Forensic Genetics and Legal Medicine 2019-2020

11th May 2020

Kinship testing (complex cases)

Complex parentage cases

✓ Deficiency case: mother's data missing; maternity testing

✓ Deficiency case: alleged parent's (AP) data missing

- Analysis of AP's archival material
- Exhumation of AP's body
- Reconstruction (derivation of AP's obligate alleles/haplotypes from available relatives; will often include lineage markers and X-STRs)

✓ Immigration cases: tested subjects are related, but degree of kinship could be different from what declared

Deficiency case (mother's data are missing)*

Child and alleged father are both "PQ"

Pr (G / P) = ???

Pr (F_{PQ}/P_{PQ}) = Pr (F_{PQ} and P_{PQ}) / Pr (P_{PQ})

Pr (G / N) = ???

Pr (F_{PQ})

Possible fathers	Freq. possible fathers	Freq. Children PP	Freq. children PQ	Freq. Children QQ
PP	p^2	p^3	p^2q	-
PQ	2pq	p^2q	pq	pq^2
QQ	q^2	-	pq^2	q^3

*the same applies for maternity tests, in which paternity and paternal alleles cannot be assumed a priori

Pr (G / P) = Pr (F_{PQ}/P_{PQ}) = Pr (F_{PQ} and P_{PQ}) / Pr (P_{PQ}) = pq / 2pq = 1/2

Pr (G / N) = Pr (F_{PQ}) = 2pq

LR = 1/(4pq)

For some genotype combinations, LR and PP values in deficiency cases are reduced, compared to standard trios (if p=0.2, LR=1.56 and PP=63%)

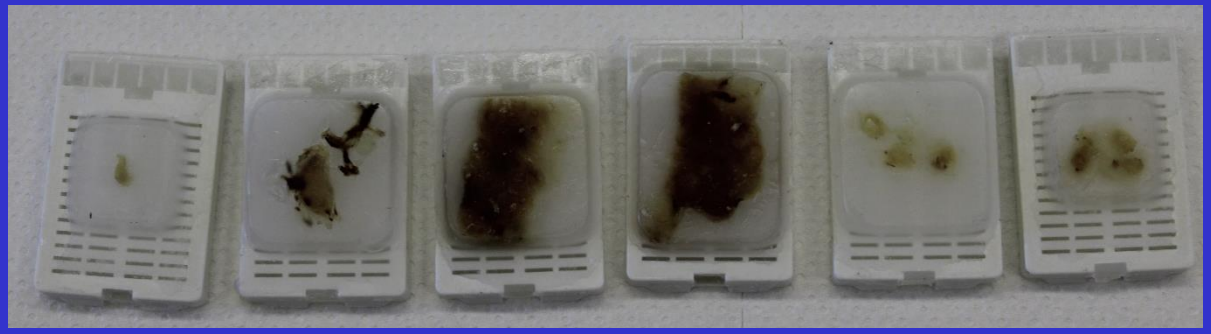
child

alleged father

C	AF	Numerator	Denominator	PI
AB	AC	0.5b	2ab	0.25/a
AB	AB	0.5(a+b)	2ab	(a+b)/4ab
AB	A	b	2ab	0.5/a
A	AC	0.5a	a^2	0.5/a
A	A	a	a^2	1/a

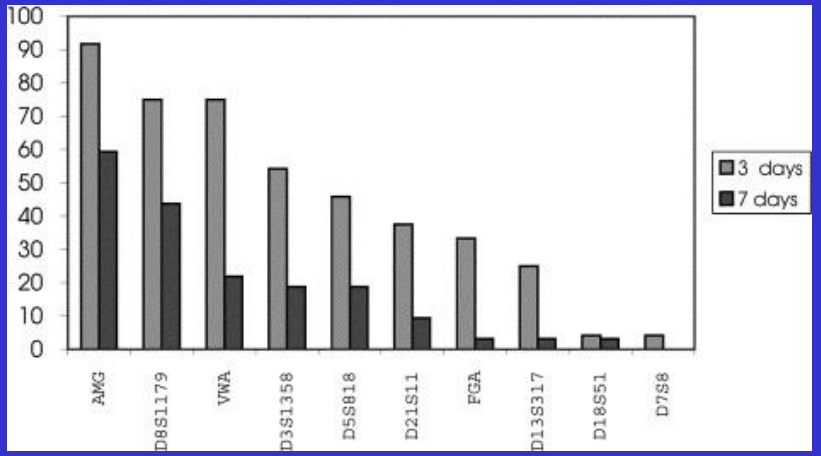
Average PI value with a standard set of 16 STRs in duo cases ~ 2 x 10^7

Archival samples



Formalin fixed paraffin embedded (FFPE) samples are often available for deceased alleged fathers.
Italian legislation requires that FFPE samples collected in hospitals are stored at least for 20 years.

Formalin is a potent and rapid DNA degrading reagent



Percentages of successful multiplex PCR after 3 and 7 days formalin fixation (Legrand et al. Forensic Sci Int 2002)

Introduction of molecular methods in pathology meant that (starting from the '90s) formalin fixation time was normally reduced, at least in clinical settings (forensic pathologists still tend to «forget» samples in formalin for longer periods…)

NB Tumor tissue FFPE samples should be avoided, since often affected by genotype anomalies («microsatellite instability», «loss of heterozigosity»), similar in apeearance to allele drop out or drop in.

Exhumation

- ✓ Soft tissues, if still available (DNA is more prone degradation, but sometimes unpredictably intact)

«corification» in zinc coffins

- ✓ Bone

- ✓ Teeth

- ✓ Other tissues and body fluids sporadically described in the literature as good sources of DNA in decomposed bodies(nails, vitreous humor)….

Which bone?

Best target in the cranium according to literature (Kulstein et al. Int J Legal Med 2018): inner part of the petrous bone

Highest bone density is found in the femur diaphysis: maximum protection against degrading agents

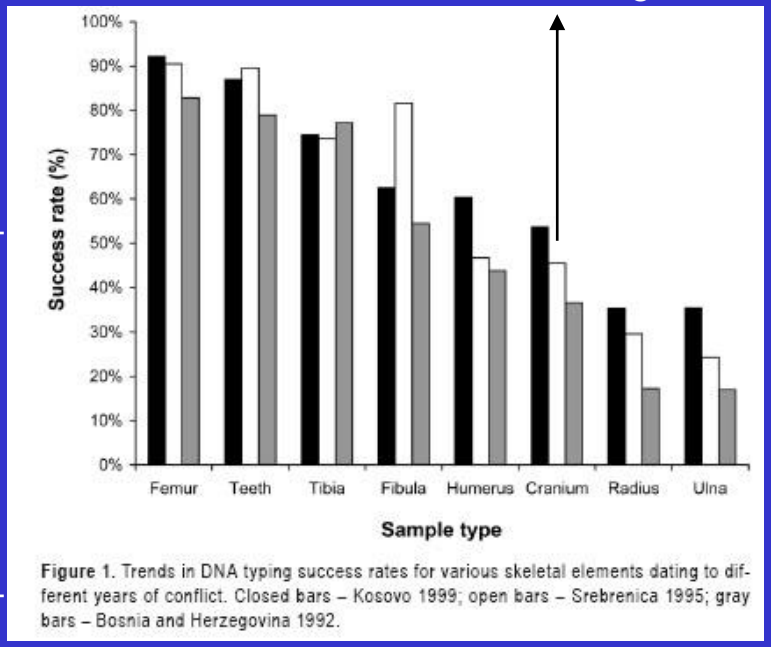
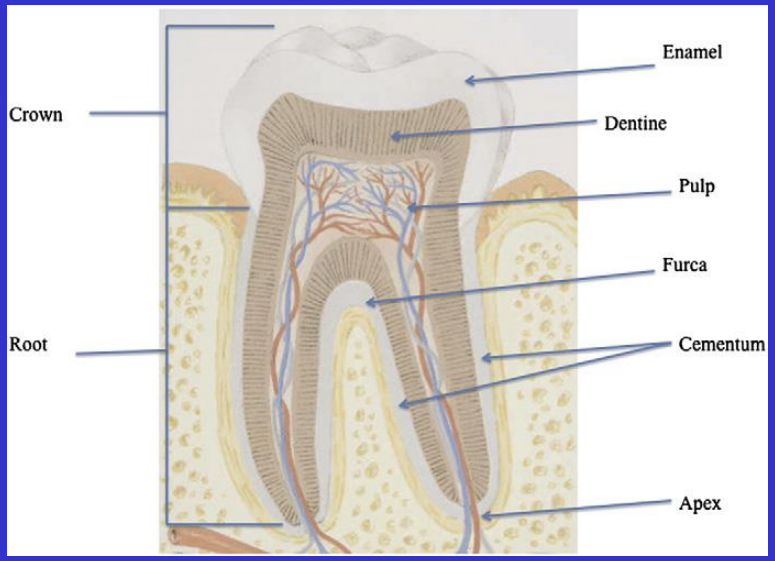


Figure 1. Trends in DNA typing success rates for various skeletal elements dating to different years of conflict. Closed bars – Kosovo 1999; open bars – Srebrenica 1995; gray bars – Bosnia and Herzegovina 1992.

NB data are for diaphysis, not epiphysis made of spongy bone that does not effectively preserve DNA

Milos et al Croat Med J 2007

Which tooth?



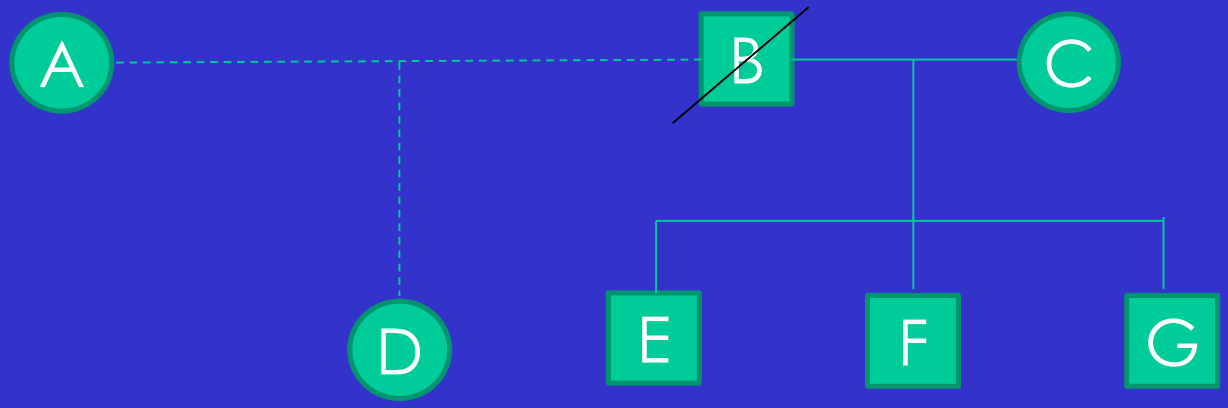
→ hardest tissue in the human body, 96% mineral, acellular

→ highly cellular

↓

Healthy teeth with larger pulp volume are preferred: molars > canines > premolars

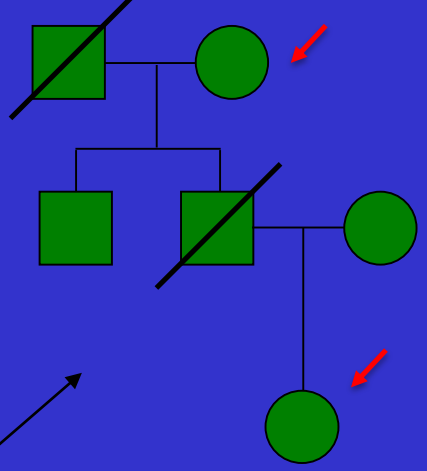
Reconstruction



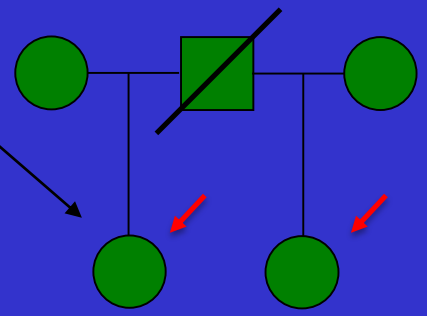
Y-STR testing can give support to the assumption that E, F and G are full sibs, and therefore their autosomal STR genotypes can be used to reconstruct those of B, deceased alleged father of D

X-STRs

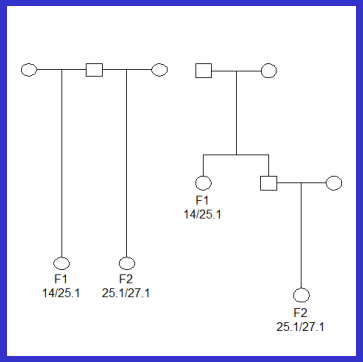
- ✓ X chromosome is transmitted from father to daughters without recombination
- ✓ X chromosome is transmitted from mother to children after regular recombination
- ✓ Analysis of X-STRs is particularly useful in specific deficiency cases
- ✓ Analysis of X-STRs is particularly useful to discriminate between some alternative pedigrees (immigration cases, disaster victim identification)



Alleged paternal grandmother and granddaughter



Alleged paternal half-sisters



	N of Identical By Descent (IBD) alleles		
	0	**1**	**2**
Autosomal STRs			
Paternal half-siblings	0.50	0.50	0
Avuncular	0.50	0.50	0
X-STRs			
Paternal half-siblings	0	1	0
Avuncular	0.50	0.50	0

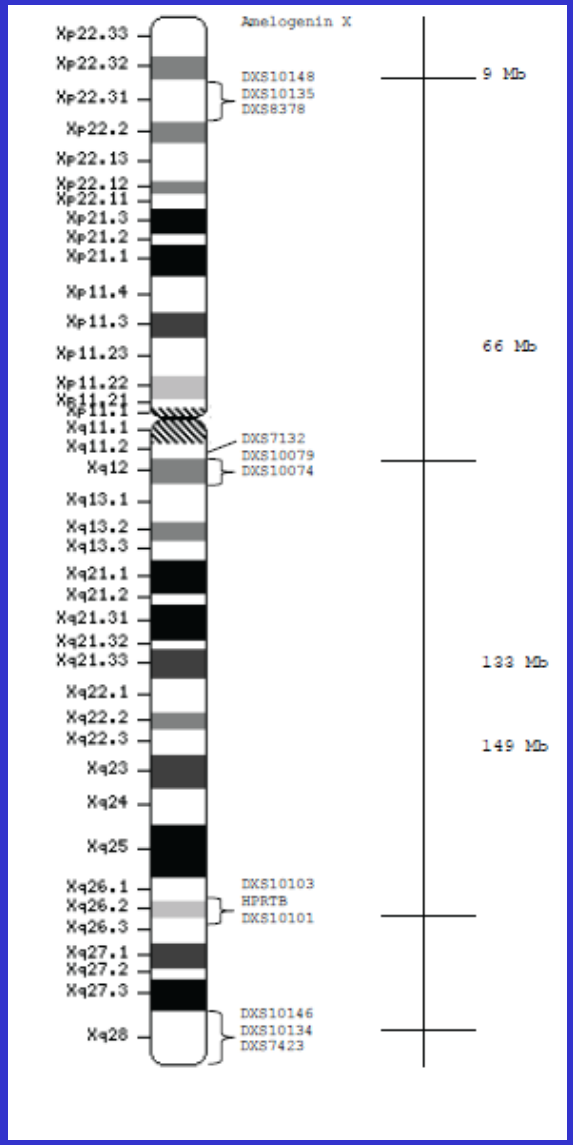
X-STRs

- X chromosome is ~150 Mb long. Since ~50 Mb distance between a pair of markers is required to assume independence (50% chance of crossing over), a test with no more than 4 independent X-STRs is possible…

- Recombination rates for X-linked STR markers are required to include >4 markers in calculations

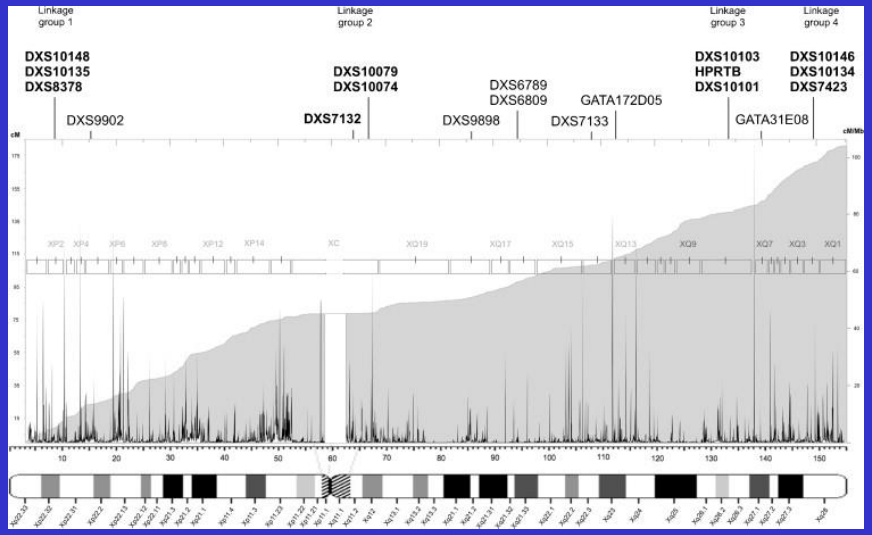
- A possible solution is to use clusters of closely linked markers, with recombination rare within clusters and near independence between clusters

- Haplotypes formed by closely linked X-STR markers can be in linkage equilibrium (LE) or in linkage disequilibrium (LD)

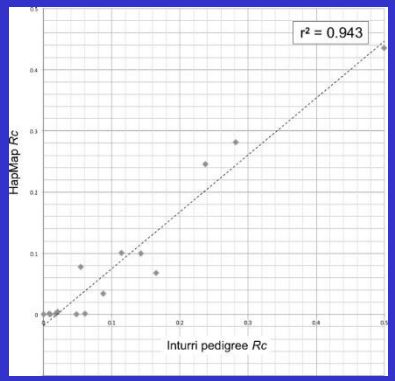


Recombination

✓ Recombination rates are not homogeneous across the X chromosome. So recombination fractions cannot be straightly derived from physical maps.



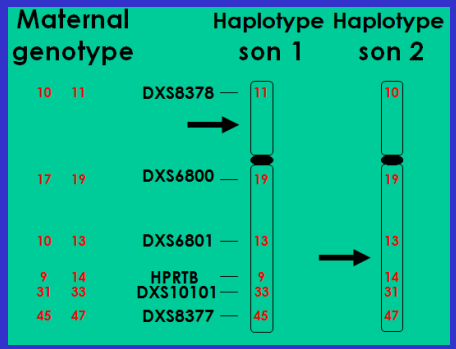
- Recombination events can be directly measured through large studies of "informative" pedigrees (e.g.: mother and two or more sons)…

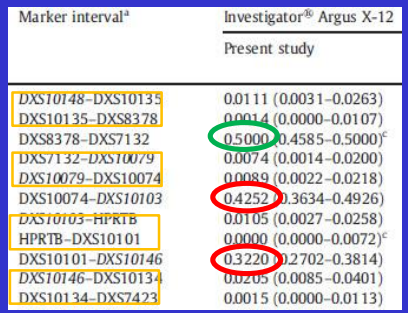


Inturri et al. Forensic Sci Int Genet 2011

Good agreement between the two estimation methods

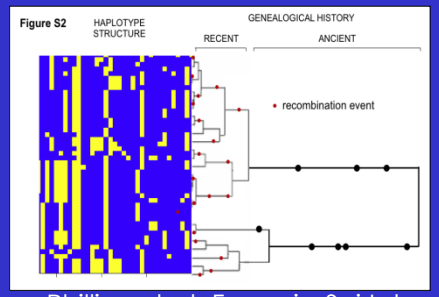
- …or derived from high density SNP data using a coalescent model





Marker interval[a]	Investigator® Argus X-12 Present study
DXS10148–DXS10135	0.0111 (0.0031–0.0263)
DXS10135–DXS8378	0.0014 (0.0000–0.0107)
DXS8378–DXS7132	0.5000 (0.4585–0.5000)[c]
DXS7132–*DXS10079*	0.0074 (0.0014–0.0200)
DXS10079–DXS10074	0.0089 (0.0022–0.0218)
DXS10074–*DXS10103*	0.4252 (0.3634–0.4926)
DXS10103–HPRTB	0.0105 (0.0027–0.0258)
HPRTB–DXS10101	0.0000 (0.0000–0.0072)[c]
DXS10101–*DXS10146*	0.3220 (0.2702–0.3814)
DXS10146–DXS10134	0.0205 (0.0085–0.0401)
DXS10134–DXS7423	0.0015 (0.0000–0.0113)

Nothnagel et al. Forensic Sci Int Genet 2012



Phillips et al. Forensic Sci Int Genet 2011

Linkage disequilibrium

X STRs A - B - C
12-29-32
12-34-33
13-35-27
14-30-29
10-32-40
10-31-34
14-29-35
14-30-29
11-28-29
12-29-29
13-35-37
12-29-32
12-30-33
11-28-33
11-30-30
12-28-29

→ Observed haplotype frequency 2/16 = 0.125
Expected haplotype frequency = p12 * p29 * p32 = (6/16 = 0.375) * (4/16 = 0.25) * (2/16 = 0.125) = 0.012

↓

Large differences between observed and expected haplotype frequencies may indicate LD

- Large haplotipic databases of X-STR clusters are needed to estimate LD
- LD can be population specific (effect of genetic drift on the X chromosome stronger than on autosomal STRs, though smaller compared to Y-STRs and mtDNA)
- In case of LD, haplotype frequencies rather than allele frequencies are needed in calculations
- Need for dedicated software capable to accomodate recombination rates, LD, mutation…

Kinship testing in disaster victim identification (DVI)

- ✓ Rarely, in DVI, reliable antemortem DNA samples of missing individuals are available. Identification is normally done via kinship testing of the missings' relatives.
- ✓ Assumptions on priors are necessary for decision making: a probability of identification (POI) threshold has to be set

I = identity, the tested human remains are from the missing relative of tested references

$$\frac{Pr(I \mid G)}{Pr(N \mid G)} = \frac{\mathbf{Pr(G \mid I)}}{\mathbf{Pr(G \mid N)}} \times \frac{Pr(I)}{Pr(N)}$$

Posterior odds → $\dfrac{\text{Posterior odds}}{\text{Posterior odds} + 1}$ = POI

- ✓ The POI threshold is set by the authorities in charge of the identification process as a trade-off between the risk for a sample to be wrongly identifying or remain unidentified
- ✓ This does not depend solely on the power of the DNA test (average expected LR*) but also on the number n of victims of the mass disaster

*also depending on the type of reference samples (e.g. mother better than sib, etc.)

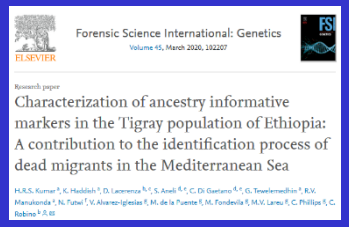
1/3000 priors were estimated in 2001 WTC attack and a posterior odds threshold of 1000 (POI > 99.9%) was set for identification (LR ≥ 3 x 10^6)

Posterior odds = LR * 1/n-1

known total number of missing persons or victims

Adjusted through additional forensic Information:

- Sex
- Age
- Postmortem interval
- Ancestry
- …


Forensic Science International: Genetics
Volume 45, March 2020, 102207

Research paper

Characterization of ancestry informative markers in the Tigray population of Ethiopia: A contribution to the identification process of dead migrants in the Mediterranean Sea

Estimation of n is difficult in «open» mass disasters (e.g. war or genocide)

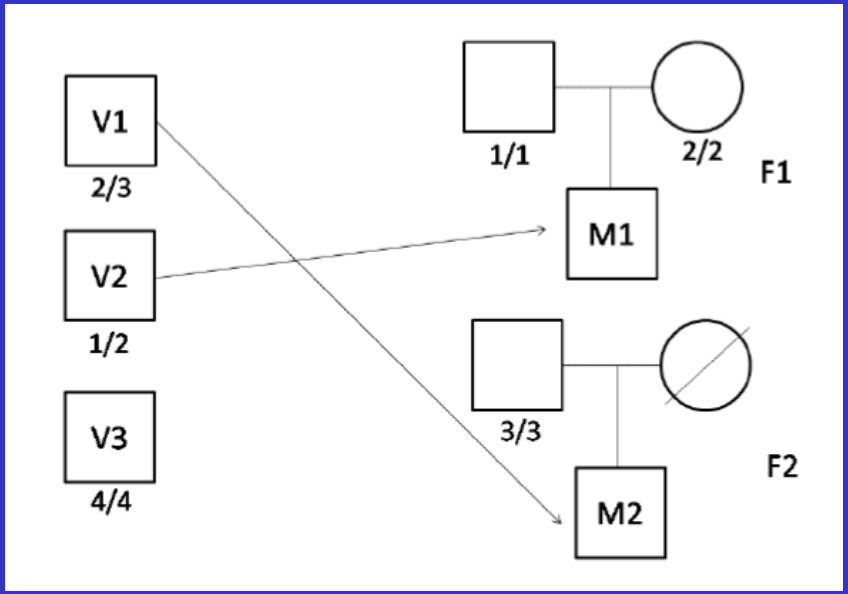
Estimation of n is easy in «closed» mass disasters (e.g. plane crash with passengers' list)

$$PPi = O'i / \Sigma j\ O'j$$

Posterior probability that sample belongs to missing individual i (POI)

Posterior odds thaf sample belongs to missing individual of reference family i

Sum of posterior odds that sample belongs to missing individual of reference families 1,…i,…j



Reference family	Body	Prior odds	LR	Posterior odds
F1	V1	1/4	0	0
F1	V2	1/4	8 1/2x0.25x0.25	2
F1	V3	1/4	0	0
F1	V4	1/4	1	1/4
F2	V1	1/4	2 0.25/2x0.25x0.25	1/2
F2	V2	1/4	0	0
F2	V3	1/4	0	0
F2	V4	1/4	1	1/4

- ✓ naval accident with n missing victims, 3 bodies found (V1,V2,V3), reference families (F1,F2) available for 2 of the missings (M1,M2)
- ✓ One STR with four alleles tested (assumed frquency of each allele in the population 0.25)

Equal priors for the missing being V1, V2, V3 or a body not yet found

POI (V1,F2) = 0.5/0.75=66.6%

POI (V2,F1) = 2/2.25=88.8%

Familial searching in criminal DNA databases

- ✓ if a perpetrator is not recorded on the database, then no match will result. However close relatives e.g. brother or father will have many alleles in common.

- Such strategy clearly has ethical implications. In some national database (UK) familial searching is routine activity. In The Netherlands a specific law allowing and regulating familial searching was issued in 2012. In the US some federal states (California, Colorado…) support familial searching, while it was banned in others (e.g. Maryland). However, most national DNA databases do not have specific regulations.

Possible strategies

- ✓ **Identical by state (IBS) alleles**
- compares the number of shared alleles between the forensic profile and the candidate profile(s) from a database

TABLE 2—*IBD distributions for relevant pairwise kinships for familial searching.*

	IBD		
Pairwise Kinship	Φ_2	Φ_1	Φ_0
Unrelated	0	0	1
Parent–child	0	1	0
Full-sib	1/4	1/2	1/4

IBD, Identity-By-Descent.

Ge et al. J Forensic Sci 2011

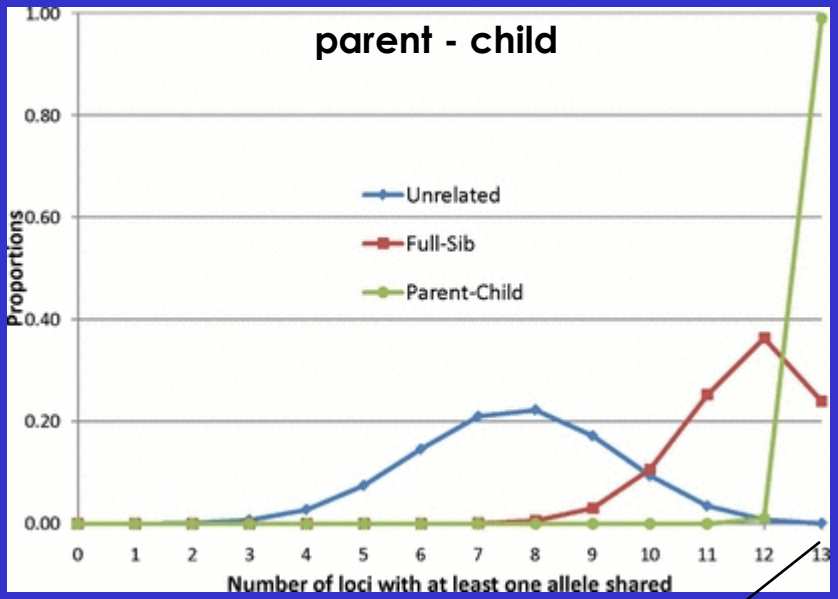
TABLE 3—*Expected probability of two individuals sharing i number of alleles (i.e., IBS) given IBD = j, $P_i(\Phi_j)$, where* i *is IBS* (i = 0, 1, 2) *and j is IBD* (j = 0, 1, 2). *α_r is the sum of the r-th power of allele frequencies at the locus, namely, $a_r = \sum_{i=1}^{k} p_i^r$, where p_i is the allele frequency, k is the number of alleles at this locus. ϑ is the population substructure parameter.*

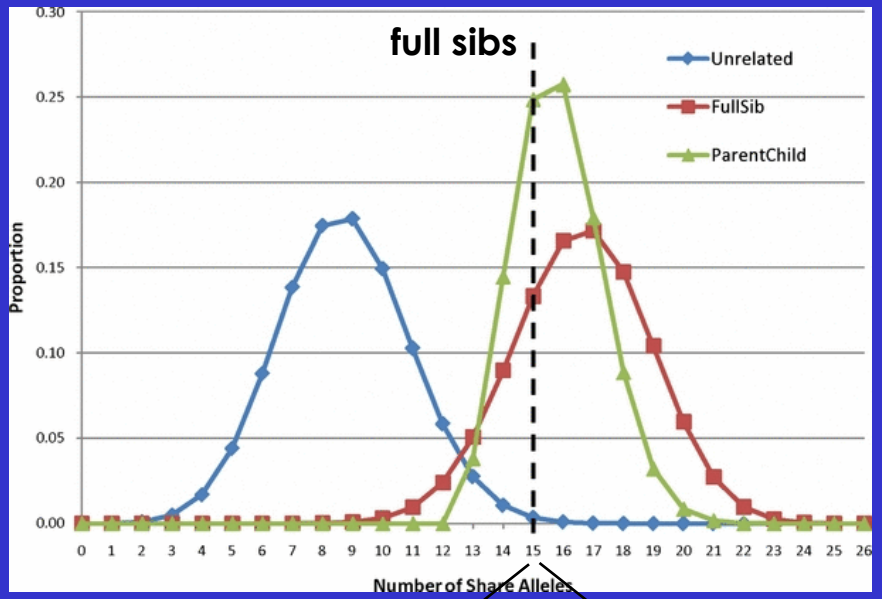
		IBD	
IBS	Φ_2	Φ_1	Φ_0
0	0	0	$\frac{\theta^2(1-\theta)(1-a_2)+2\theta(1-\theta)^2(1-2a_2+a_3)+(1-\theta)^3(1-4a_2+4a_3+2a_2^2-3a_4)}{(1+\theta)(1+2\theta)}$
1	0	$(1-\theta)$ $(1-a_2)$	$\frac{8\theta^2(1-\theta)(1-a_2)+4\theta(1-\theta)^2(1-a_3)+4(1-\theta)^3(a_2-a_3-a_2^2+a_4)}{(1+\theta)(1+2\theta)}$
2	1	$\theta+(1-\theta)a_2$	$\frac{6\theta^3+\theta^2(1-\theta)(2+9a_2)+2\theta(1-\theta)^2(2a_2+a_3)+(1-\theta)^3(2a_2^2-a_4)}{(1+\theta)(1+2\theta)}$

IBD, Identity-By-Descent; IBS, Identity-By-State.

IBS

✓ Simulations with 13 CODIS STRs for Caucasian population data
(Ge et al. J Forensic Sci 2011)



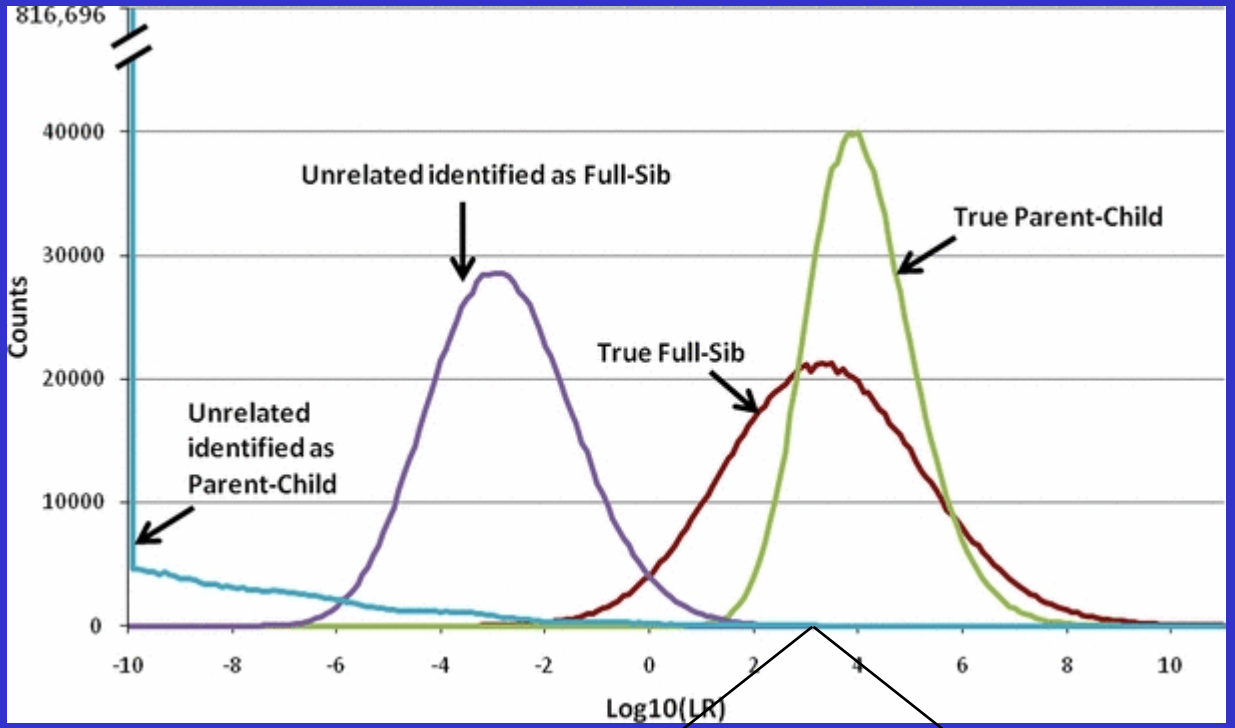


- 0.077% unrelated individuals not excluded as parents (770 possible hits in a database of one million individuals)

- 0.45% unrelated individuals not excluded
- ~18% true parent-child and full sibs excluded

LR

- ✓ probabilities of the forensic and candidate profiles given that the hypothesis the donors are related (parent–child or full-sib) versus unrelated



Ge et al. J Forensic Sci 2011

~0.01% of unrelated pairs had LR > 1,000 in favor of parent-child or full sib relation

~14% of parent-child had LR < 1,000 in favor of parent-child relation
and ~42% of full sibs had LR < 1,000 in favor of full sib relation

Familial searching in criminal DNA databases (STR loci)

- ✓ By combining IBS and LR strategy it is possible to reduce the false positive rate, but the % of false negatives is high

	Unrelated Identified As			
	Parent–Child (%)	Full-sib (%)	True Parent–Child (%)	True Full-sib (%)
IBS ≥ 14; KI ≥ 100	0.0421	0.0742	4.8	23.8
IBS ≥ 14; KI ≥ 1000	0.0128	0.0093	16.6	42.6
IBS ≥ 14; KI ≥ 10,000	0.0014	0.0007	50.5	63.2
IBS ≥ 15; KI ≥ 1000	0.0098	0.0089	27.3	43.1
IBS ≥ 15; KI ≥ 10,000	0.0010	0.0007	21.8	63.3
IBS ≥ 16; KI ≥ 1000	0.0063	0.0075	55.4	45.7
IBS ≥ 16; KI ≥ 10,000	0.0007	0.0007	65.9	63.9
IBS ≥ 16; KI ≥ 100,000	0.0001	0.0003	86.9	80.4

IBS, Identity-By-State. **KI = LR**

false negatives

Ge et al. J Forensic Sci 2011

- ✓ Familial searching strategies based on 15-20 STR loci DNA profiles found in criminal databases have several limitations
- ✓ Search for relationships more distant than parent-child and full sibship is not feasible

How to solve a 40-year-old cold case with 99$

The capture of the Golden State killer