INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Vzdělávání v oblasti forenzní genetiky
reg. č. CZ.1.07/2.3.00/09.0080

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Familias: A recently upgraded program for determining familial relationships

Petter Mostad
Chalmers University, Gothenburg, Sweden
mostad@chalmers.se

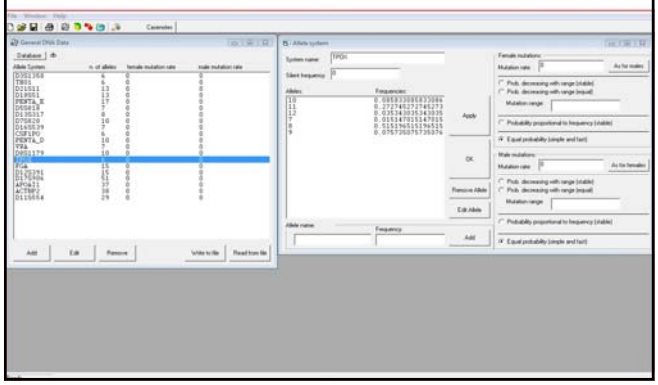
Overview

- Inference of familial relationships with the Familias program
- Newly added features (new homepage etc)
- Using Familias with multiple hypotheses
- Some lesser known features:
  - Generation of pedigrees
  - Priors
  - Etc.
- Future developments: R-package etc

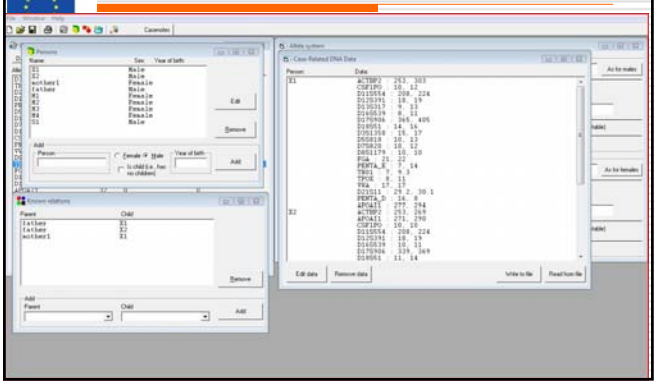
The familias program

- Project initiated around 1992: computation of probabilities relating to inferring family relations
  - RMI – Forensic medicine in Oslo, Björnar Olaisen, Margurethe Stenersen, Berit Mevåg.
  - Thore Egeland & myself at the Norwegian Computing Center
- Released free on the internet 1995
- Validated by Jiří Drábek (2008)

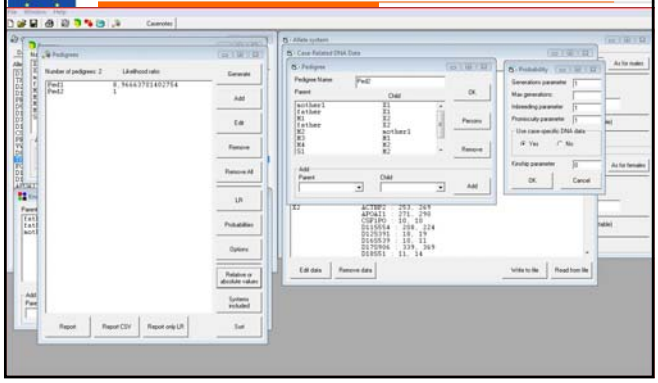
Familias user interface



Familias user interface, cont.

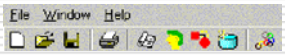


Familias user interface, cont.



Some Familias features

- General pedigrees
- Computations of likelihoods, LR, posterior probabilities.
- Multiple hypotheses
- Different mutation models
- Kinship
- Silent alleles
- Automatic pedigree generation
- Simple Windows user interface

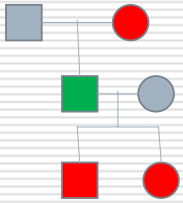


Examples of uses

- ”Deficient” paternity cases, where data is available only from more distant relatives
- Disaster identification (Olaisen 1996)
- Immigration cases (Tillmar 2007)
- Animal breeding
- …and probably many uses we don’t know about

Largest user problem: Defining pedigrees correctly

- Example: Identifying remains of man based on tests of his mother and his children
- In alternative pedigree, an alternative missing man must take his place



Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

The recent familias upgrade

- Improved input and output of data formats (Excel-type)
- Improved presentation of results, and improved design of some forms
- Ongoing improvements in new homepage www.familias.name
- Improved stability
  - Two bugs in user interface fixed
- Larger extensions has been moved to a coming R implementation

Two or more hypotheses?

- Bayesian philosopical view: Separation between DNA-based evidence and non-DNA evidence.
- There may be *several* different scenarios, each with its *separate* probability for observing the given data.
- Each should be investigated and reported.

Example: The brother of the alledged father is the father

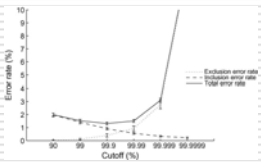
- The LR between a man and a ”randomly selected man” in a certain population may be quite high.
- A close relative may be expected to fit almost equally well
- Case-dependent non-DNA evidence indicate whether close relatives should be considered

Example: Immigration cases

- Immigration rules (in Sweden) specify exactly which relationships a residence permit may be based on
- For some immigrants, biological parenting may be less important than social parenting
- For an immigrant family, a number of possible pedigrees may have to be investigated

Immigration cases

Tillmar et al (2007) investigated the risk of erroneous conclusions in immigration cases, using simulation (and Familias)



- Using mutations in computations lower error rates somewhat
- Improper databases increases rates slightly (but case LR values may be quite influenced)
- Including an investigation of a standard set of 5 hypotheses decreases rates slightly

Example: Large number of possible pedigrees

- ”Given DNA-data on a set of persons, what is their relationship?”
- A set of *possible* pedigrees must be specified (and given prior probabilities)
- Familias can generate the set of possible pedigrees, given some constraints.
- A crude type of prior can also be computed.

Example: The Romanov family

- The Romanov family (parents, five children, and their doctor) disappeared in connection with the Russian revolution.
- The remains in a grave was later shown to contain the family, by using mtDNA and STR markers (Gill et al, 1994)
- In Egeland et al (2000) it was shown how to use Familias to combinatorially find the right pedigree.

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

An example extending Familias

- When using non-zero mutation rates, computations with Familias may become too slow on large pedigrees with ”loops”.
- If mutations are disregarded, some loops may be replaced by simple conditional probabilities of whether or not alleles are IBD
- Computations may be implemented by combining Familias with some external calculations, for example in Excel.

Example, continued.

- Case: Do two men share the same father or not? Their mothers are cousins once removed.
- Solution: condition on whether paternal and/or maternal alleles are IBD. Sum and multiply over all autosomal markers
- Can be extended to similar situations
- BUT: This also illustrates the limiations of Familias and its user interface

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Future extensions of Familias planned in new framework

- We are working on an implementation of the core functions of Familias in an R package: OpenFamilias
- This will make future extensions much faster and simpler to do
- More flexible for advanced users!
- A bit more difficult for novices

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Some current work which may be implemented this way

- Computational methods for X chromosome markers in LD (Tillmar, submitted)
- Computational models for SNP data
- Evaluation of STR mixture data

Future work:

- Support for multiple databases
- Inclusion of Y data
- Inclusion of mtDNA data
- Simulations for power computations
- Missing persons evaluation in catastrope scenarios

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Acknowledgements

- Oslo, Norway: Thore Egeland, Öyvind Skare, Margurethe Stenersen, Bente Mevåg, Björnar Olaisen, …
- Linköping, Sweden: Gunilla Holmlund, Andreas Tillmar, Bertil Lindblom, Jenny Velander, Daniel Kling,…
- Göteborg, Sweden (where I am): Iftikhar Ahmad, Qi Cao, …

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

References

- Gill et al (1994): Identification of the remains of the Romanov family by DNA analysis (Nat.Genet.)
- Olaisen (1996): Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster (Nature Genetics)
- Egeland et al (2000): Beyond traditional paternity and identification cases: Selecting the most probable pedigree (FSI)
- Tillmar (2007): DNA testing for immigration cases: The risk of erroneous conclusions (FSI)
- Jiří Drábek (2008): Validation of software for calculating the likelihood ratio for parentage and kinship (FSI)

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.