Applications of Relatedness

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.

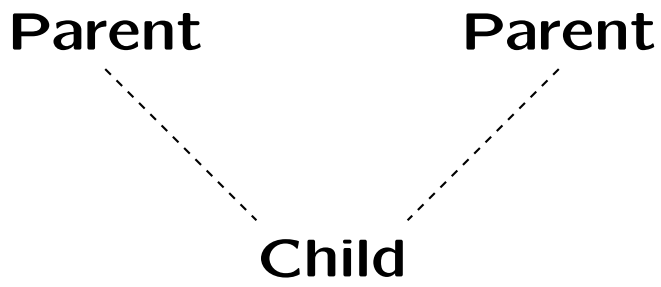
Missing Persons

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

H_p: The sample is from the missing person.

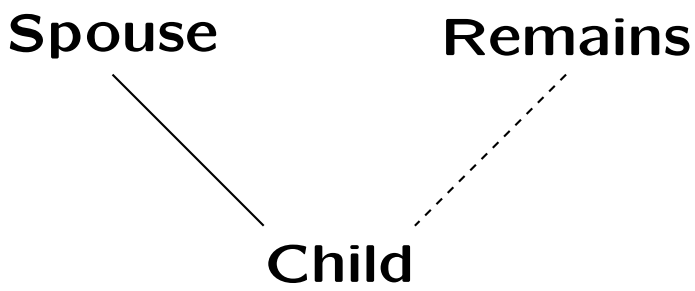
H_d: The sample is from some unknown person.

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.

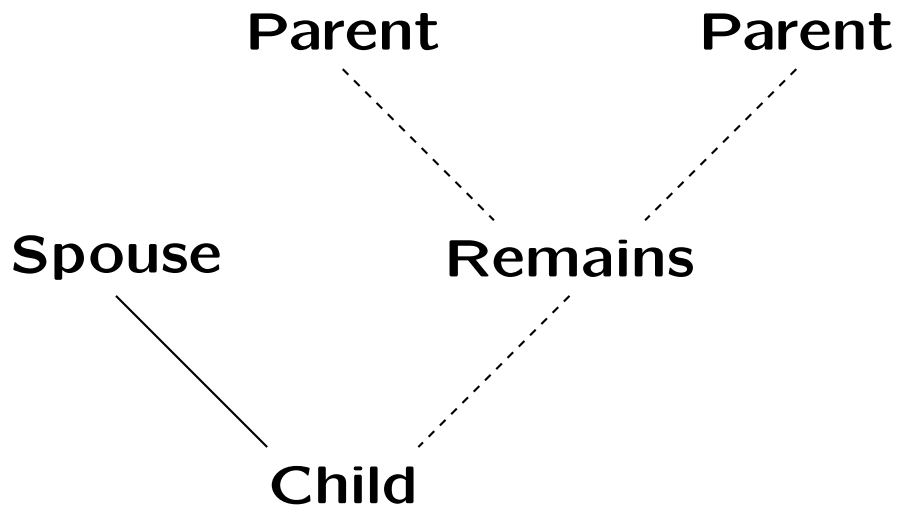


The likelihood ratios are the same as in a paternity case where X is the alleged father of child C who has mother S

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 - Strategies

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 (als in this context also called a kinship index (KI):

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

Familial Searching - Effectiveness

- 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 - Considerations

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

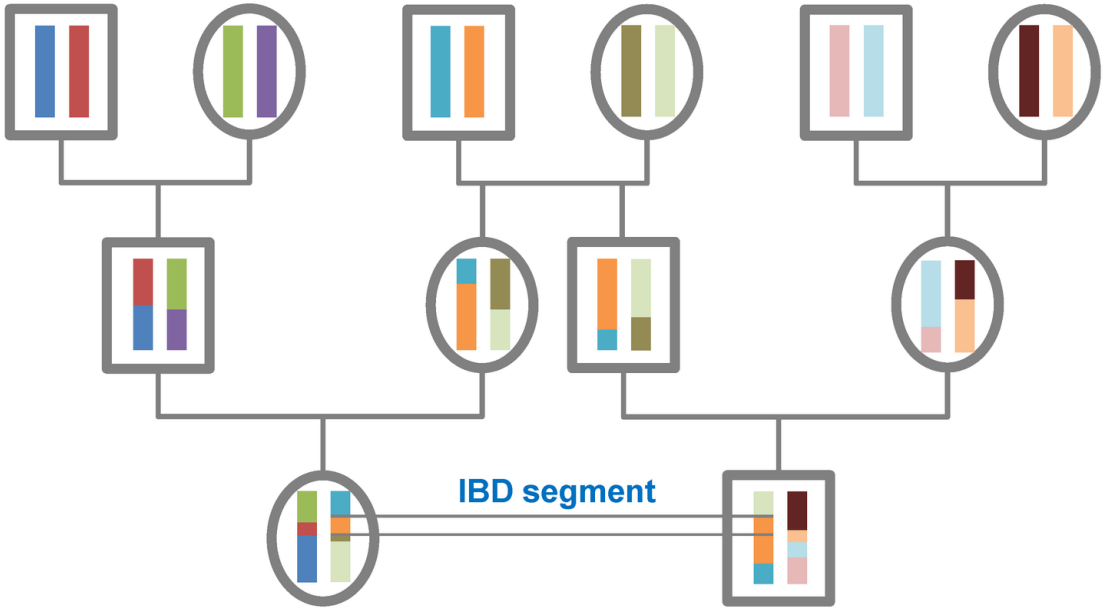
Genealogical Searching

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 - Strategy

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`

Genealogical Searching - Strategy

Whereas policies largely restrict of 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	3rd 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 - 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.