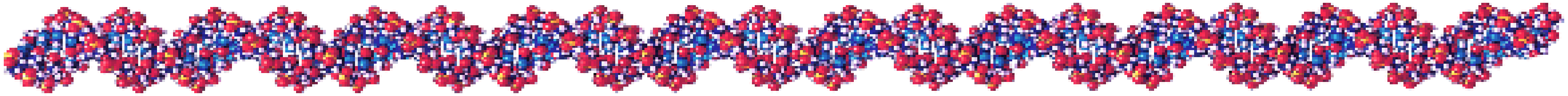
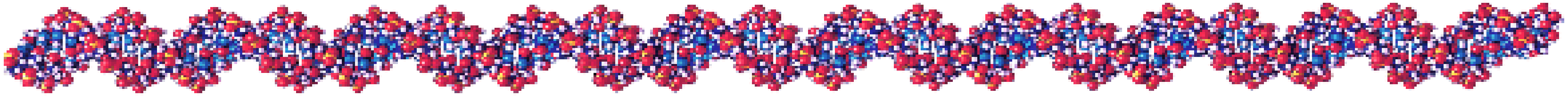


Statistical Methods for Kinship Analysis



Kristen E. Lewis, Ph.D.

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

Towson University
April 26, 2010

Overview

- Measures of relatedness
- Inbreeding in a population
- Probability of observed genotypes from related individuals
- Likelihood ratio calculations for kinship testing
- Bayesian statistics
- Background relatedness

Measures of Relatedness

- Identity by descent
- Coancestry coefficient
- Inbreeding coefficient

Basic Concepts

- Two genes which are copies of a common ancestral gene are said to be **identical by descent (IBD)**
  - IBD is unobservable

- Two genes that are not shared by descent but have the same allelic state are said to be **identical by state (IBS)**
  - Allelic states are observable

- Observation of genotype → inference of relatedness

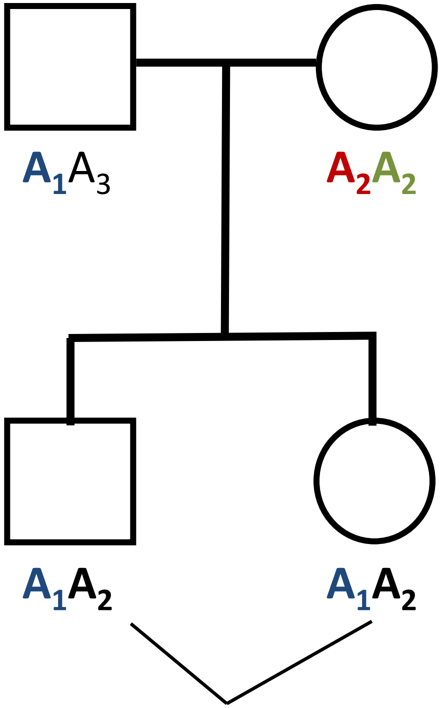
Example: The observed genotypes of unidentified remains and the brother of a missing man can be used to answer the question of whether the body is that of the missing man

IBD vs. IBS

IBS: Alleles of the same type

- Same base for a SNP
- Same number of STR repeats

IBS does not generally equate to IBD



A_1A_3 A_2A_2

A_1A_2 A_1A_2

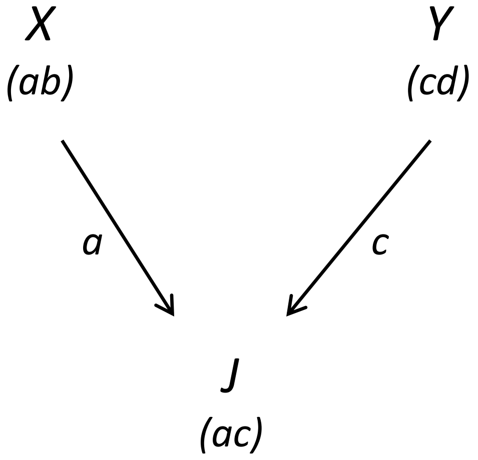
A_1 is IBD

A_2 or A_2

A_2 is IBS
IBD status is unknown

Coancestry Coefficient (θ)

Probability that an allele taken at random from one individual is IBD to an allele taken at random from another individual



X
(ab)

Y
(cd)

a

c

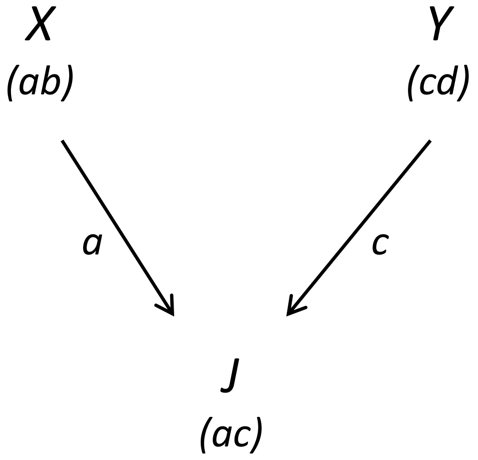
J
(ac)

$\theta_{XY} = \Pr(a \equiv c)$

Inbreeding Coefficient (*F*)

Probability that an individual carries two IBD alleles at a locus

Inbreeding coefficient of an individual is the coancestry of its parents



X *(ab)*

Y *(cd)*

a

c

J *(ac)*

$$F_J = \theta_{XY}$$

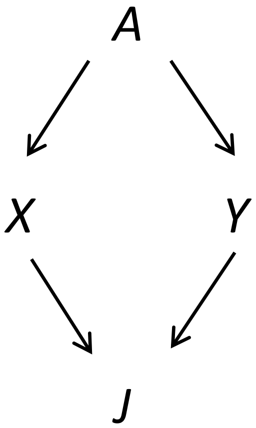
Calculating Relatedness Coefficients

Path Counting Method

Probabilities of identity are defined relative to an ancestor common to both maternal and paternal lineages of an individual

Inbreeding Coefficient: n = number of individuals in the pathway that link the **parents** to the common ancestor, including the parents

Coancestry Coefficient: n = number of individuals in the pathway that link the **individuals** to the common ancestor, including the individuals



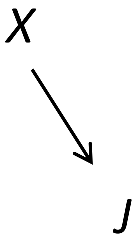
$$F_J = \theta_{XY} = (1/2)^n (1 + F_A)$$

F_A = inbreeding coefficient of common ancestor

Expression summed over all pathways and common ancestors

Path Counting

How related are Parent-Offspring?



Common ancestor of *X* and *J*: *X*

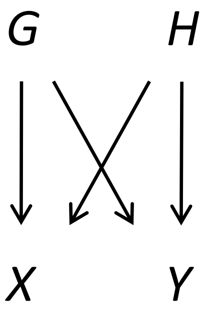
Path linking *X*,*J* to common ancestor: *XJ*

Number of individuals in path: 2

Coancestry of *X* and *J*: $\theta_{XJ} = (1/2)^2 = 1/4$

Path Counting

How related are Full Siblings?



Common ancestor of *X* and *Y*: *G* and *H*

Path linking *X*,*Y* to common ancestor: *X*-*G*-*Y* and *X*-*H*-*Y*

Number of individuals in path: 3

Coancestry of *X* and *Y*: $\theta_{XY} = (1/2)^3 + (1/2)^3 = 1/4$

Coancestry for Common Relationships

Relationship	**Coancestry**
Unrelated	0
Identical twins	1/2
Parent-child	1/4
Full siblings	1/4
Half siblings	1/8
Uncle-nephew	1/8
Grandparent-grandchild	1/8
Double first cousins	1/8
First cousins	1/16

Effect of Inbreeding in a Population

Individual Genotype Frequencies

Assume all individuals in a population have an inbreeding coefficient, *F*

- Probability an individual has two IBD alleles is *F*
- Probability an individual does not have two IBD alleles is 1 – *F*
- Probability that any allele is type A is p_A

Probability that an individual is homozygous:

$$P_{AA} = \underbrace{\boxed{F \times p_A}}_{\text{2 alleles IBD}} + \underbrace{\boxed{(1 - F) \times p_A^2}}_{\text{no alleles IBD}}$$

Rearranged similar to Hardy-Weinberg equation:

$$P_{AA} = \underbrace{\boxed{p_A^2}}_{\text{HWE}} \underbrace{\boxed{+ \; Fp_A(1 - p_A)}}_{\text{inbreeding}}$$

Individual Genotype Frequencies with Inbreeding

Homozygous:

$$P_{AA} = p_A^2 + Fp_A(1 - p_A)$$

Heterozygous individuals must have non-IBD alleles:

$$P_{Aa} = \underbrace{\boxed{2(1 - F)\, p_A p_a}}_{\text{no alleles IBD}}$$

Rearranged similar to Hardy-Weinberg equation:

$$P_{Aa} = \underbrace{\boxed{2p_A p_a}}_{\text{HWE}} \underbrace{\boxed{- 2Fp_A p_a}}_{\text{inbreeding}}$$

Example: Effect of Inbreeding in Population

If every person in the population had parents who were first cousins, inbreeding coefficient $F = 1/16 = 0.0625$

$$P_{AA} = p_A^{\,2} + 0.0625 p_A(1 - p_A)$$

$$P_{Aa} = 2p_A p_a - 0.125 p_A p_a$$

If $p_A = p_a = 0.2$,

$P_{AA} = 0.040 + 0.01 = \mathbf{0.050}$

$P_{Aa} = 0.080 - 0.005 = \mathbf{0.075}$

How does inbreeding influence **allelic diversity**? What about **disease rates**?

Inbreeding…
Increases homozygosity
Decreases heterozygosity

15 IBD Possibilities for Any Two Individuals

For individuals *X* (*ab*) and *Y* (*cd*):

none

$a \equiv b$

$c \equiv d$

$a \equiv c$

$a \equiv d$

$b \equiv c$

$b \equiv d$

$a \equiv b \equiv c$

$a \equiv b \equiv d$

$a \equiv c \equiv d$

$b \equiv c \equiv d$

$a \equiv b, c \equiv d$

$a \equiv c, b \equiv d$

$a \equiv d, b \equiv c$

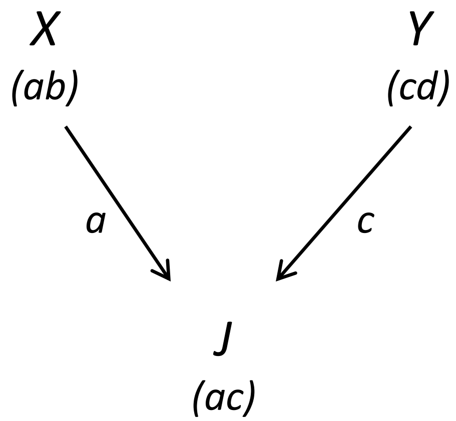
$a \equiv b \equiv c \equiv d$

- A probability value is associated with each of these IBD possibilities (e.g., δ_{ac} or $\delta_{ac.bd}$)
- When neither individual is inbred, the 15 IBD possibilities can be reduced to seven possible IBD relationship. See: Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80
- The seven can be reduced to a three-parameter set (next slide)

Seven IBD Possibilities for Two Non-Inbred Individuals

For non-inbred individuals, there are seven IBD possibilities that can be described by **three k-coefficients** corresponding to probabilities that individuals have zero, one, or two IBD alleles

Recall that the coancestry θ_{XY} is the probability that a random allele from *X* is IBD to a random allele from *Y*



X (ab) Y (cd)

a c

J (ac)

$\theta_{XY} = \Pr(a \equiv c)$

IBD Alleles	**Probability**
none	$k_0 = \delta_0$
a,c or *a,d* or *b,c* or *b,d*	$k_1 = \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}$
(*a,c* and *b,d*) or (*a,d* and *b,c*)	$k_2 = \delta_{ac.bd} + \delta_{ad.bc}$

k-coefficients sum to 1

Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80

So far…

We have defined measures of relatedness:

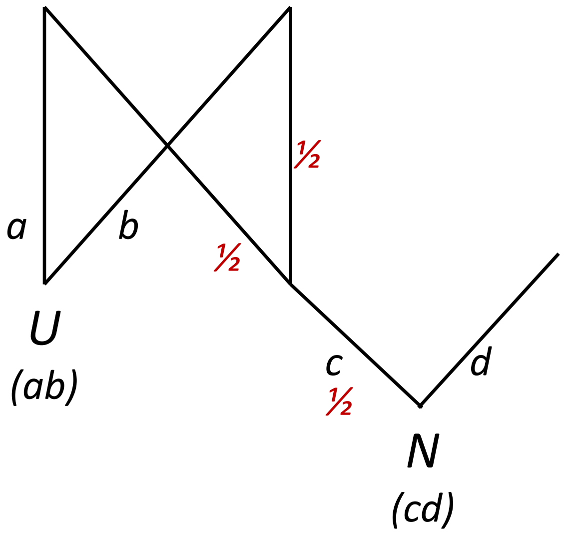
- Coancestry coefficient, θ
- Inbreeding coefficient, F
- IBD states (e.g., $a \equiv b$) and related k-coefficients

We know there are IBD probabilities (δ) associated with the three IBD k-coefficients

But how are these probabilities calculated for specific relationships?

How are the IBD coefficients determined?

Uncle-Nephew example



Determine allele transmission frequencies:

50/50 chance of a parent transmitting one of its alleles to an offspring

$$\boldsymbol{k_0} = \tfrac{1}{2}$$

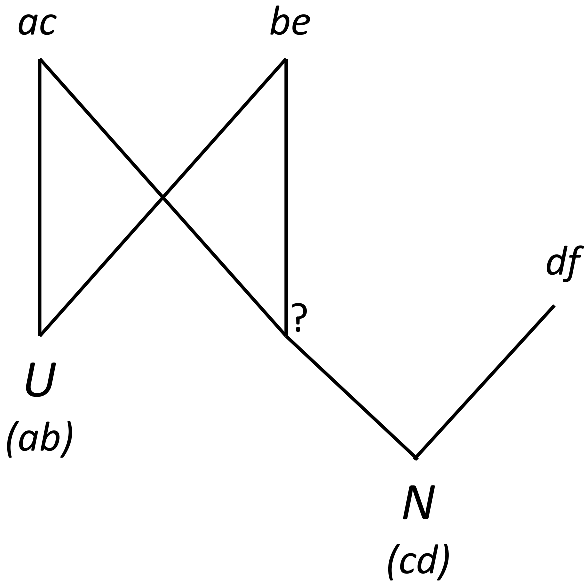
$$\left.\begin{array}{l} \Pr(a \equiv c) = (\tfrac{1}{2})\,(\tfrac{1}{2}) = \tfrac{1}{4} \\ \Pr(b \equiv c) = (\tfrac{1}{2})\,(\tfrac{1}{2}) = \tfrac{1}{4} \end{array}\right\} \boldsymbol{k_1} = \tfrac{1}{2}$$

$$\boldsymbol{k_2} = 0$$

Can also draw a matrix of all possible genotypes of offspring (next slide)

How are the IBD coefficients determined?

Uncle-Nephew example



Determine genotype matrix:

Possible genotypes of {[child of (ac x be)] x df}

Possible genotypes of (ac x be)

	ad	af	bd	bf	cd	cf	ed	ef
ab	1	1	1	1	0	0	0	0
cb	0	0	1	1	1	1	0	0
ae	1	1	0	0	0	0	1	1
ce	0	0	0	0	1	1	1	1

$\mathbf{k_0 = 16/32 = \tfrac{1}{2}}$ $\mathbf{k_1 = 16/32 = \tfrac{1}{2}}$ $\mathbf{k_2 = 0}$

1. Determine possible offspring combinations from the mating of U’s parents (ac x be)
2. Determine possible offspring combinations from the mating of N’s parents (df x each of the genotypes from Step 1)
3. Count number of shared (IBD) alleles and divide by the total possible

IBD Allele Sharing Values for Common Relationships

Relationship	k_0	k_1	k_2	**Coancestry** $\theta = \frac{1}{2} k_2 + \frac{1}{4} k_1$
Unrelated	1	0	0	0
Identical twins	0	0	1	1/2
Parent-child	0	1	0	1/4
Full siblings	1/4	1/2	1/4	1/4
Half siblings	1/2	1/2	0	1/8
Uncle-nephew	1/2	1/2	0	1/8
Grandparent-grandchild	1/2	1/2	0	1/8
Double first cousins	9/16	3/8	1/16	1/8
First cousins	3/4	1/4	0	1/16

k-coefficients: Probability of sharing zero, one, or two IBD alleles

Let’s consider the genotypes of two individuals…

What are the genotype probabilities when jointly considering genotypes from two individuals?

It is **not** p^2 or 2pq for two unrelated individuals! Why?

In this case, determine probability of observing allele *i* under different IBD states when you jointly consider the genotypes from two homozygous individuals.

IBD Relationship	$X = A_iA_i$ $Y = A_iA_i$
δ_0	p_i^4
δ_{ac}	p_i^3
δ_{ad}	p_i^3
δ_{bc}	p_i^3
δ_{bd}	p_i^3
$\delta_{ac.bd}$	p_i^2
$\delta_{ad.bc}$	p_i^2

Probability of observing allele *i* for two **unrelated**, non-inbred individuals with 0 IBD alleles

… for two non-inbred individuals with 1 pair of IBD alleles

… for two non-inbred individuals with 2 pairs of IBD alleles

Genotype probabilities when jointly considering genotypes from two individuals

Probability of observing alleles *i*,*j*,*k*,*l* under different IBD states for all possible two-person genotype combinations.

Joint Genotypes	k_0	k_1	k_2
ii,ii	p_i^4	p_i^3	p_i^2
ii,jj	$p_i^2p_j^2$		
ii,ij	$2p_i^3p_j$	$p_i^2p_j$	
ii,jk	$2p_i^2p_jp_k$		
ij,ij	$4p_i^2p_j^2$	$p_ip_j(p_i + p_j)$	$2p_ip_j$
ij,ik	$4p_i^2p_jp_k$	$p_ip_jp_k$	
Ij,kl	$4p_ip_jp_kp_l$		

Now…

We have defined measures of relatedness:

- coancestry coefficient, θ
- inbreeding coefficient, *F*
- IBD states (e.g., $a \equiv b$) and related *k*-coefficients

We can calculate IBD probabilities (δ) associated with the three *k*-coefficients for various relationships

We can define probabilities of observing different allele combinations for two individuals

How do we bring all of these parameters together to determine the probability of observing two genotypes if the individuals are related?

Joint Genotypic Probabilities (non-inbred)

Determine probability of individuals' genotypes when their degree of relationship is known or assumed

Express genotype probabilities for pairs of individuals as a function of the allele probabilities and IBD relationship

Genotypes	**Genotypic state**	**Non-inbred Joint Genotypic Probability**
A_iA_i, A_iA_i	Hom/hom	$k_2P_i^2 + k_1P_i^3 + k_0P_i^4$
A_iA_i,A_jA_j	Hom/hom	$k_0P_i^2P_j^2$
A_iA_i, A_iA_j	Hom/het	$k_1P_i^2P_j + 2k_0P_i^3P_j$
A_iA_i, A_jA_m	Hom/het	$2k_0P_i^2P_jP_m$
A_iA_j,A_iA_j	Het/het	$2k_2P_iP_j + k_1P_iP_j(P_i + P_j) + 4k_0P_i^2P_j^2$
A_iA_j,A_iA_m	Het/het	$k_1P_iP_jP_m + 4k_0P_i^2P_jP_m$
A_iA_j,A_mAl	Het/het	$4k_0P_iP_jP_mP_l$

Adapted from Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80

Example of Joint Genotypic Probabilities and Accounting for Relatedness

Probability that two individuals are A_iA_i if **full siblings**:

$$\Pr(A_iA_i, A_iA_i) = k_2 p_i^2 + k_1 p_i^3 + k_0 p_i^4$$

$$\Pr(A_iA_i, A_iA_i \mid \text{full sibs}) = (1/4)p_i^2 + (1/2)p_i^3 + (1/4)p_i^4$$

$$\Pr(A_iA_i, A_iA_i \mid \text{full sibs}) = \boldsymbol{p_i^2(1 + p_i)^2/4}$$

Compare this to the probability for two **unrelated** A_iA_i individuals: $\boldsymbol{p_i^4}$

If $p_i = 0.2$,

$$\Pr(A_iA_i, A_iA_i \mid \text{full sibs}) = 0.0144$$

$$\Pr(A_iA_i, A_iA_i \mid \text{unrelated}) = 0.0016$$

Distinguishing between Putative Relationships

- Paternity testing
  - Is the alleged father the true father of the child or unrelated?

- Human remains identification
  - Are the remains from a person who has a specified relationship to a family member of a missing person? Or unrelated to family member?

Are these individuals related?

- Compare the probabilities of the observed genotypes under alternative hypotheses about relationship

- Use a **likelihood ratio** to determine which competing putative relationship makes the observed genotypes more probable

Likelihood Ratio

Probability of the evidence under different relatedness scenarios

$$LR = \frac{Pr(E|H_1)}{Pr(E|H_2)}$$

$$LR = \frac{Pr(X,Y|S)}{Pr(X,Y|U)}$$

Example:

H_1: X is the sibling of Y

H_2: X is unrelated to Y

Calculate LR for each locus. Multiply LRs across independent loci for combined LR.

Analogous to multiplying Paternity Indices (= LRs) to generate a final Combined Paternity Index for paternity testing.

General Formula of a Likelihood Ratio

Human remains with genotype G_{HR}
Putative parents of missing person are Mother (G_M) and Father (G_F)
Pedigree information denoted by I

H_1: Remains are from a biological child of M and F
H_2: Remains are from an unknown person unrelated to M and F

$$\mathrm{LR} = \frac{\Pr(G_{HR}, G_F, G_M | H_1, I)}{\Pr(G_{HR}, G_F, G_M | H_2, I)}$$

Using the third law of probability, condition G_{HR} on the other information

$$\mathrm{LR} = \frac{\Pr(G_{HR} \mid G_F, G_M, H_1, I)\,\Pr(G_F, G_M | H_1, I)}{\Pr(G_{HR} \mid G_F, G_M, H_2, I)\,\Pr(G_F, G_M | H_2, I)}$$

General Formula of a Likelihood Ratio (cont.)

$$\mathrm{LR} = \frac{\Pr(G_{HR} \mid G_F, G_M, H_1, I)\,\Pr(G_F, G_M \mid H_1, I)}{\Pr(G_{HR} \mid G_F, G_M, H_2, I)\,\Pr(G_F, G_M \mid H_2, I)}$$

If we assume that the joint probabilities of observing the genotypes of the parents do not depend on the two hypotheses, then

$$\Pr(G_F, G_M \mid H_1, I) = \Pr(G_F, G_M \mid H_2, I)$$

$$\mathrm{LR} = \frac{\Pr(G_{HR} \mid G_F, G_M, H_1, I)}{\Pr(G_{HR} \mid G_F, G_M, H_2, I)}$$

Defining the Numerator of a LR:

Joint probabilities for two related non-inbred individuals

Probability of observing all possible two-person genotype combinations under **different IBD states**.

Joint Genotypes	k_0	k_1	k_2
ii,ii	p_i^4	p_i^3	p_i^2
ii,jj	$p_i^2 p_j^2$		
ii,ij	$2p_i^3 p_j$	$p_i^2 p_j$	
ii,jk	$2p_i^2 p_j p_k$		
ij,ij	$4p_i^2 p_j^2$	$p_i p_j (p_i + p_j)$	$2p_i p_j$
ij,ik	$4p_i^2 p_j p_k$	$p_i p_j p_k$	
Ij,kl	$4p_i p_j p_k p_l$		

For *ii,ii* genotypes: $\Pr(ii,ii|H_1) = k_0\, p_i^4 + k_1\, p_i^3 + k_2\, p_i^2$

Adapted from Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80

Defining the Denominator of a LR:

Joint probabilities for two unrelated individuals

Probability of observing all possible two-person genotype combinations for **unrelated, non-inbred individuals**. Unrelated individuals have $k_0 = 1$, $k_1 = k_2 = 0$

Joint Genotypes	k_0
ii,ii	p_i^4
ii,jj	$p_i^2 p_j^2$
ii,ij	$2p_i^3 p_j$
ii,jk	$2p_i^2 p_j p_k$
ij,ij	$4p_i^2 p_j^2$
ij,ik	$4p_i^2 p_j p_k$
Ij,kl	$4p_i p_j p_k p_l$

For ii,ii genotypes: $\Pr(ii,ii|\mathrm{H}_2) = k_0\, p_i^4$

Adapted from Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80

Likelihood ratios for two non-inbred individuals

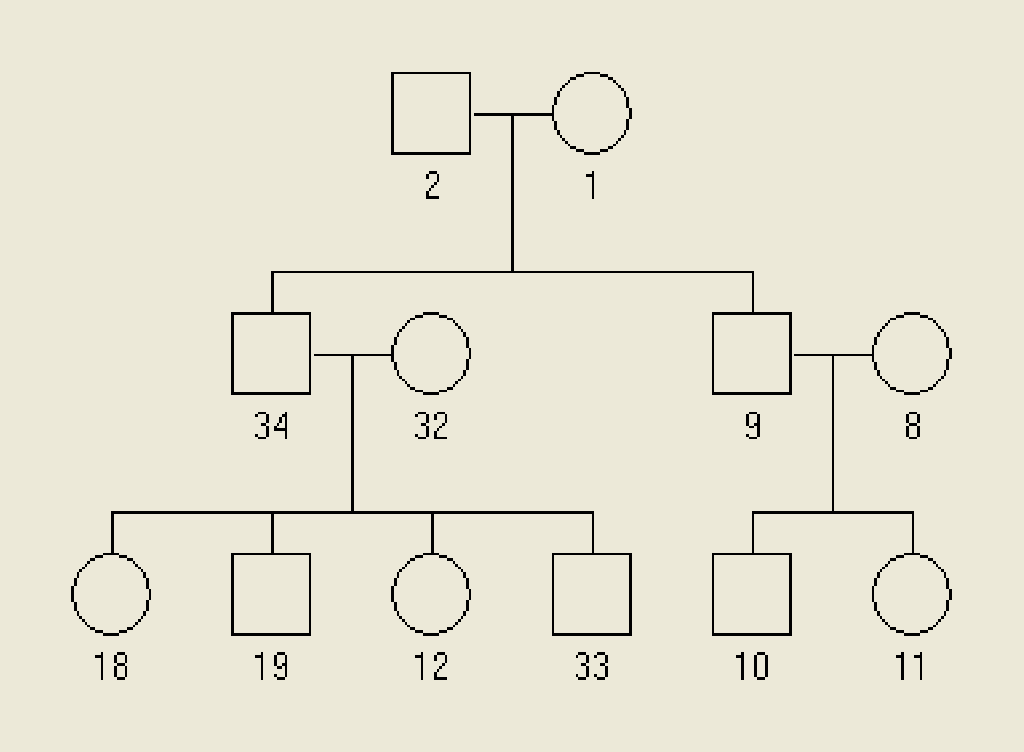
Likelihood ratios of observing all possible two-person genotype combinations under hypotheses of relatedness (defined by IBD states) versus unrelated.

Joint Genotypes	k_0	k_1	k_2
ii,ii	1	$1/p_i$	$1/p_i^2$
ii,jj	1		
ii,ij	1	$1/2p_i$	
ii,jk	1		
ij,ij	1	$(p_i + p_j)/(4p_ip_j)$	$1/(2p_ip_j)$
ij,ik	1	$1/4p_i$	
Ij,kl	1		

For *ii,ii* genotypes: $$\text{LR} = \frac{\Pr(ii,ii\,|\,\text{full sibs})}{\Pr(ii,ii\,|\,\text{unrelated})} = 1/4 + 1/2p_i + 1/4p_i^2$$

Adapted from Weir B., Anderson A., Hepler A., Nat. Rev. Genet. (2006) 7:771-80

Complex Kinship Testing



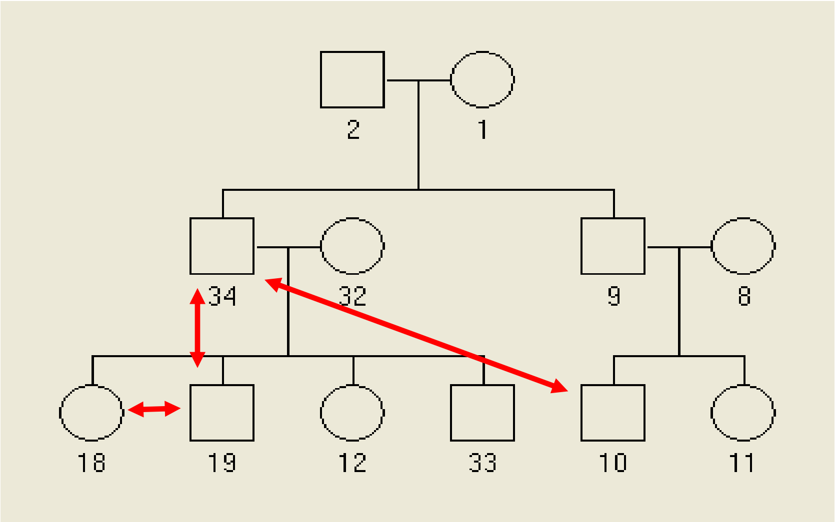
The statistical power for complex kinship testing significantly decreases compared to one-to-one matching

Requirements:

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

- 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

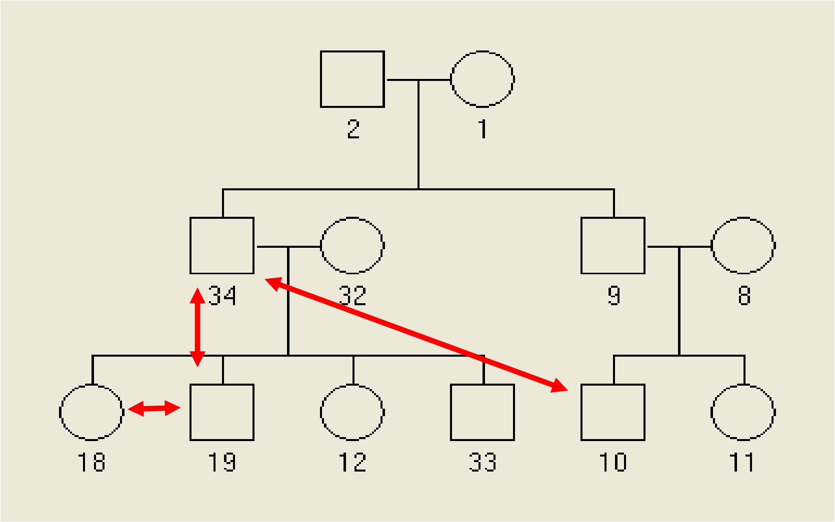


Comparison	LR for 34 & 19	LR for 18 & 19	LR for 34 & 10
Parent-Child	1.28E+06		
Full Siblings		2.76E+07	
Half Siblings			
Cousins			
Uncle-Nephew			6.65E-01
Grandparent-Grandchild			
	Parent/ Child	Full Sibs	Uncle/ Nephew

15 STR loci typed with commercial Identifiler kit

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

15 STR loci typed with commercial Identifiler kit

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

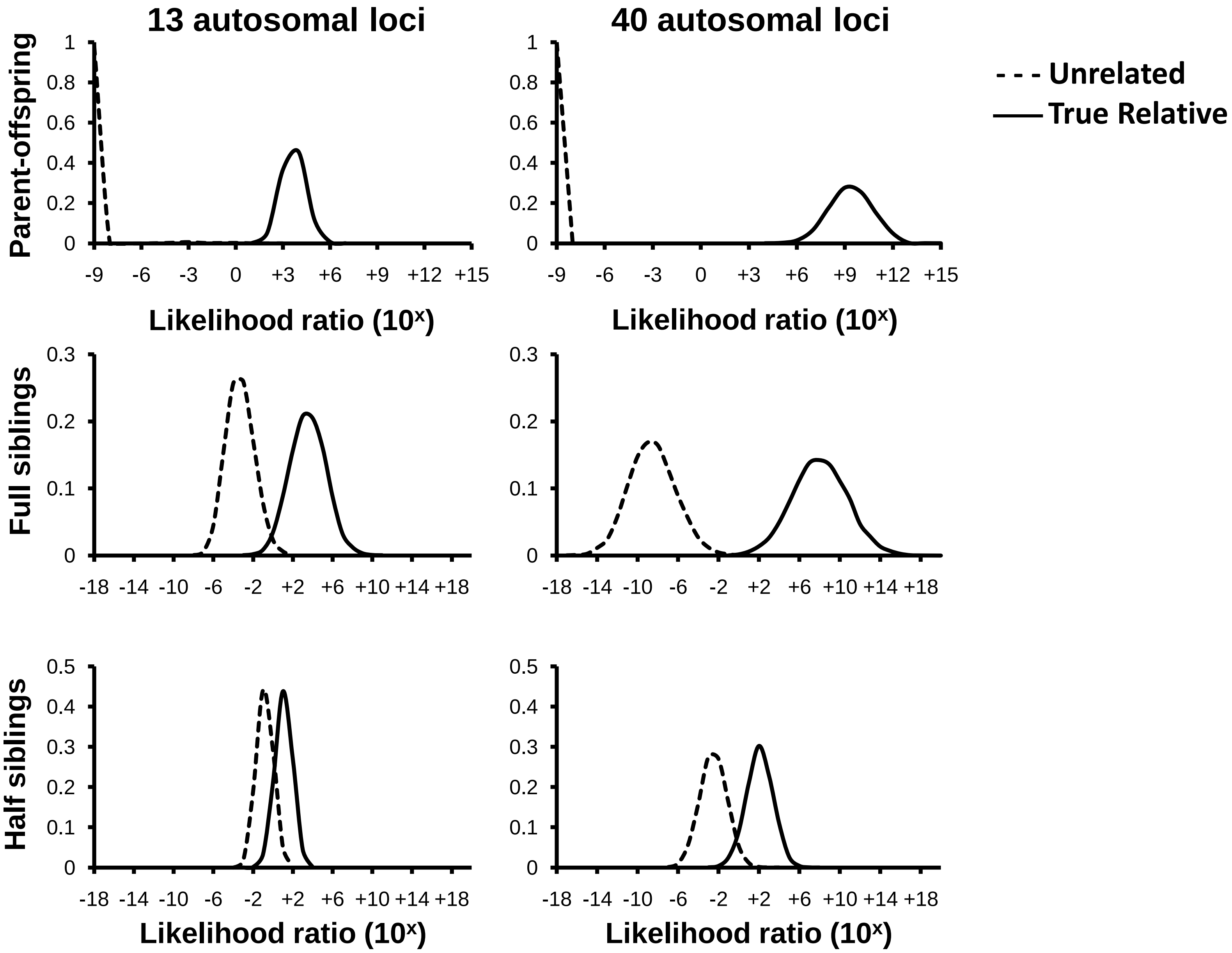
Hill, C.R., Butler, J.M., Vallone, P.M. (2009) A 26plex autosomal STR assay to aid human identity testing. J. Forensic Sci. 54(5): 1008-1015.

Distributions of Likelihood Ratios

Probabilistic nature of inferring relationships based on genotypes

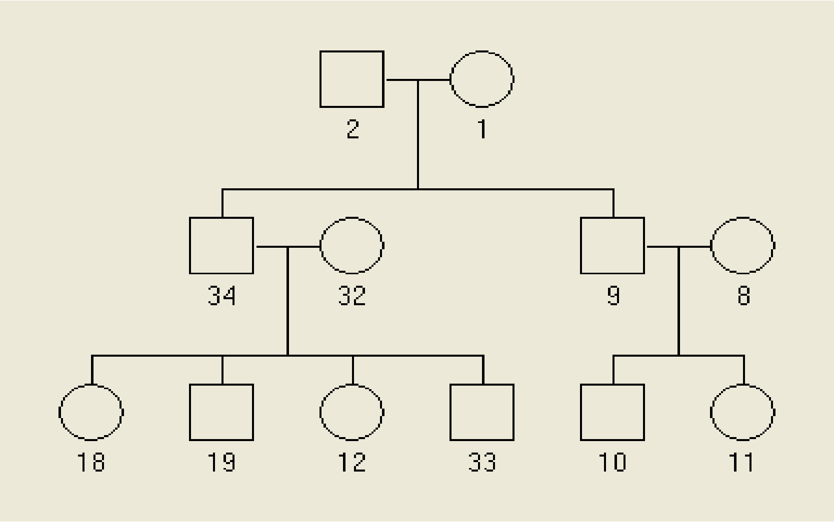
- Unrelated individuals can appear related (false positives)
    - Incorrect relationships may be suggested
- Related individuals can appear unrelated (false negatives)

Likelihood Ratio Distributions



Complex Pedigree Analysis

- For all genotypes and putative relationships, calculate the probabilities of the genotypes under both hypotheses
- Multiply the LRs of each independent piece of the pedigree by each other (DNA-VIEW does this) **OR** consider the likelihood of the entire pedigree under the different hypotheses
- Final LR determines how probable the pedigree is for H_1 vs H_2

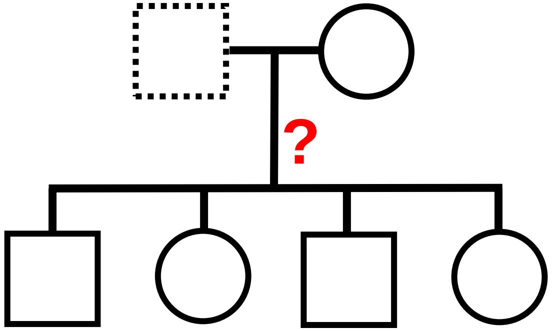


Brenner C., Symbolic kinship program. Genetics (1997) 145:535-42

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

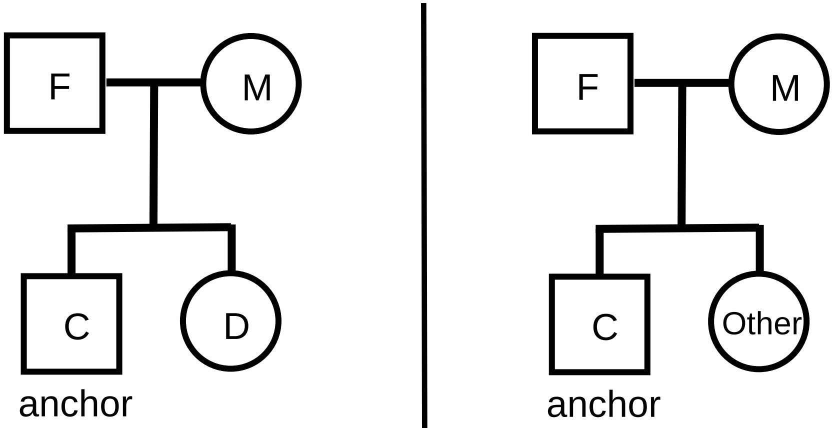
- Refugee/asylum cases
  - 400 applications processed per day
  - 60,108 refugees admitted in 2008
  - 34,753 were relatives of applicant



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

Chris Miles, DHS

Immigration Scenario



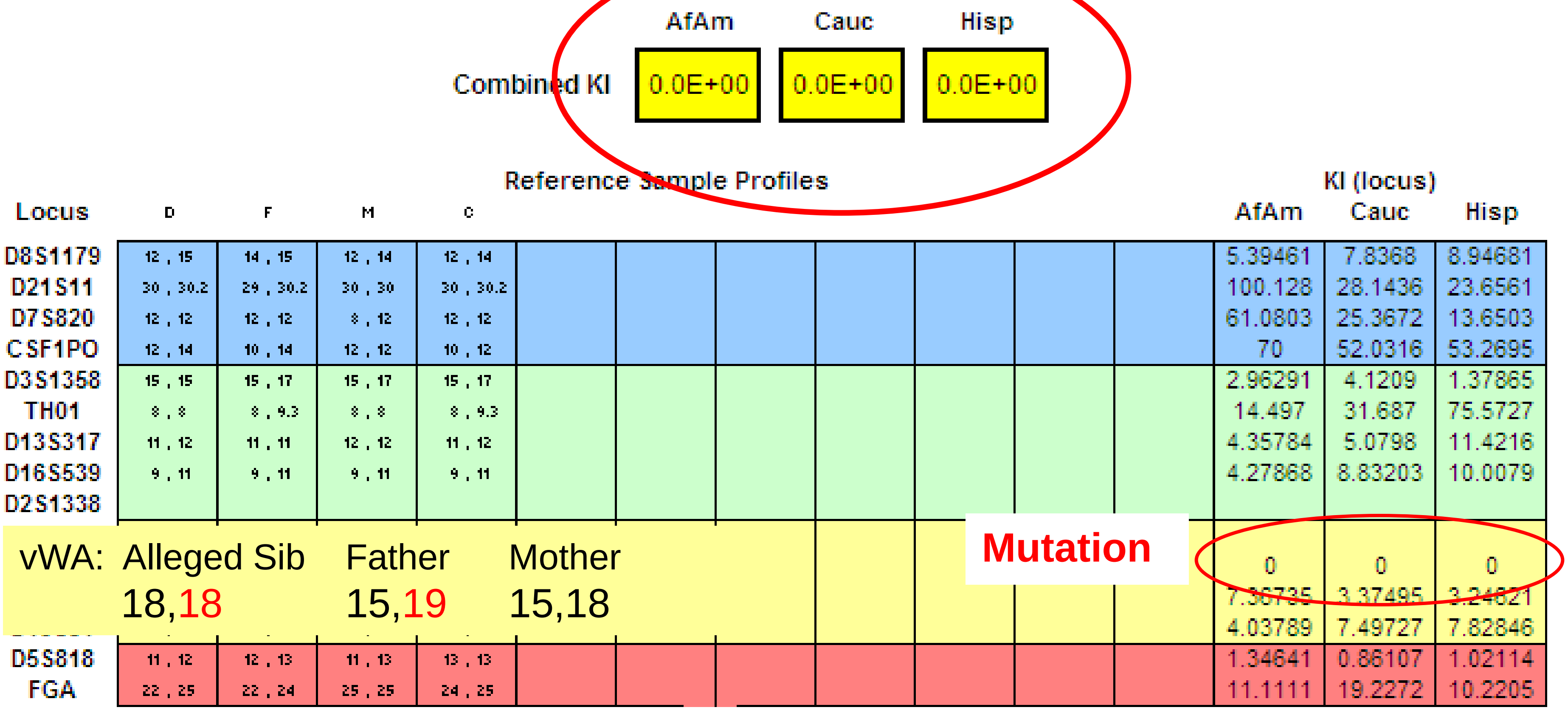
Hypothesis 1
C and D are full siblings

Hypothesis 2
D is unrelated to C

We know the genotypes for each individual.

Compute the LR for this situation using software programs.

Combined likelihood ratio



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

Reference Sample Profiles

Locus	D	F	M	C								KI (locus) AfAm	KI (locus) Cauc	KI (locus) 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	3.24821
												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

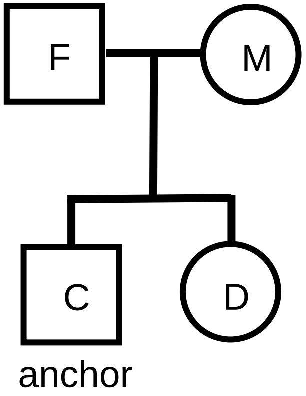
Mutation

Per locus likelihood ratios

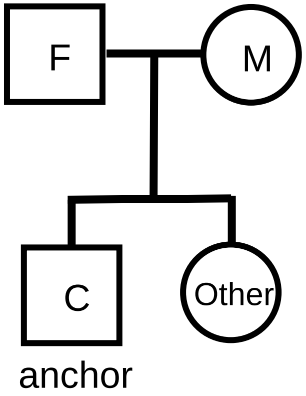
KIn CALc

Steven Myers, CA DOJ

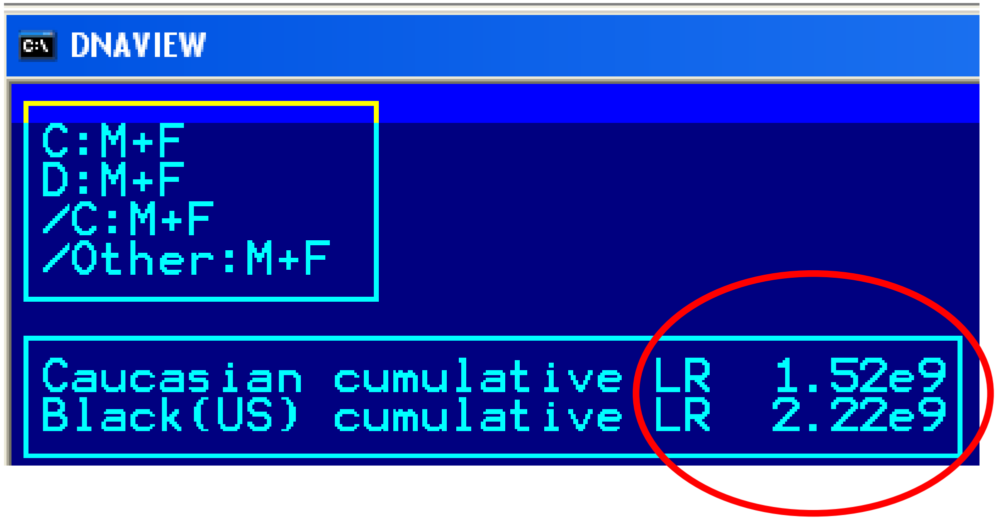
Immigration Scenario



Hypothesis 1
C and D are full siblings



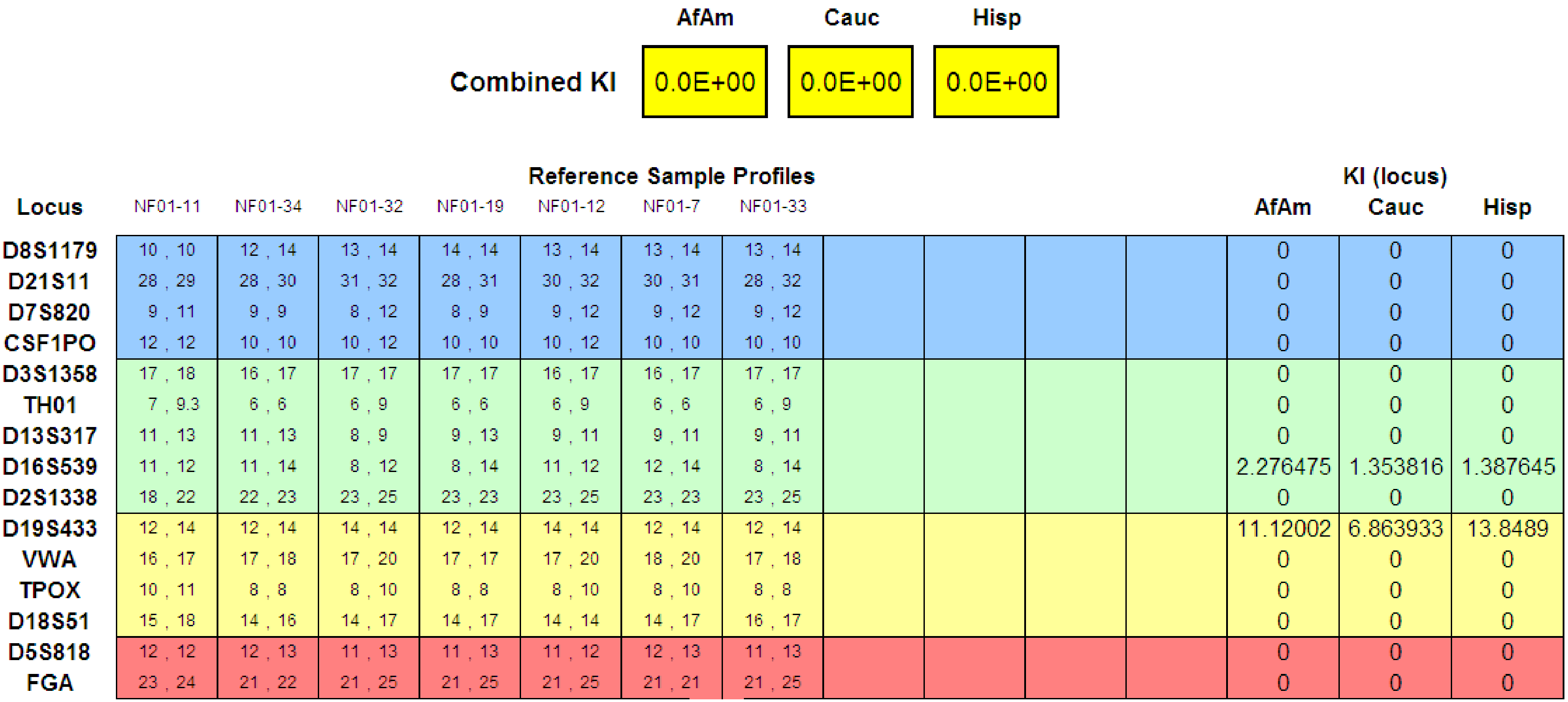
Hypothesis 2
D is unrelated to C



```
C:M+F
D:M+F
/C:M+F
/Other:M+F
```

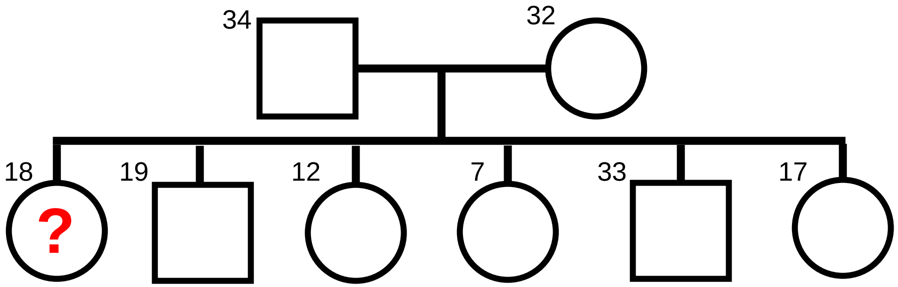
```
Caucasian cumulative LR   1.52e9
Black(US) cumulative LR   2.22e9
```

Complex Pedigree Analysis (1)



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

	Reference Sample Profiles											KI (locus)		
Locus	NF01-11	NF01-34	NF01-32	NF01-19	NF01-12	NF01-7	NF01-33					**AfAm**	**Cauc**	**Hisp**
D8S1179	10 , 10	12 , 14	13 , 14	14 , 14	13 , 14	13 , 14	13 , 14					0	0	0
D21S11	28 , 29	28 , 30	31 , 32	28 , 31	30 , 32	30 , 31	28 , 32					0	0	0
D7S820	9 , 11	9 , 9	8 , 12	8 , 9	9 , 12	9 , 12	9 , 12					0	0	0
CSF1PO	12 , 12	10 , 10	10 , 12	10 , 10	10 , 12	10 , 10	10 , 10					0	0	0
D3S1358	17 , 18	16 , 17	17 , 17	17 , 17	16 , 17	16 , 17	17 , 17					0	0	0
TH01	7 , 9.3	6 , 6	6 , 9	6 , 6	6 , 9	6 , 6	6 , 9					0	0	0
D13S317	11 , 13	11 , 13	8 , 9	9 , 13	9 , 11	9 , 11	9 , 11					0	0	0
D16S539	11 , 12	11 , 14	8 , 12	8 , 14	11 , 12	12 , 14	8 , 14					2.276475	1.353816	1.387645
D2S1338	18 , 22	22 , 23	23 , 25	23 , 23	23 , 25	23 , 23	23 , 25					0	0	0
D19S433	12 , 14	12 , 14	14 , 14	12 , 14	14 , 14	12 , 14	12 , 14					11.12002	6.863933	13.8489
VWA	16 , 17	17 , 18	17 , 20	17 , 17	17 , 20	18 , 20	17 , 18					0	0	0
TPOX	10 , 11	8 , 8	8 , 10	8 , 8	8 , 10	8 , 10	8 , 8					0	0	0
D18S51	15 , 18	14 , 16	14 , 17	14 , 17	14 , 14	14 , 17	16 , 17					0	0	0
D5S818	12 , 12	12 , 13	11 , 13	11 , 13	11 , 12	12 , 13	11 , 13					0	0	0
FGA	23 , 24	21 , 22	21 , 25	21 , 25	21 , 25	21 , 21	21 , 25					0	0	0



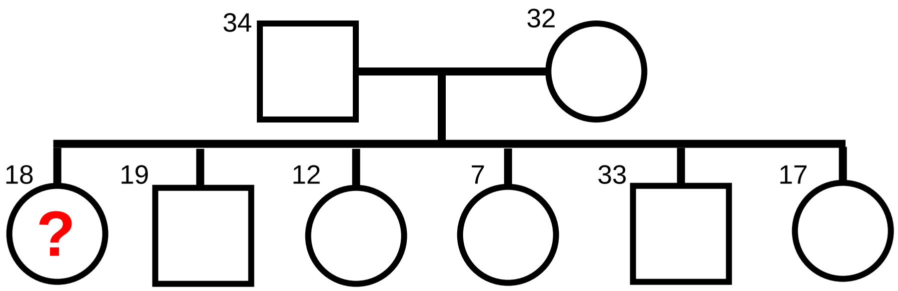
KIn CALc

Steven Myers, CA DOJ

Complex Pedigree Analysis (2)

	AfAm	Cauc	Hisp
Combined KI	9.5E+16	6.2E+13	3.2E+15

	Reference Sample Profiles											KI (locus)		
Locus	NF01-18	NF01-34	NF01-32	NF01-19	NF01-12	NF01-7	NF01-33					**AfAm**	**Cauc**	**Hisp**
D8S1179	12 , 14	12 , 14	13 , 14	14 , 14	13 , 14	13 , 14	13 , 14					3.461538	4.265601	4.205
D21S11	30 , 32	28 , 30	31 , 32	28 , 31	30 , 32	30 , 31	28 , 32					10.01281	4.797203	2.795726
D7S820	8 , 9	9 , 9	8 , 12	8 , 9	9 , 12	9 , 12	9 , 12					9.153176	10.40631	53.26951
CSF1PO	10 , 10	10 , 10	10 , 12	10 , 10	10 , 12	10 , 10	10 , 10					6.786704	7.768687	7.775187
D3S1358	17 , 17	16 , 17	17 , 17	17 , 17	16 , 17	16 , 17	17 , 17					12.5	11.14359	31.10075
TH01	6 , 6	6 , 6	6 , 9	6 , 6	6 , 9	6 , 6	6 , 9					41.68242	9.737476	9.28494
D13S317	9 , 13	11 , 13	8 , 9	9 , 13	9 , 11	9 , 11	9 , 11					35.60111	14.88992	4.134129
D16S539	12 , 14	11 , 14	8 , 12	8 , 14	11 , 12	12 , 14	8 , 14					56.00128	11.45536	18.17815
D2S1338	23 , 23	22 , 23	23 , 25	23 , 23	23 , 25	23 , 23	23 , 25					25.60973	13.7442	16.46041
D19S433	12 , 14	12 , 14	14 , 14	12 , 14	14 , 14	12 , 14	12 , 14					11.12002	6.863933	13.8489
VWA	17 , 18	17 , 18	17 , 20	17 , 17	17 , 20	18 , 20	17 , 18					5.009276	2.143511	2.897961
TPOX	8 , 10	8 , 8	8 , 10	8 , 8	8 , 10	8 , 10	8 , 8					7.272894	12.43107	13.44858
D18S51	14 , 14	14 , 16	14 , 17	14 , 17	14 , 14	14 , 17	16 , 17					61.24764	8.307958	8.655535
D5S818	13 , 13	12 , 13	11 , 13	11 , 13	11 , 12	12 , 13	11 , 13					4.183884	11.7036	27.09336
FGA	21 , 21	21 , 22	21 , 25	21 , 25	21 , 25	21 , 21	21 , 25					16	8.307958	14.67035



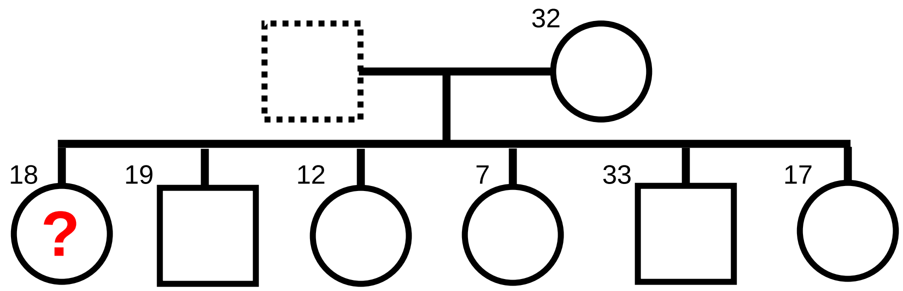
KIn CALc

Steven Myers, CA DOJ

Complex Pedigree Analysis (3)

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

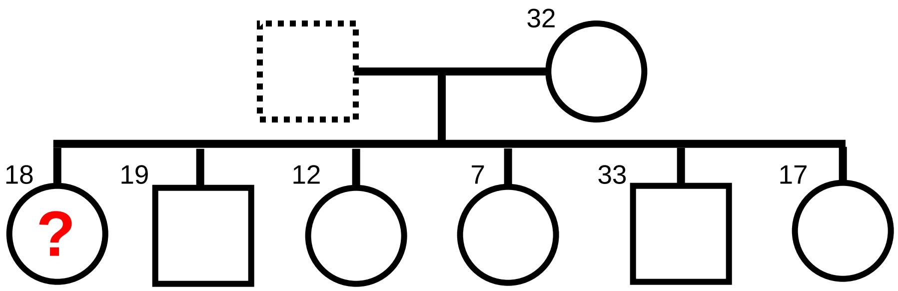
	Reference Sample Profiles											KI (locus)	
Locus	NF01-11	NF01-32	NF01-19	NF01-12	NF01-7	NF01-33						**AfAm**	**Cauc**	**Hisp**
D8S1179	10 , 10	13 , 14	14 , 14	13 , 14	13 , 14	13 , 14						0	0	0
D21S11	28 , 29	31 , 32	28 , 31	30 , 32	30 , 31	28 , 32						0	0	0
D7S820	9 , 11	8 , 12	8 , 9	9 , 12	9 , 12	9 , 12						0	0	0
CSF1PO	12 , 12	10 , 12	10 , 10	10 , 12	10 , 10	10 , 10						0.520833	0.495932	0.402338
D3S1358	17 , 18	17 , 17	17 , 17	16 , 17	16 , 17	17 , 17						0	0	0
TH01	7 , 9.3	6 , 9	6 , 6	6 , 9	6 , 6	6 , 9						0	0	0
D13S317	11 , 13	8 , 9	9 , 13	9 , 11	9 , 11	9 , 11						0	0	0
D16S539	11 , 12	8 , 12	8 , 14	11 , 12	12 , 14	8 , 14						2.276475	1.353816	1.387645
D2S1338	18 , 22	23 , 25	23 , 23	23 , 25	23 , 23	23 , 25						0	0	0
D19S433	12 , 14	14 , 14	12 , 14	14 , 14	12 , 14	12 , 14						11.12002	6.863933	13.8489
VWA	16 , 17	17 , 20	17 , 17	17 , 20	18 , 20	17 , 18						0	0	0
TPOX	10 , 11	8 , 10	8 , 8	8 , 10	8 , 10	8 , 8						0.347187	0.687504	0.748578
D18S51	15 , 18	14 , 17	14 , 17	14 , 14	14 , 17	16 , 17						0	0	0
D5S818	12 , 12	11 , 13	11 , 13	11 , 12	12 , 13	11 , 13						0	0	0
FGA	23 , 24	21 , 25	21 , 25	21 , 25	21 , 21	21 , 25						0	0	0



Complex Pedigree Analysis (4)

	AfAm	Cauc	Hisp
Combined KI	8.7E+14	5.5E+11	1.2E+13

	Reference Sample Profiles											KI (locus)		
Locus	NF01-18	NF01-32	NF01-19	NF01-12	NF01-7	NF01-33						**AfAm**	**Cauc**	**Hisp**
D8S1179	12 , 14	13 , 14	14 , 14	13 , 14	13 , 14	13 , 14						0.076705	0.129605	0.1015
D21S11	30 , 32	31 , 32	28 , 31	30 , 32	30 , 31	28 , 32						10.01281	4.797203	2.795726
D7S820	8 , 9	8 , 12	8 , 9	9 , 12	9 , 12	9 , 12						7.316314	8.226783	34.27196
CSF1PO	10 , 10	10 , 12	10 , 10	10 , 12	10 , 10	10 , 10						5.695983	6.426598	6.377532
D3S1358	17 , 17	17 , 17	17 , 17	16 , 17	16 , 17	17 , 17						12.5	11.14359	31.10075
TH01	6 , 6	6 , 9	6 , 6	6 , 9	6 , 6	6 , 9						29.13263	7.73316	7.580934
D13S317	9 , 13	8 , 9	9 , 13	9 , 11	9 , 11	9 , 11						35.60111	14.88992	4.134129
D16S539	12 , 14	8 , 12	8 , 14	11 , 12	12 , 14	8 , 14						56.00128	11.45536	18.17815
D2S1338	23 , 23	23 , 25	23 , 23	23 , 25	23 , 23	23 , 25						36.02725	20.39331	24.66132
D19S433	12 , 14	14 , 14	12 , 14	14 , 14	12 , 14	12 , 14						11.12002	6.863933	13.8489
VWA	17 , 18	17 , 20	17 , 17	17 , 20	18 , 20	17 , 18						5.009276	2.143511	2.897961
TPOX	8 , 10	8 , 10	8 , 8	8 , 10	8 , 10	8 , 8						6.76564	11.90222	12.8936
D18S51	14 , 14	14 , 17	14 , 17	14 , 14	14 , 17	16 , 17						61.24764	8.307958	8.655535
D5S818	13 , 13	11 , 13	11 , 13	11 , 12	12 , 13	11 , 13						2.022977	3.074218	5.031624
FGA	21 , 21	21 , 25	21 , 25	21 , 25	21 , 21	21 , 25						22.21359	12.58377	19.99133



Bayesian Statistics

Estimating Relationship from Genotypes

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

Bayes Theorem

$$\text{Posterior odds} = \frac{\Pr(H_p|E)}{\Pr(H_d|E)}$$

What we want to know!

$$\text{Posterior odds} = \quad \text{LR} \times \text{Prior odds}$$

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \quad \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

What we can measure!

Bayes Theorem Example

- Airplane crash
- 300 victims (unidentified remains)
- DNA profile from parent of one of the victims
- Prior odds could be 1/300 for a parent-child relationship for each of the 300 remains

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

Accounting for Background Relatedness

Background Relatedness

- Any two individuals in a finite population are related
  - Must have a common ancestor at some point in the past
- Any relatedness between individuals occurs against a background level of relatedness in the population
  - In humans, background relatedness is low but should be considered

Background Relatedness

- Population substructure
  - Groups of individuals within a population that have some degree of reproductive isolation
  - Allele frequencies are likely to be different from the population as a whole

- Can use population-wide allele frequency estimates with θ correction

Theta (θ) Correction

- Probability that any 2 alleles in the same subpopulation are IBD
- Equal to coancestry coefficient (θ) for any 2 individuals
- Equal to the inbreeding coefficient (*F*) of any single individuals in that group

Conditional Match Probability

- Probability that an unknown member of a population (e.g., perpetrator of crime) would have the profile given that a known person (e.g., suspect who has been wrongly accused) has been found to be of that type

- Probability that an untyped relative will have a certain genotype, given the observed genotype of a relative

Theta (θ) Correction in Forensics

$$\Pr(A_iA_i|A_iA_i) = \frac{[3\theta+(1-\theta)P_i][2\theta+(1-\theta)P_i]}{(1+\theta)(1+2\theta)}$$

$$\Pr(A_iA_j|A_iA_j) = \frac{2[\theta+(1-\theta)P_i][\theta+(1-\theta)P_j]}{(1+\theta)(1+2\theta)}$$

Conditional match probabilities are greater than unconditional profile probabilities $\Pr(A_iA_i)$ and $\Pr(A_iA_j)$

References

Weir B., Anderson A., Hepler A., Genetic relatedness analysis: Modern data and new challenges. Nat. Rev. Genet. (2006) 7:771-80

Weir B., Anderson A., Interpreting DNA Evidence. Summer Institute in Statistical Genetics, Seattle, WA (2006)

Buckleton J., Triggs C., Walsh S, Forensic DNA Evidence Interpretation. CRC Press (2005)

Fung W., Hu, Y-Q, Statistical DNA Forensics. Theory, Methods and Computation. Wily (2008)

Thank you for your attention

Contact:

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

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