

# Forensic Genetics

Module 19 – Topic 8

# Schedule – Day 3

<b>Topic 7</b>	Y-STR profiles	8:00-8:50
<b>Review</b>	Topic 6 and 7 exercises	9:05-9:55
<b>Topic 8</b>	Incorporating relatives	10:10-11:00
<b>Review</b>	Topic 8 exercises	11:30-12:20
<b>Topic 9</b>	NGS data and other techniques	12:35-1:25
<b>Review</b>	Recap	1:40-2:30

# Relatedness

- Incorporating Relatives
  - Identity By Descent
  - Match Probabilities and LR<sub>s</sub> for relatives
- Applications of Relatedness
  - Paternity Testing
  - Missing Persons
  - Familial Searching
  - Genealogical Searching

# LR Challenges

- A traditional LR considers an alternative proposition with unrelated individuals (which usually favors the prosecution).
  - Where does this individual come from? From the same population and sub-population, from a different sub-population, or a different population?
  - What if someone who is related to the suspect is the source of the DNA sample?
- The LR applies only to one specific defendant.

# Relatives

Because DNA profiles are inherited, relatives are more likely to share a DNA profile than unrelated individuals.

$H_p$ : The DNA in the sample came from the suspect.

$H_d$ : The DNA in the sample came from an unrelated individual.

$H_p$ : The DNA in the sample came from the suspect.

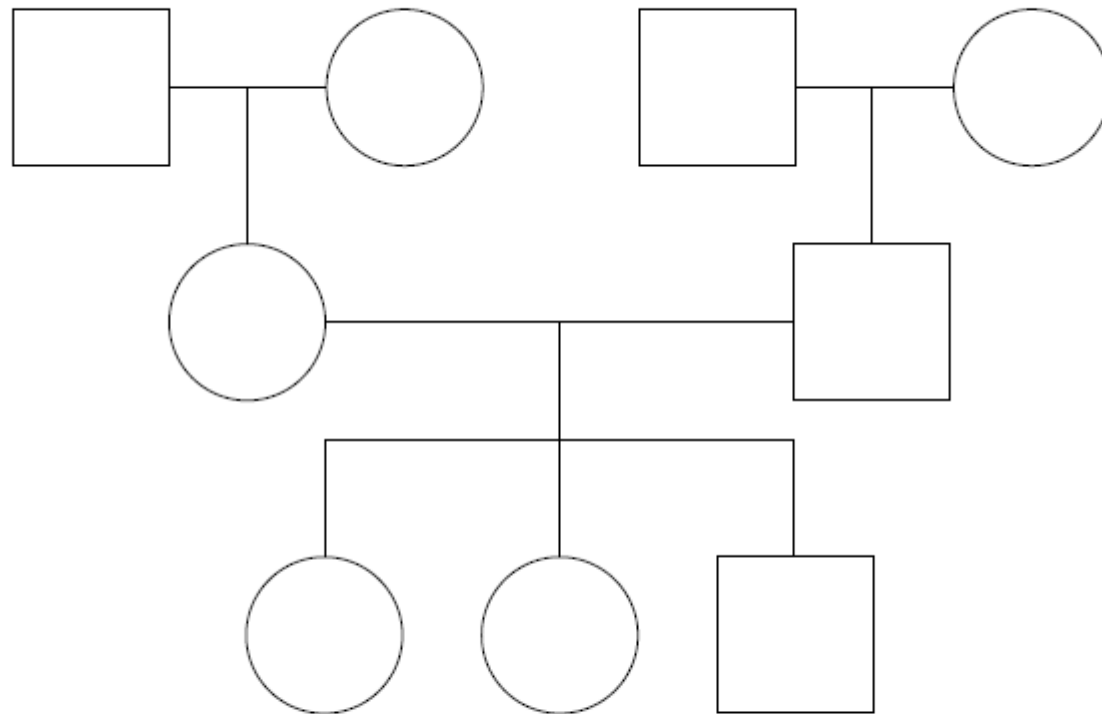
$H_d$ : The DNA in the sample came from a brother of the suspect.

The relationship type can be anything: parent, child, sibling, uncle, cousin, etc.

The more distant the relationship, the closer the value will become to the LR considering unrelated individuals.

# Pedigrees

Pedigrees provide a graphical representation of relationships.



Individuals are said to be related if they share a common ancestor. Relationships can be unilateral (one-sided) or bilateral (two-sided).

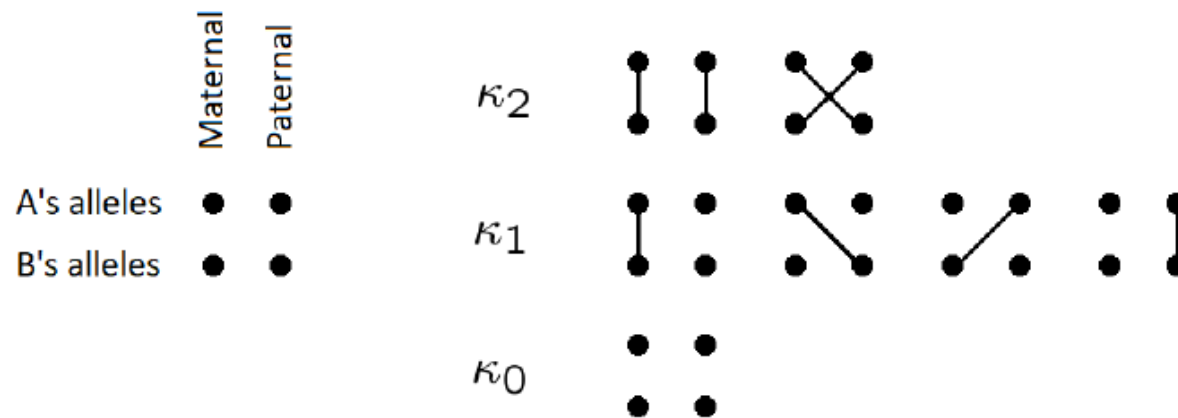
# Identity By Descent

- Relatives are similar because they share alleles that are *identical by descent* (IBD).
- IBD alleles are copies of the same allelic type inherited through a common ancestor (and ignores mutation).
- A pedigree or relationship determines IBD probabilities, which determine probabilities of joint genotypes.

# IBD Coefficients

For *non-inbred* relatives, there are three IBD classes. We write  $\kappa_i$  to denote the IBD probabilities:

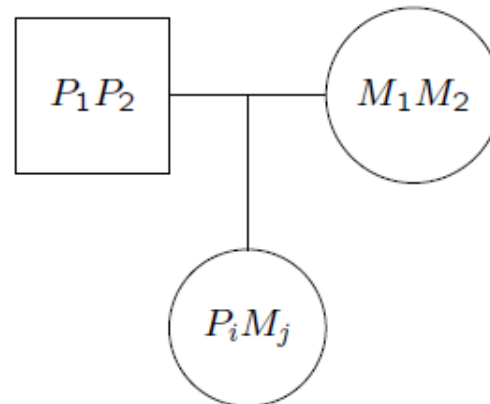
$$\kappa_i = \Pr(i \text{ alleles IBD})$$





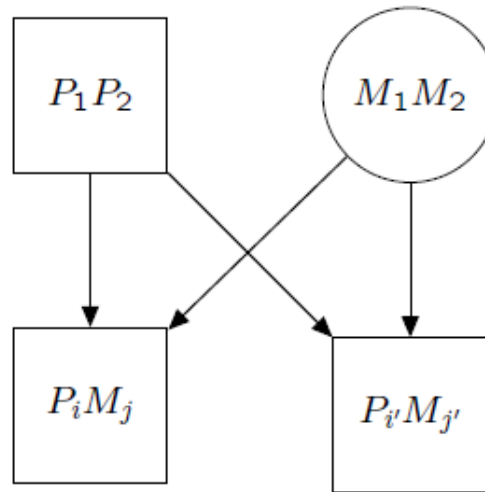
# IBD for Parent-Offspring

- Mendel's law states that one of the two alleles from a parent will be passed down to a child;
- Both alleles have equal probability  $\frac{1}{2}$  of being passed down.



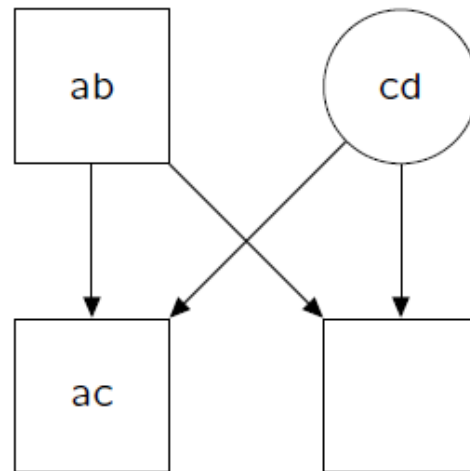
# IBD for Siblings

What about siblings?



# IBD for Siblings

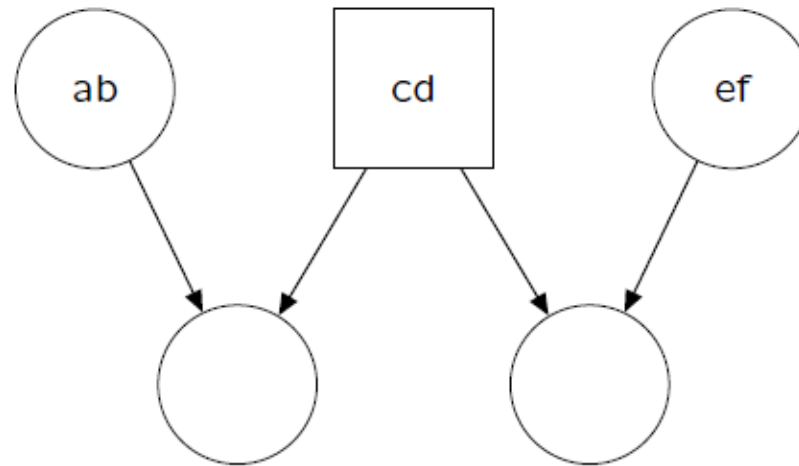
They share either both, one or none of the alleles IBD.



Sib 1	Sib 2	Alleles IBD		
		0	1	2
ac	ac			✓
	bc		✓	
	ad		✓	
	bd	✓		
<b>Total</b>		1/4	1/2	1/4

# IBD for Half-Sibs

What are the IBD coefficients for half-sibs?



# IBD Coefficients

The following table shows IBD probabilities for common relationships:

<b>Relationship</b>	$\kappa_0$	$\kappa_1$	$\kappa_2$
Unrelated	1	0	0
Parent/child	0	1	0
Identical twins	0	0	1
Siblings	1/4	1/2	1/4
Half-sibs	1/2	1/2	0
First cousins	3/4	1/4	0

These IBD probabilities give the expected relatedness between individuals (the realized relatedness is variable).

# Match Probabilities for Relatives

If  $\kappa_0 = 1$ , we are in the original situation and write  $M_2$  for the appropriate match probability:

$$M_2 = \begin{cases} p_A^2, & \text{for homozygous loci } AA, \\ 2p_A p_B, & \text{for heterozygous loci } AB. \end{cases}$$

If  $\kappa_1 = 1$ , the match probability  $M_1$  changes to:

$$M_1 = \begin{cases} p_A, & \text{for homozygous loci } AA, \\ \frac{1}{2}(p_A + p_B), & \text{for heterozygous loci } AB. \end{cases}$$

If  $\kappa_2 = 1$ , both alleles are IBD and the match probability is 1.

# Match Probabilities for Relatives

Combining the terms leads to the overall single-locus match probability for relatives:

$$\kappa_2 + \kappa_1 M_1 + \kappa_0 M_2,$$

which yields a standard match probability of  $M_2$  for unrelated individuals.

# LRs for Relatives

With this approach we can incorporate specific relatives. But what if no specific alternative is available?

$H_d$ : The DNA in the sample came from an unrelated individual.

$H_d$ : The DNA in the sample came from a brother of the suspect.

$H_d$ : The DNA in the sample came from an unknown individual from the population.



# LRs Including Relatives

- We can model a situation where relatives of the suspect make up a small proportion of the total population.
- Background information may be used to assess plausible values for the number of relatives in each category.
- An overall LR can be calculated as a weighted average over the sets.

# The Island Problem

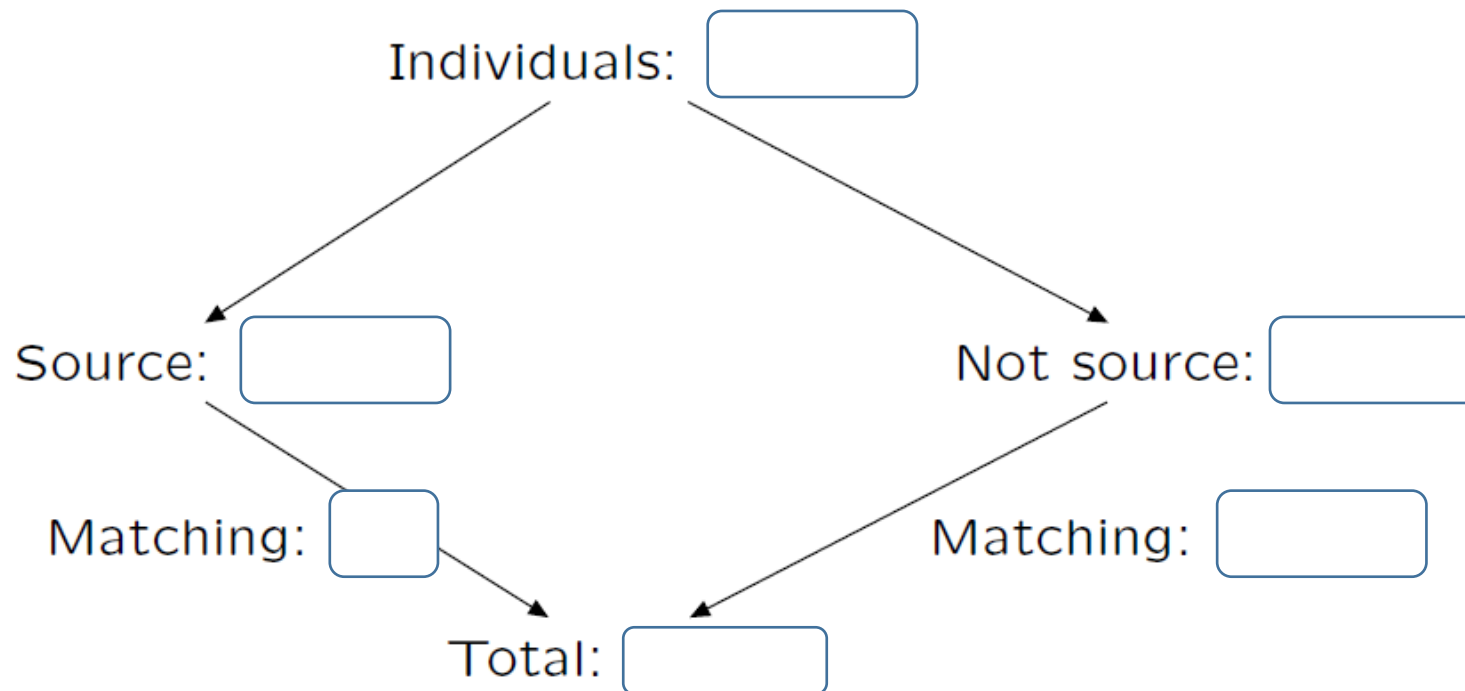
Suppose there is a crime committed on a remote island with a population of size 1001. A suspect  $Q$  is found to match the crime scene profile. What is the probability that  $Q$  is the source of the profile, assuming that:

- All individuals are equally likely to be the source.
- The DNA profiles of all the other individuals are unknown.
- The match probability for unrelated individuals is  $5 \times 10^{-6}$ .

Source: Weight-of-Evidence for Forensic DNA Profiles (Balding, 2015)

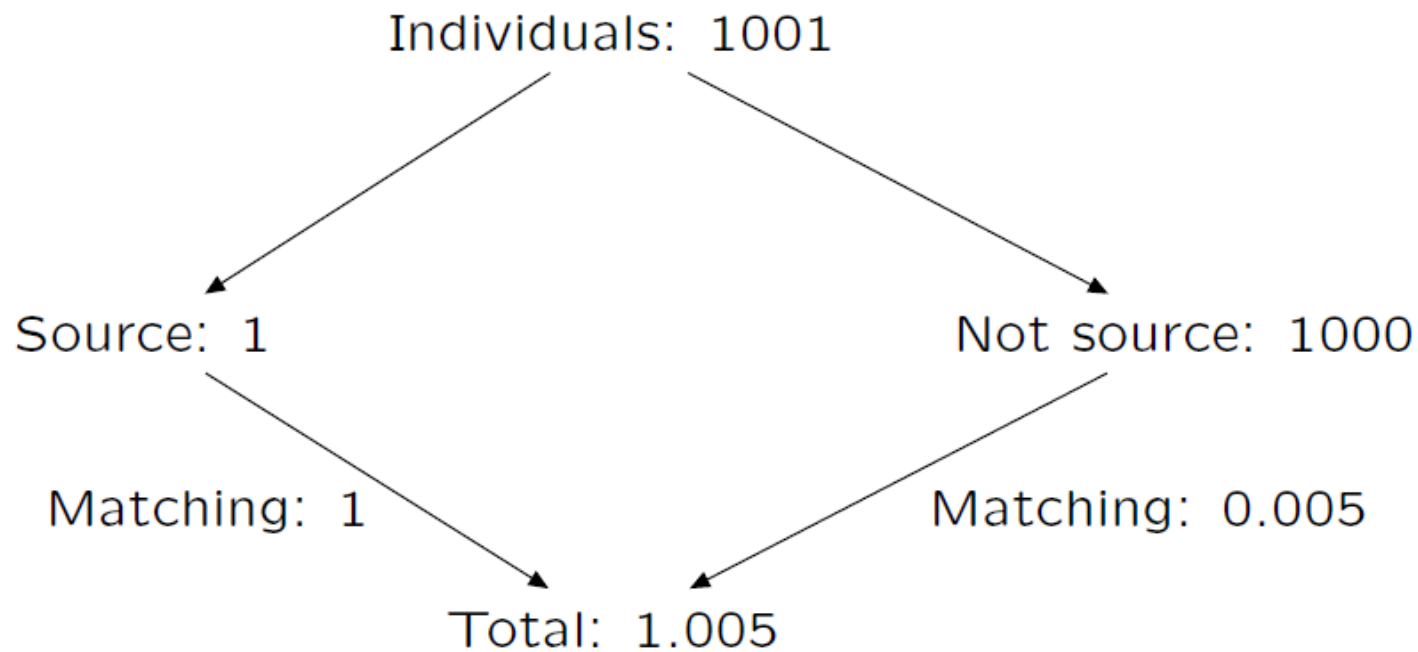
# The Island Problem

Assuming  $Q$  has no relatives on the island, there is a  chance that  $Q$  is the source.



# The Island Problem

Assuming  $Q$  has no relatives on the island, there is a  $\frac{1}{1.005} \approx 99.5\%$  chance that  $Q$  is the source.



# The Island Problem – Relatives

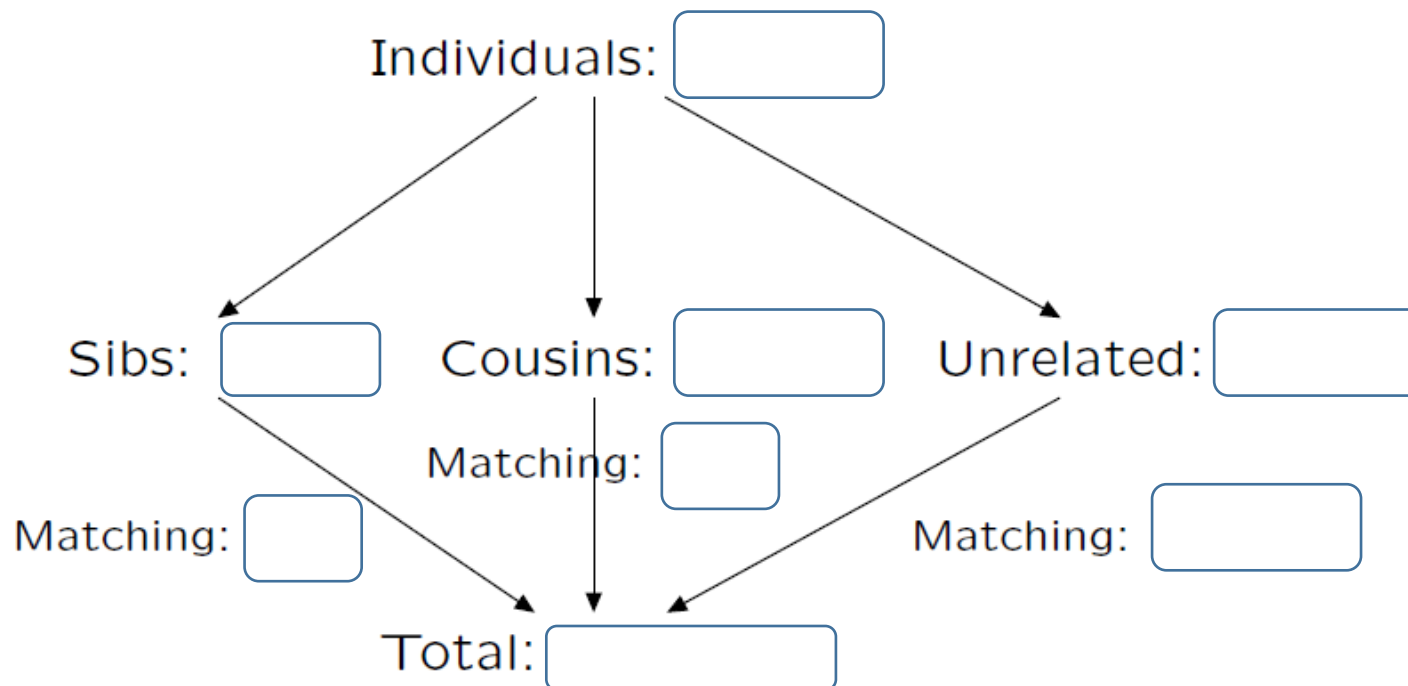
Now suppose that  $Q$  has one sibling and 20 cousins on the island, and no other relatives. What is now the probability that  $Q$  is the source, using match probabilities of:

- 1 in 1000 for a cousin;
- 1 in 100 for a sibling;
- and  $5 \times 10^{-6}$  for unrelated individuals.

Source: Weight-of-Evidence for Forensic DNA Profiles (Balding, 2015)

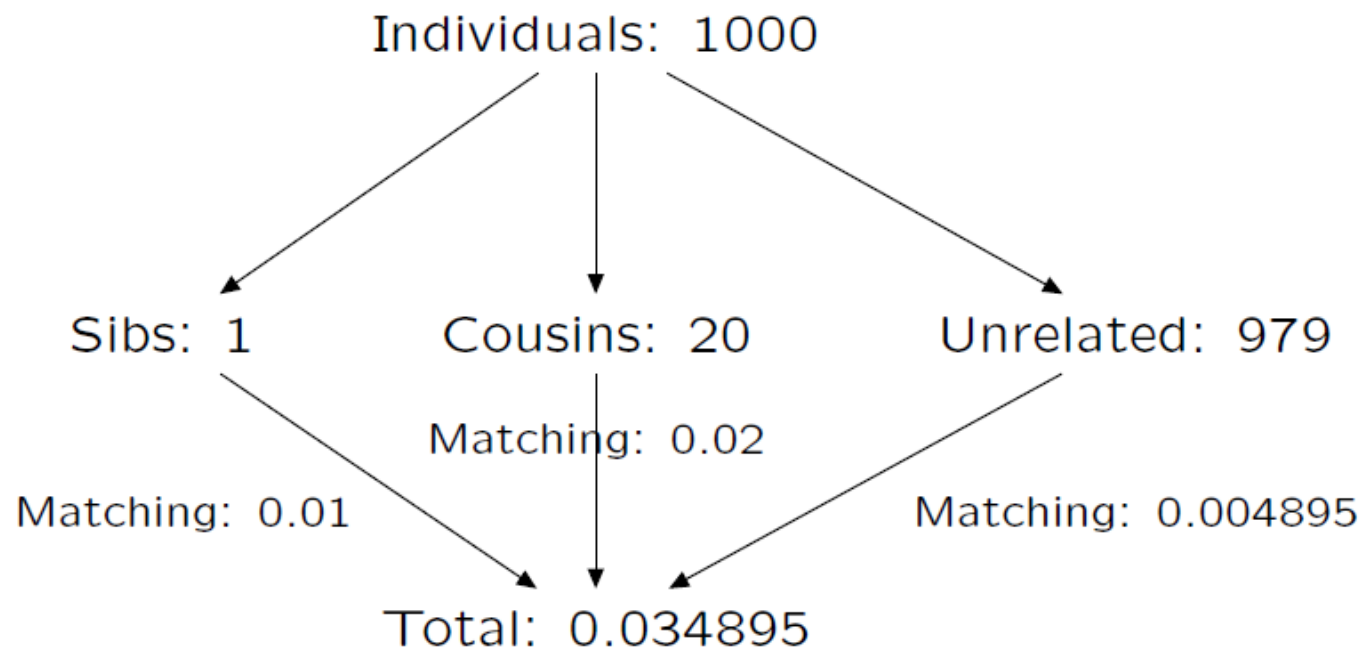
# The Island Problem – Relatives

In this case the probability that  $Q$  is the source decreases to



# The Island Problem – Relatives

In this case the probability that  $Q$  is the source decreases to  $\frac{1}{1.034895} \approx 96.6\%$ .



# The Island Problem – Relatives

Note how the LR for unrelated individuals ( $LR_U = 200\,000$ ), the LR for cousins ( $LR_C = 1\,000$ ), and the LR for siblings ( $LR_S = 100$ ), can be combined as a weighted average of the match probabilities:

$$\left( \frac{979}{1000} \times 5 \times 10^{-6} + \frac{20}{1000} \times \frac{1}{1000} + \frac{1}{1000} \times \frac{1}{100} \right)^{-1} \approx 28\,650.$$

With prior odds of  $\frac{1}{1000}$ , the probability that  $Q$  is not the source decreases from  $\frac{1}{201} \approx 0.5\%$  to  $\frac{1}{29.65} \approx 3.4\%$ .

What if we were not given any information about the relatives of  $Q$ ?



# Other Applications

The concept of relatedness is important for, and benefits, many applications:

- Paternity testing
- Missing persons
- Familial searching
- Genealogical searching

# Paternity Testing

Paternity and familial identification can provide evidence in criminal context and during civil litigation. For a paternity case, the two propositions could be:

$H_p$ : The alleged father (AF) is the true father.

$H_d$ : Some other (unrelated) man is the father.

The likelihood ratio is in this case often referred to as the paternity index (PI).

# Paternity Testing

A case can be extended to allow for more complex situations:

- Unavailability of the mother;
- Relatedness between the mother and alleged father;
- A relative of the alleged father is the true father;
- Incorporating profiles of (alleged) relatives (e.g. for half-sibs or when alleged father is unavailable);
- Multiple children;
- Incorporating mutations, substructure, silent alleles, non-autosomal DNA, etc.

# Missing Persons

The discussed methods for evidence evaluation are also applicable to other situations, such as disaster victim identification and immigration cases.

A comparison must in these cases be carried out between a profile obtained from unidentified remains, or an applicant, and a missing person's profile.

It is, however, often the case that a sample from the missing person is not available, in which case it might be possible to make use of surrogate samples (e.g. obtained through a medical institution).

Alternatively, relatives can be used for testing purposes.

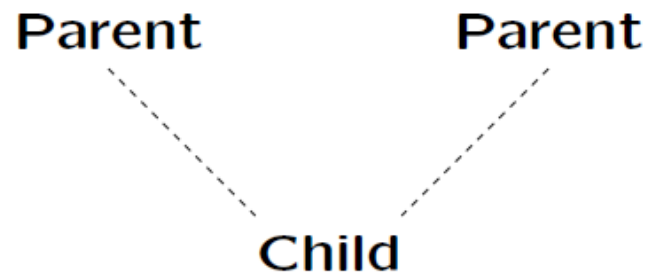
# Missing Persons

For a missing person case, the two propositions could be:

$H_p$ : The sample is from the missing person.

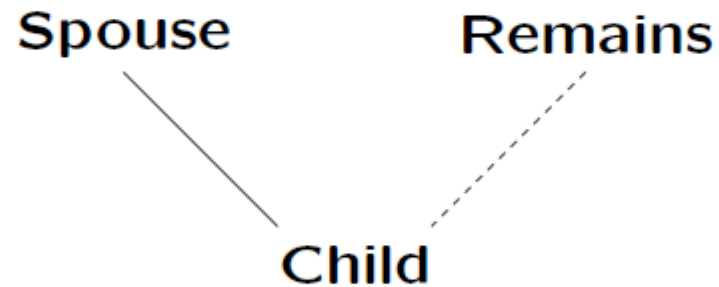
$H_d$ : The sample is from some unknown person.

The genetic evidence  $E$  may consist of the genotype from a sample that has come from some person  $X$  who may be the missing person, together with the genotypes from the parents of the missing person.



# Missing Persons

If, instead, the genotypes of the spouse  $S$  and child  $C$  of the missing person are available, the situation is similar to evidence evaluation in case of paternity testing.

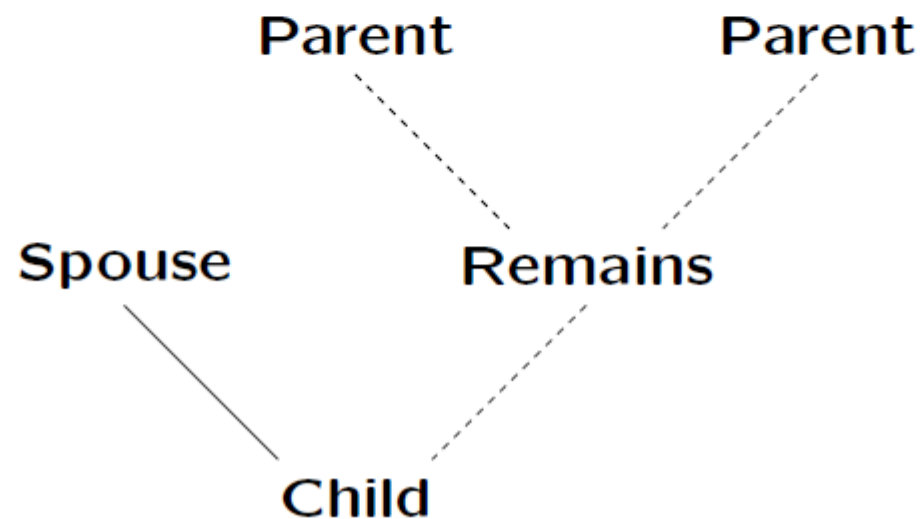


The likelihood ratios are the same as in a paternity case where  $X$  is the alleged father of child  $C$  who has mother  $S$

# Missing Persons

It may be the case that people apart from the spouse and child of the missing person are typed. The general procedure is the same: the probabilities of the set of observed genotypes under two explanations are compared.

The likelihood ratio is arranged to involve probabilities of genotypes conditional on previous generations. If both parents of an individual have been typed, there is no need to condition on the grandparents of that individual.



# Familial Searching

- A database may be used to compare crime scene profiles to known offenders when investigators lack a suspect.
- A high stringency search requires a full match of the DNA profiles, and might not always return a hit.
- Lowering the search stringency level may lead to a partial match, and has the potential to identify close relatives.
- *Familial searching* refers to the process where investigators look for close relatives in the DNA database in order to open up new investigative leads.



# Familial Searching

A certain strategy is required to select a potential relative of the unknown donor from the database. Two general methods are available, both resulting in a ranked list of candidates to investigate further:

- **IBS method:** simply counts the number of shared alleles between two DNA profiles.
- **LR method:** likelihood under two competing hypothesis (also in this context also called a kinship index (KI):

$$KI = \frac{\sum_{i=0,1,2} \Pr(G_C, G_R | IBD = i) \Pr(IBD = i | \text{relationship})}{\sum_{i=0,1,2} \Pr(G_C, G_R | IBD = i) \Pr(IBD = i | \text{unrelated})}$$

# Familial Searching

- LR methods outperform the IBS method.
- It is slightly easier to locate parent-child relationships, although siblings more often obtain a number one ranking.
- More loci improve the effectiveness of familial searching, especially in case of extra highly polymorphic loci.
- Ranked lists can be refined based on lineage markers.
- The methods can be extended by using a combination of IBS and LR, setting thresholds, or using a weighted approach.

It is important to note that the effectiveness depends on the assumption that a true close relative of the donor is actually present in the database.

# Familial Searching

A serial killer nicknamed the Grim Sleeper (due to a 14-year break) was responsible for the death of at least 10 young women in Los Angeles between 1985 and 2007.

When traditional forensic methods failed, investigators turned to novel partial-match DNA search methods authorized in 2008, eventually leading to a positive result for a recently convicted young man. Together with other evidence this led to the suspicion of the father.

The L.A. police was notified by investigators and got a DNA sample from a discarded piece of pizza. Lonnie Franklin was found to match, leading to an arrest in July 2010 and eventual conviction in May 2016.

# Familial Searching

Familial searching has proven to be a successful tool in several cases, but it also raises privacy and legal policy concerns:

- Disproportional attention to members of populations that are over-represented in the database.
- False positives may lead to the investigation of innocent people.
- Might reveal the presence of a family member in the database.
- Might reveal the presence of a previously unknown genetic link.
- Might reveal the absence of a genetic link.
- Crimes might go unreported (in case of searches against victim profiles).

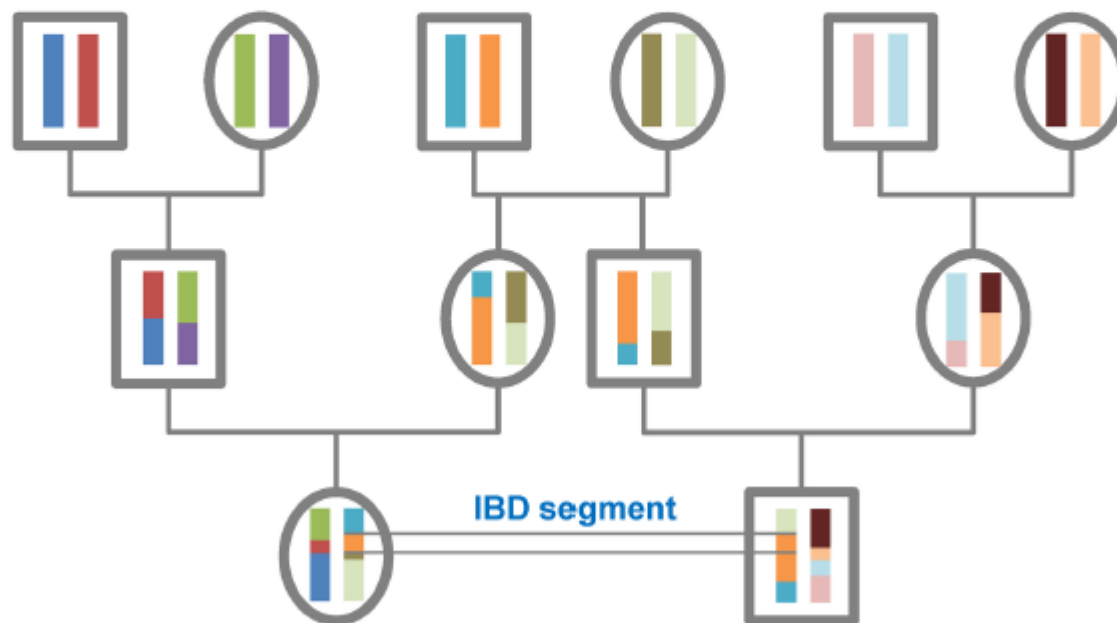
# Genealogical Searching

The limited set of STR markers does not allow for finding relatives beyond first and second degree relationships. SNP panels, with up to a million SNPs allow distinguishing even distant cousins.

A different statistical measure is used, that takes (lack of) recombination into account. Genetic linkage is measured in centimorgans (cM), which can be used to determine the level of relatedness between SNP profiles.

# Genealogical Searching

One Morgan is the length along a chromosome in which 1 recombination event is expected to occur. IBD segments occur when people share matching DNA segments that have been inherited from a common ancestor without any intervening recombination.



Source: [https://isogg.org/wiki/Identical\\_by\\_descent](https://isogg.org/wiki/Identical_by_descent)

# Genealogical Searching

Whereas policies largely restrict or even prohibit the practice of familial searching, these limitations do not explicitly extend to civilian DNA databases. With the emergence of consumer genomics tools, genealogical searching has become far more powerful.

Service	Database size	DTC provider	Relative finder	3 <sup>rd</sup> party support
23andMe	5M	•	•	
Ancestry	9M	•	•	
DNA.Land	100K		•	•
FTDNA	1M	•	•	•
GEDmatch	1M		•	•
LivingDNA	n/a	•		
MyHeritage	1.4M	•	•	•

Source: Re-identification of genomic data using long range familial searches (Erlich et al., 2018).

# Genealogical Searching – Case Example

The Golden State Killer is a serial killer, rapist, and burglar who committed at least 13 murders, more than 50 rapes, and over 100 burglaries in California from 1974 to 1986. He is believed to be responsible for three crime sprees, each of which spawned a different nickname (the Visalia Ransacker, the East Area Rapist, and the Original Night Stalker) before it became evident that they were committed by the same person.

Crime scene evidence was used to obtain a profile that mimicked the format of regular direct-to-consumer (DTC) providers in order to upload it to GEDmatch.



# Genealogical Searching – Case Example

A search identified 10 to 20 distant relatives of the perpetrator, which eventually led to the arrest of Joseph James DeAngelo. It took five genealogists four months to trace back the identity of the suspected perpetrator.

DeAngelo, a former police officer, was arrested in April 2018, after a DNA sample collected from the door handle of his car confirmed a match. In April 2019, the case prosecutors announced that they would seek the death penalty.

# Genealogical Searching - Implications

In May 2019, in response to privacy concerns, GEDmatch updated their policy. Users must now explicitly opt in to allow their profiles to be used in law enforcement investigations.

## Genealogy databases could reveal the identity of most Americans

Keeping your DNA private is getting harder

BY TINA HESMAN SAEY 4:12PM, OCTOBER 12, 2018

At the same time, the service authorized law enforcement to upload DNA data to identify a perpetrator of a violent crime against another individual. The policy defined a violent crime to include only homicide and sexual assault.