H4	12
DYS437	15
DYS438	12
DYS448	19

All males in this specific linage have the identical Y-STR profile

Likelihood Ratios with Y-STR Loci



Y STR	Allele
DYS456	17
DYS389I	13
DYS390	24
DYS389II	29
DYS458	18
DYS19	14
DYS385	11
DYS385	15
DYS393	13
DYS391	11
DYS439	13
DYS635	23
DYS392	13
Y_GATA_H4	12
DYS437	15
DYS438	12
DYS448	19

Y-STR profile frequency = 0.00009

For father-son (34 vs 19):
LR = 1 / 0.00009 = 1.1E+04
LR with 40 autoSTRs = 6.68E+16

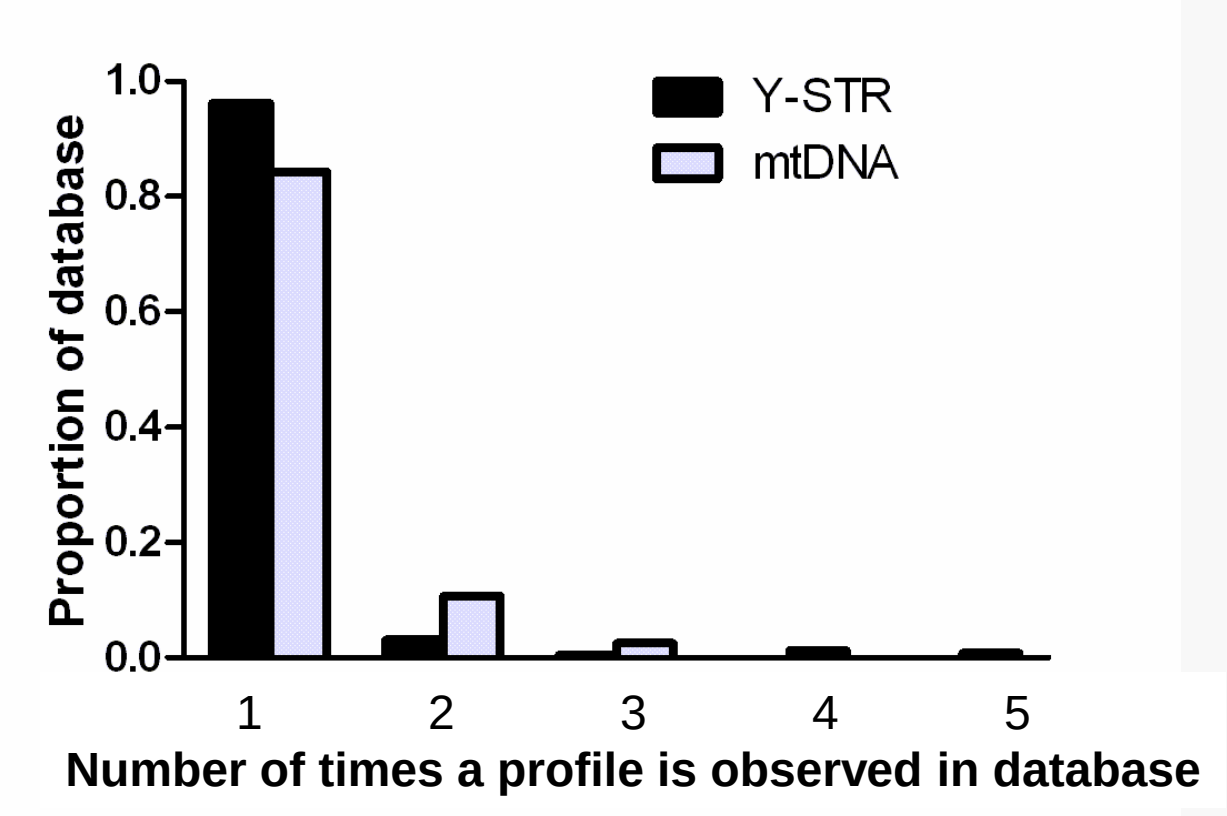
For uncle-nephew (34 vs. 10):
LR = 1 / 0.00009 = 1.1E+04
LR with 40 autoSTRs = 8.98E+05

BUT…

For brother vs. sister (19 vs 18):
LR = 0
LR with 40 autoSTRs = 1.57E+19

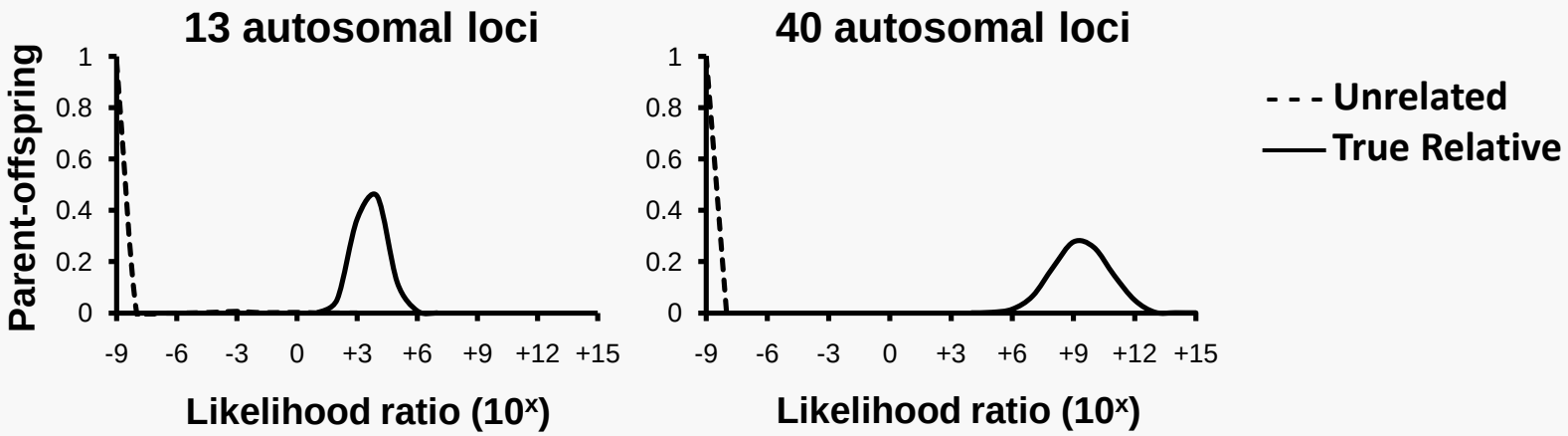
Lineage markers may not be as statistically robust as autosomal markers, but they are lineage specific and can link distant relatives.

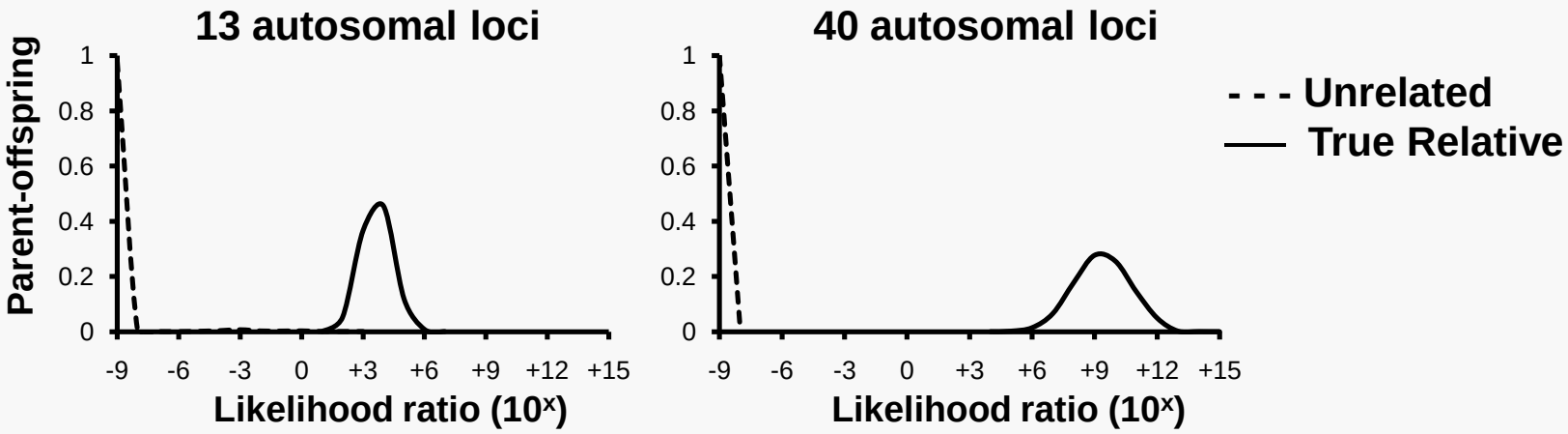
Frequency of Y-STR and mitochondrial DNA profiles in the NIST population database (n = 572)



96% of Y-STR profiles are unique
84% of mtDNA profiles are unique

Simulated Datasets

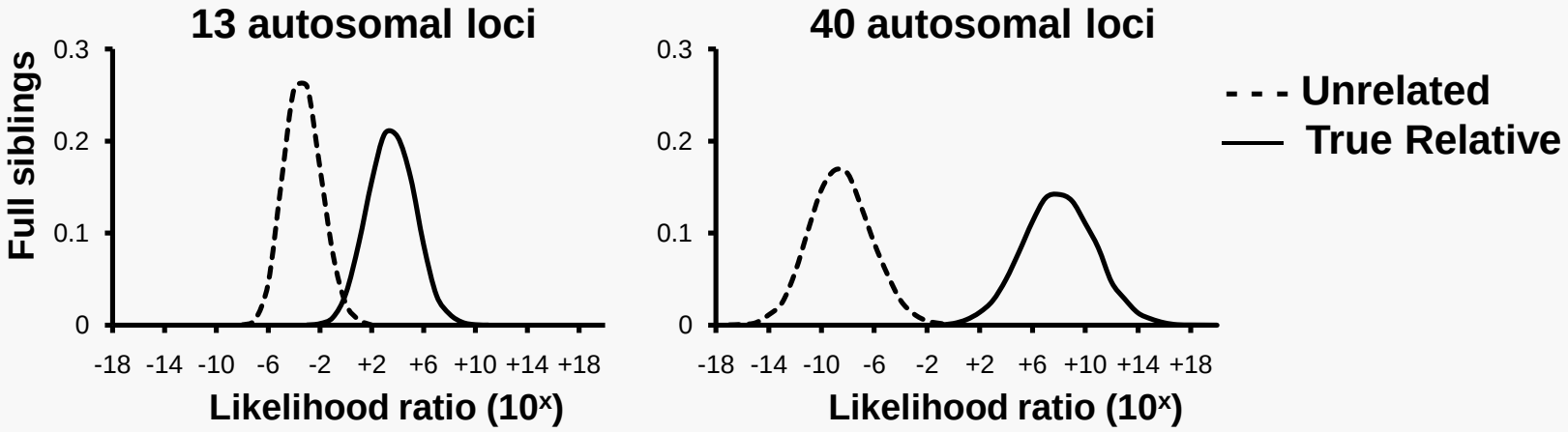




Probability of finding a true parent-offspring relative and the number of false positives (LR threshold ≥ 100 = 10^2)

Loci	Probability of true relative	Number of false positives
13	0.95	18
40	1.00	0

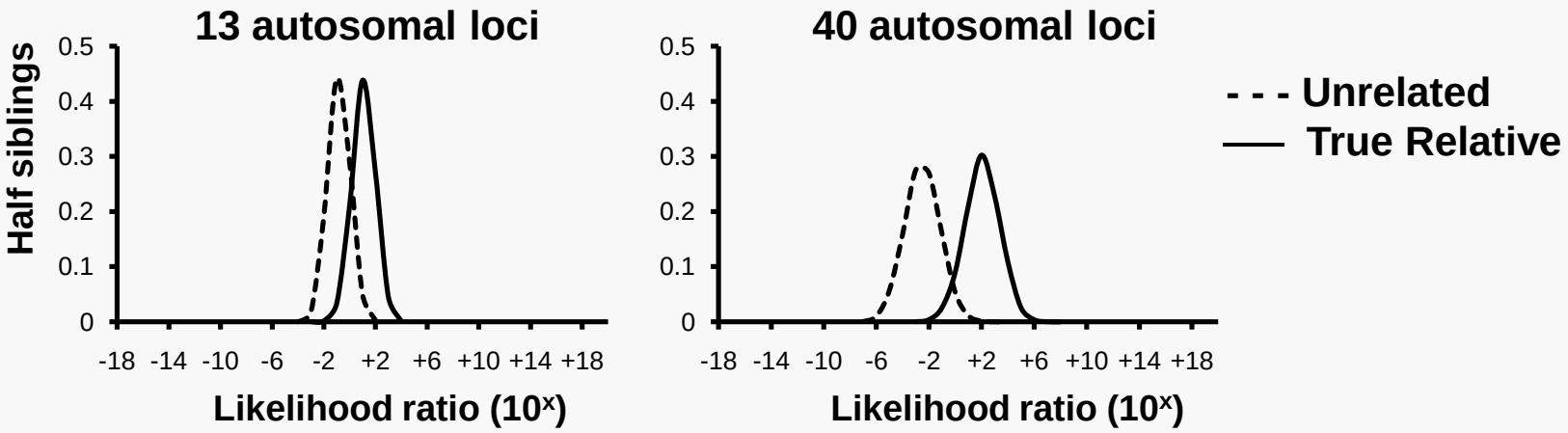
Based on a database of 100,000 individuals with U.S. allele frequencies.



Probability of finding a true full sibling relative and the number of false positives (LR threshold ≥ 100 = 10^2)

Loci	Probability of true relative	Number of false positives
13	0.70	5
40	0.97	0

Based on a database of 100,000 individuals with U.S. allele frequencies.



Probability of finding a true half sibling relative and the number of false positives (LR threshold ≥ 10 = 10^1)

Loci	Probability of true relative	Number of false positives
13	0.73	4,570
40	0.83	607

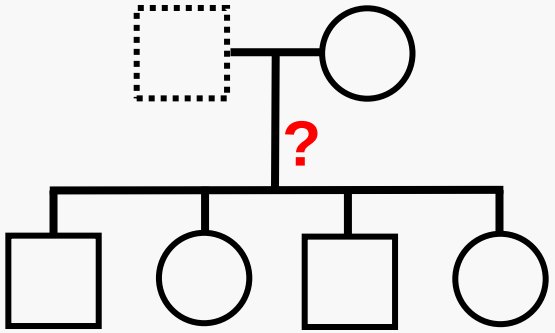
Based on a database of 100,000 individuals with U.S. allele frequencies.

Immigration:
Kinship Analysis as a DNA Biometric

US Citizenship and Immigration Services (USCIS)

- Immigration cases
  - 1,107,126 obtained legal permanent residence in US in 2008
  - 103,456 were relatives of US alien resident

- Refugee/asylum cases
  - 400 applications processed per day
  - 60,108 refugees admitted in 2008
  - 34,753 were relatives of applicant



- Support relationship claim with interview and documents
- Fraudulent claims (79%)

Chris Miles, DHS

Current: Optional DNA Testing for Immigration

- DNA may be used if interview and documents are insufficient
- Number of cases:
  - ~ One dozen accredited labs in US
  - ~ 3,500/yr for small lab, ~10,000/yr large lab
- Time: Minimum two-day lab analysis;
  up to six months to coordinate with embassy
- Cost: $600-$1500 for private lab testing
  - Applicant pays (usually US resident, “anchor”)

Future: Requiring DNA for Immigration Testing

- Easy, rapid collection
  - Trained embassy/field office staff vs. accredited physician
  - Buccal swab
- Faster, cheaper testing
  - Not necessary to have one-hour turnaround time
  - Relationship in question could dictate the markers to be analyzed
- Analysis
  - Expert system may not give definitive answer
    - May help to flag DNA quality or genotyping issues
  - Analysts trained in statistics and kinship analysis may be necessary
- Perform on-site or in US government lab? Outsource to private lab?
- Many more questions…
  - Volume? Cost? Population allele frequencies? Level of certainty?

Analysis Software

Expert Systems

- Programs that interpret data as a human expert would

- Expert systems for forensic typing (NDIS-approved)
  - FSS-I^3 (FSS/Promega)
  - GeneMapper ID-X (Applied Biosystems)
  - TrueAllele (CyberGenetics)

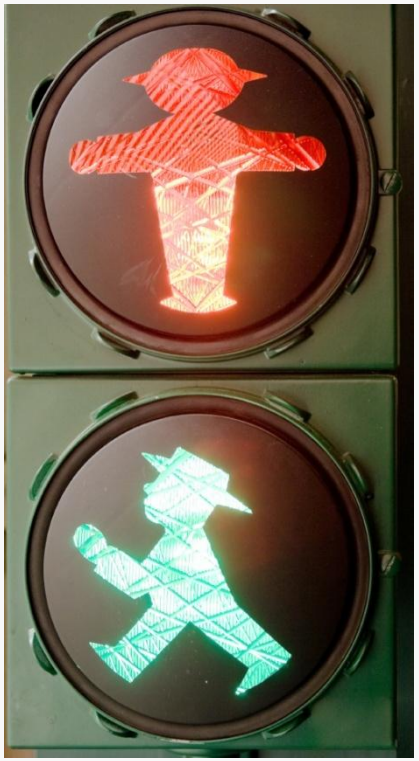
- Expert systems for kinship analysis
  - No program has been designated as an expert system for automated kinship determination

NIJ Expert System Test (NEST) Project, http://forensics.marshall.edu/NEST/default.html

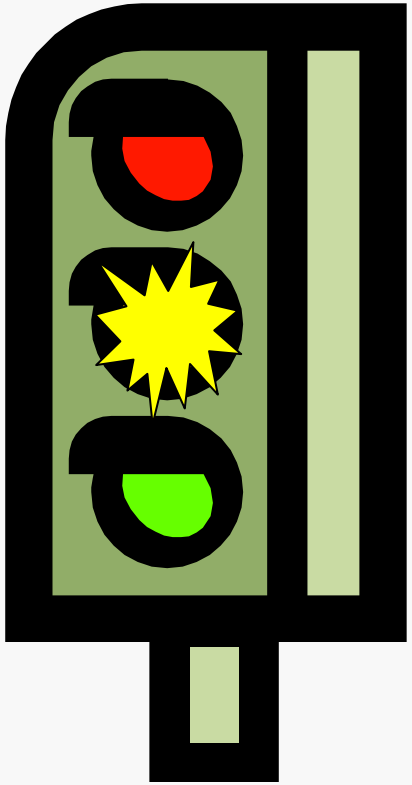
Single-Source Genotyping

non-match

match



Complex Kinship Analysis



Who defines hypotheses?

Threshold?

Inconclusive?

Missing relatives?

Kinship Analysis Software

- Public/commercial availability
  1. GeneMarker® HID v1.90 (SoftGenetics)
  2. DNA-VIEW™ v29.11 (Charles Brenner)
  3. LISA (Future Technologies Inc.)
  4. KIn CALc v3.1 (CA DOJ)

- Restricted availability
  5. CODIS 6.0 (FBI)

- Coming soon
  6. FSS-ibd (Forensic Science Service)

(1) GeneMarker® HID

- Fragment analysis program, primarily
- Kinship module
  - Draw pedigree tree (show inheritance conflicts)
  - LR approach
  - Pairwise analysis only
- Databases of allele frequencies, loci
  - Autosomal STRs, Y-STRs
  - No mitochondrial DNA

SoftGenetics: http://www.softgenetics.com/

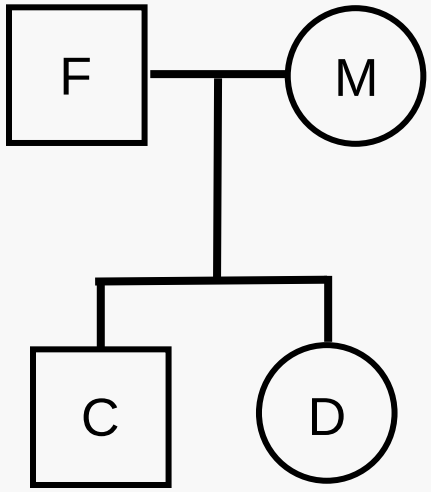
(2) DNA-VIEW™

- DOS, command-line interface
- Kinship modules
  - Paternity and complex kinship analyses
  - Pedigree simulations
  - Define relationships (pedigree tree) with symbols
  - Use up to 10 relatives in analysis
  - LR approach
- Databases of allele frequencies, loci
  - Autosomal STRs or Y-STRs
  - No mitochondrial DNA for kinship
  - Mutation considered

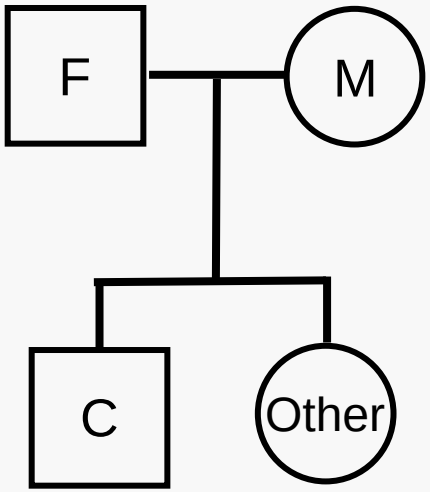
Dr. Charles Brenner: http://dna-view.com/

Immigration Scenario with DNA-VIEW

Is “D” the sister of “C,” as is claimed?



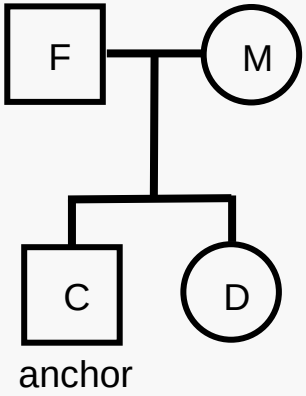
anchor



anchor

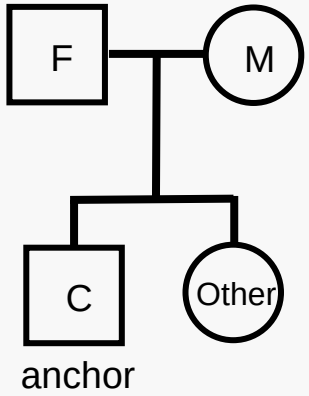
Hypothesis 1
C: M + F
D: M + F

Hypothesis 2
C: M + F
Other: M + F



anchor

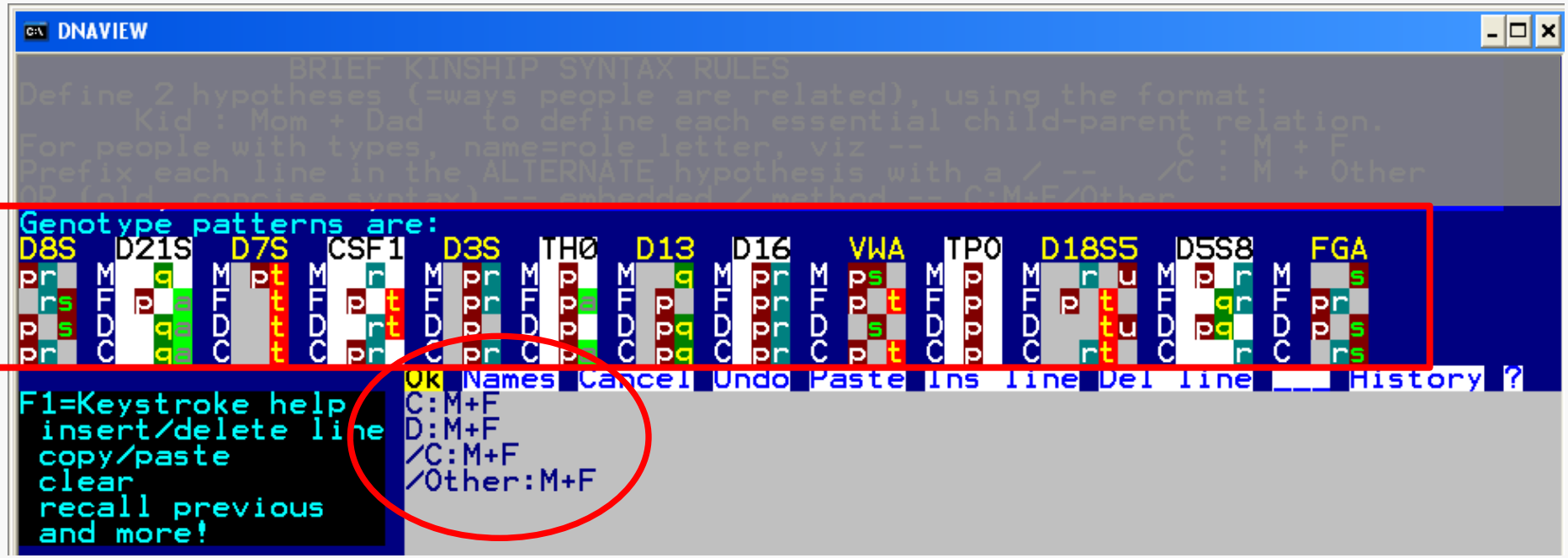
Hypothesis 1
C: M + F
D: M + F

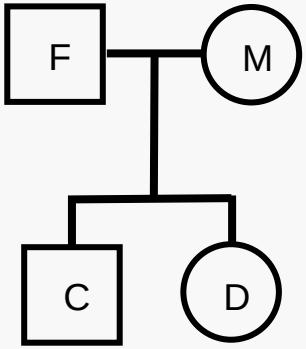


anchor

Hypothesis 2
C: M + F
Other: M + F

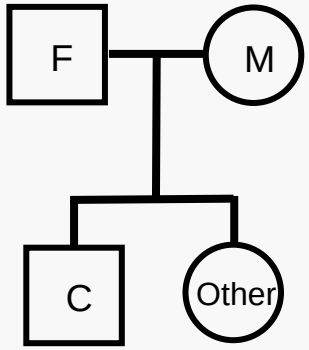
Is “D” the sister of “C”?





anchor

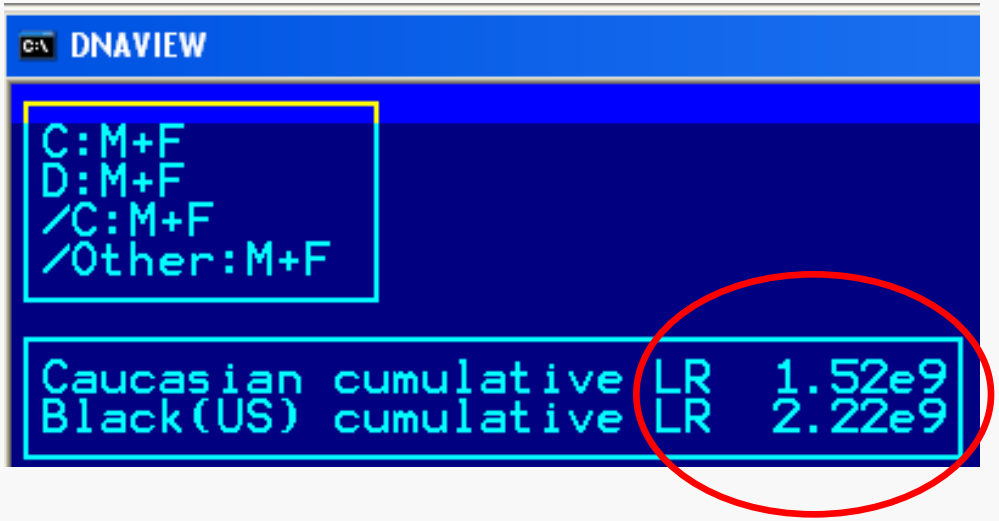
Hypothesis 1
C: M + F
D: M + F



anchor

Hypothesis 2
C: M + F
Other: M + F

Is “D” the sister of “C”?



(3) LISA

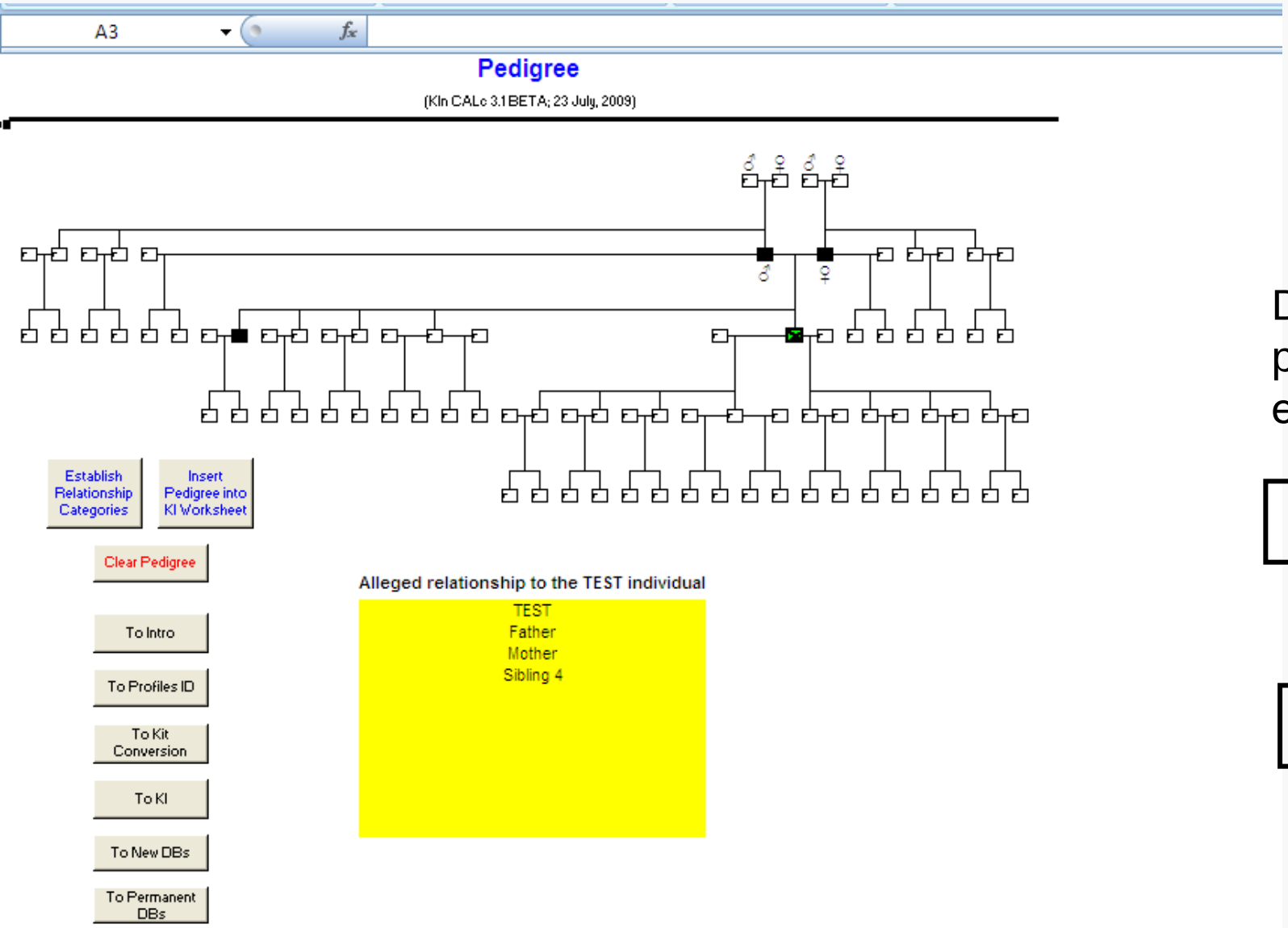
- Laboratory Information Systems Application
- Originally developed for AFDIL
- Graphical interface for DNA-VIEW™
- Uses Progeny® to draw pedigree trees
- Autosomal STRs, Y-STRs, mtDNA

Future Technologies Inc.: http://www.ftechi.com/

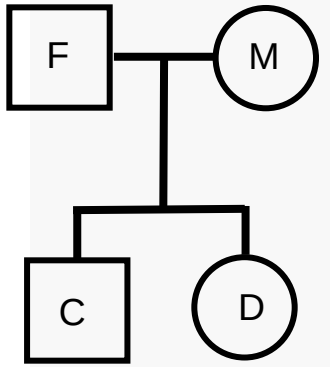
(4) KIn CALc

- Excel, macro-driven program
- Kinship module
  - Define relationships on pedigree tree
  - Use up to 10 relatives in analysis
  - LR approach based on DNA-VIEW algorithms
  - No mutation calculation (flags mutation)

KIn CALc



DNA-VIEW pedigree example…



KIn CALc

Combined likelihood ratio

	AfAm	Cauc	Hisp
Combined KI	0.0E+00	0.0E+00	0.0E+00

Locus	Reference Sample Profiles D	F	M	C	KI (locus) AfAm	Cauc	Hisp
D8S1179	12 , 15	14 , 15	12 , 14	12 , 14	5.39461	7.8368	8.94681
D21S11	30 , 30.2	29 , 30.2	30 , 30	30 , 30.2	100.128	28.1436	23.6561
D7S820	12 , 12	12 , 12	8 , 12	12 , 12	61.0803	25.3672	13.6503
CSF1PO	12 , 14	10 , 14	12 , 12	10 , 12	70	52.0316	53.2695
D3S1358	15 , 15	15 , 17	15 , 17	15 , 17	2.96291	4.1209	1.37865
TH01	8 , 8	8 , 9.3	8 , 8	8 , 9.3	14.497	31.687	75.5727
D13S317	11 , 12	11 , 11	12 , 12	11 , 12	4.35784	5.0798	11.4216
D16S539	9 , 11	9 , 11	9 , 11	9 , 11	4.27868	8.83203	10.0079
D2S1338							
					0	0	0
					7.36735	3.37495	3.24621
					4.03789	7.49727	7.82846
D5S818	11 , 12	12 , 13	11 , 13	13 , 13	1.34641	0.86107	1.02114
FGA	22 , 25	22 , 24	25 , 25	24 , 25	11.1111	19.2272	10.2205

vWA: Alleged Sib 18,18 Father 15,19 Mother 15,18

Mutation

Per locus likelihood ratios

(5) CODIS 6.0

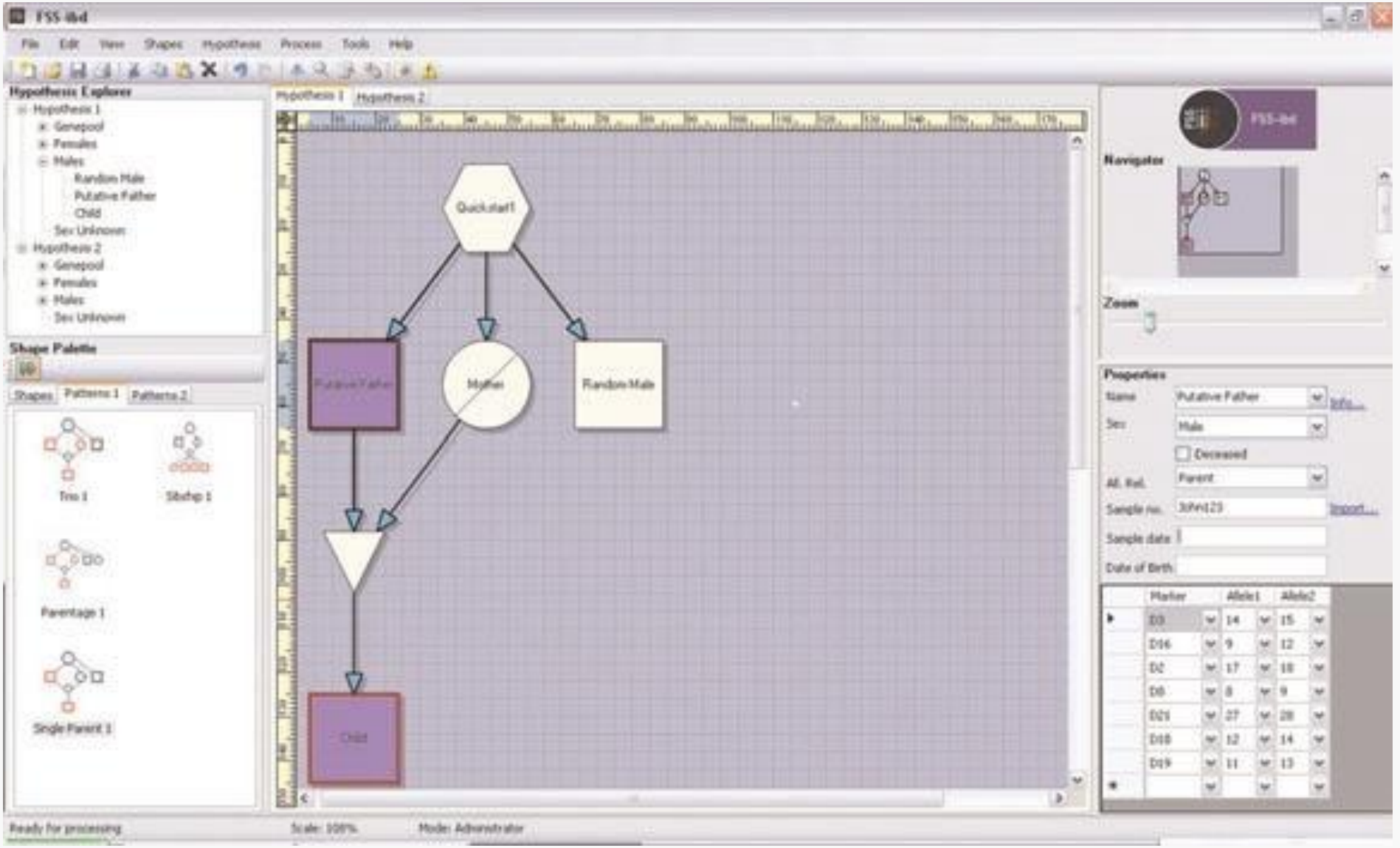
- FBI-developed program for missing persons and unidentified human remains identification
- Kinship module
  - Draw pedigree tree to define relationships
- Joint Pedigree LR
  - Ranks putative relatives
- Database of allele frequencies, loci
  - Autosomal STRs, Y-STRs, mtDNA
  - Metadata in v. 7.0

(6) FSS-ibd

- Automated relationship testing application
- Graphical user interface
- Simple paternity to complex kinship
- LR approach
- Database of allele frequencies, loci
  - Provided and user-defined
  - Autosomal STRs
  - Mutation considered

FSS: http://www.forensic.gov.uk/

FSS-ibd



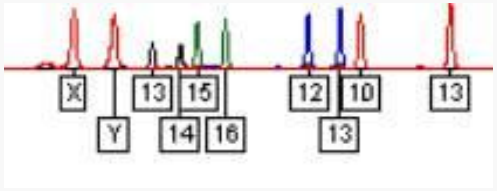
Maguire, C.N., Woodward, M. (2008) DNA-based kinship analysis, Profiles in DNA 11 (1) 4-6.

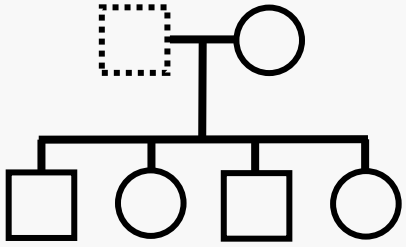
Software Validation

Validation information: http://www.cstl.nist.gov/strbase/validation.htm

- Developmental validation (manufacturer)
- Internal validation (individual labs)

Rapid DNA Testing Platform





Instrument (generates data) → Expert system (data analysis) → Kinship Software (stats and conclusion)

Internal Validation Considerations for Kinship Software

- Who should recommend validation guidelines for DNA Biometrics and kinship analysis?
  - For forensic typing: Scientific Working Group on DNA Analysis Methods
  - For paternity testing: AABB
- Define case examples to test software
  - Paternity trios, deficient cases, complex kinship, false relationships
  - How many cases are enough?
  - Real data, simulated, or both?
- Technical considerations specific to testing
  - Marker systems, allele frequency data, mutation, LR threshold, etc.
- How do you know your answer is correct?
  - Compare to hand-calculated answer or to another kinship program?

Drábek J, Validation of software for calculating the likelihood ratio for parentage and kinship, FSI:Genetics 3 (2009) 112–118

Conclusions

- DNA is a powerful tool to confirm or refute alleged familial relationships
  - Important to define specific kinship questions
- Simulations with appropriate allele frequency databases can model expected LR values for specific relationships
- Software must be “fit for purpose”
- Can there be an expert system for kinship analysis?
- What are the validation parameters? Who should define validation guidelines?

Thank you for your attention

Acknowledgments:

Dr. Peter Vallone
Erica Butts

Funding:

NRC – Postdoctoral Fellowship to Kristen Lewis

FBI – Evaluation of Forensic DNA Typing as a Biometric Tool

NIJ – Interagency Agreement with the Office of Law Enforcement Standards

Contact:

kristen.lewis@nist.gov
301-975-5205

peter.vallone@nist.gov
301-975-4872