PRELIMINARY VERSION, THE PRESENTATION WILL BE SUBJECT TO CHANGES

Familias 3.0 – Exploring new features

Daniel Kling
Norwegian Institute of Public Health *and* Norwegian University of Life Sciences
daniel.kling@fhi.no

Copenhagen, Oct 8-9, 2013

Familias 3.0 - Outline

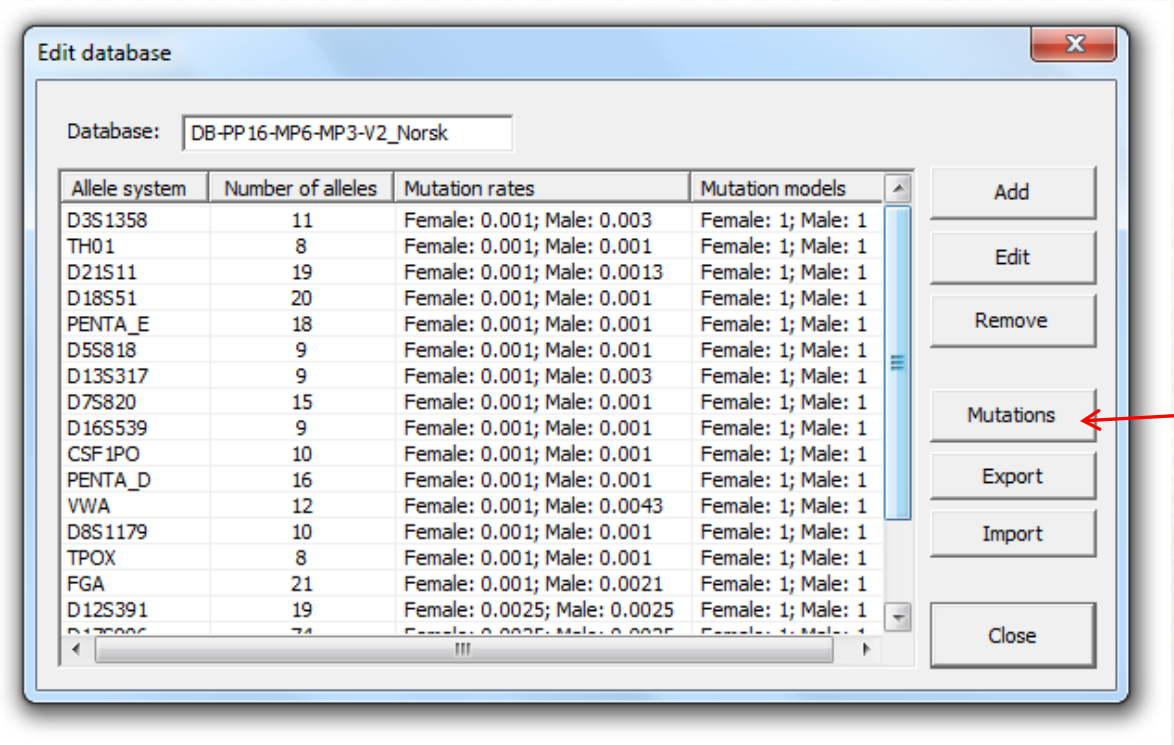
- Brief overview
- New mutation model
- Simulation interface
- DVI module
- Blind search module

Familias 3.0 – Introduction

- Developed from the old core
- Based on user input
- Validation article (Drabek)
- Subject to further improvements based on user input

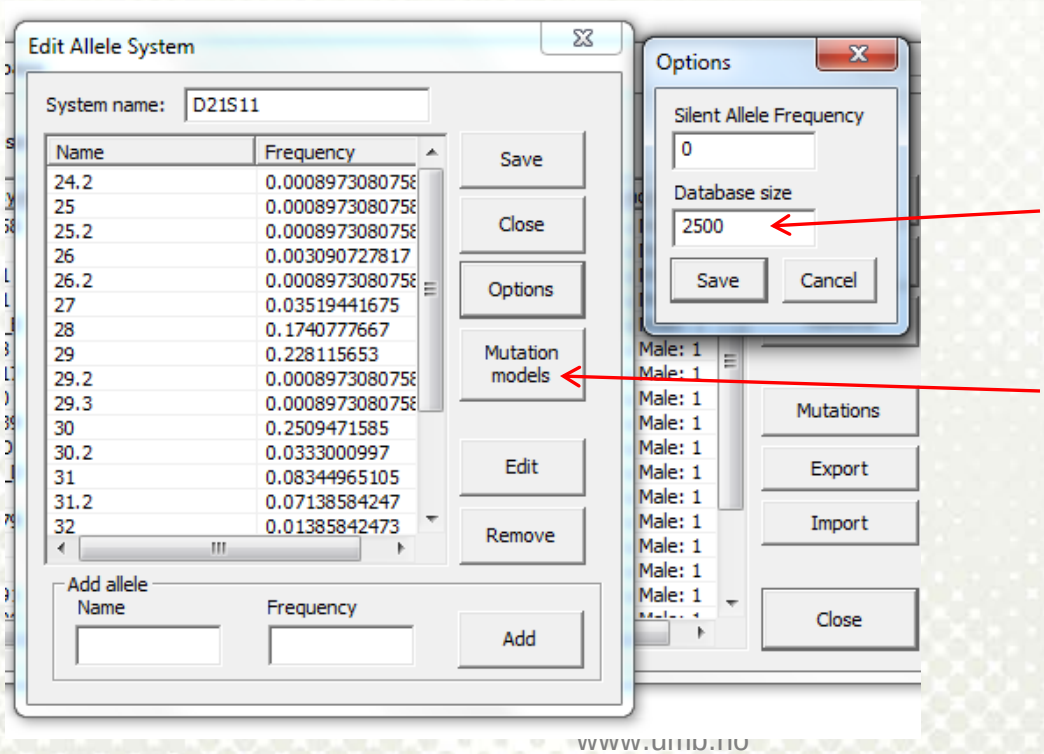
Familias 3.0 – At a glance

- Define general database
- Change mutations



Familias 3.0 – At a glance

- Define marker specific database sizes
- New allele dialog
- New mutation dialog

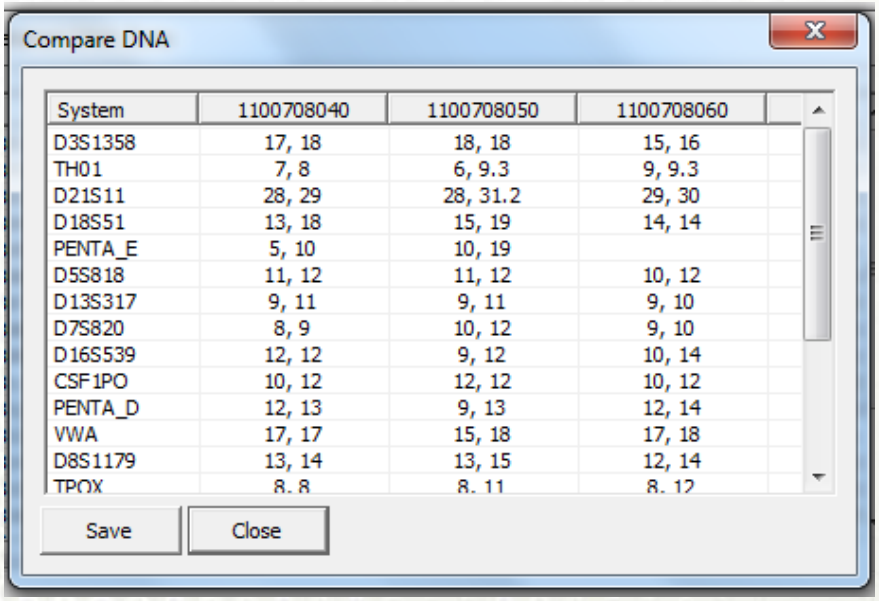


Familias 3.0 – At a glance

- New import options
- Case data dialog
- Compare profiles

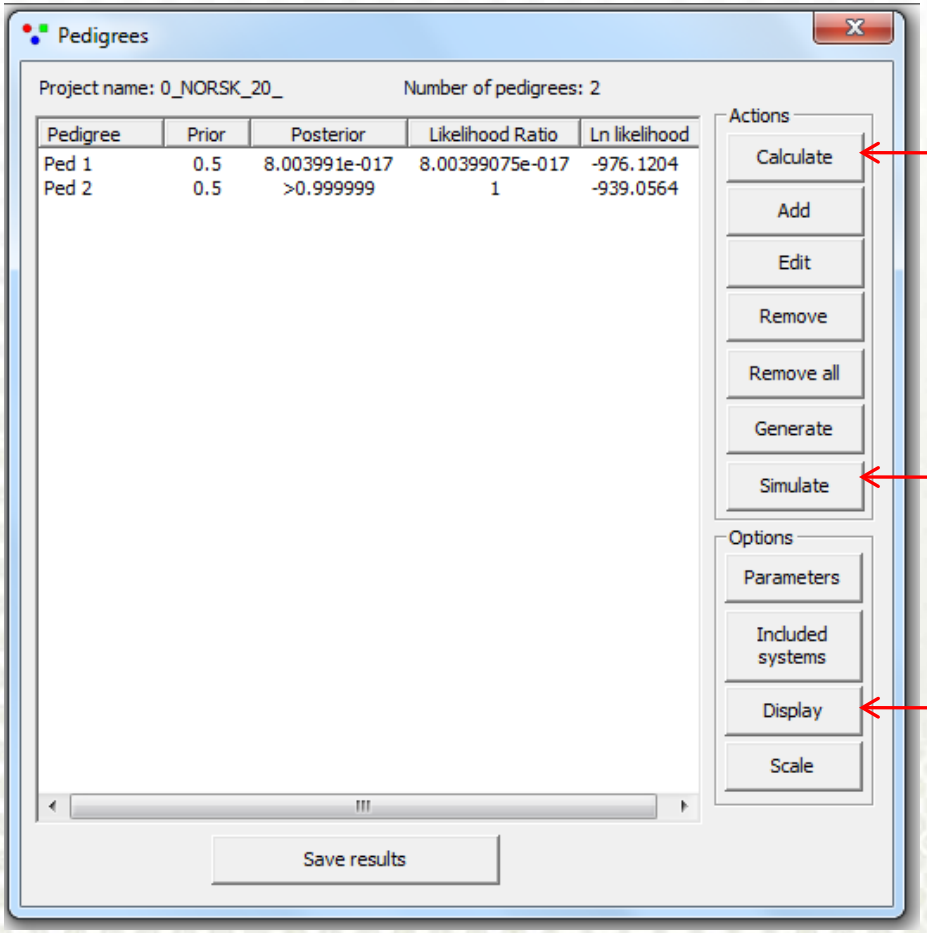
Familias 3.0 – At a glance

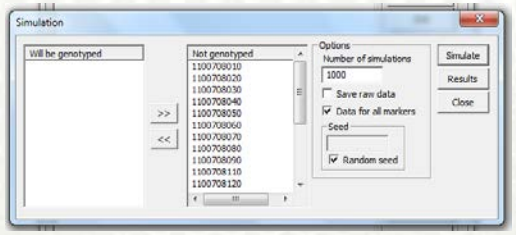
- Compare data

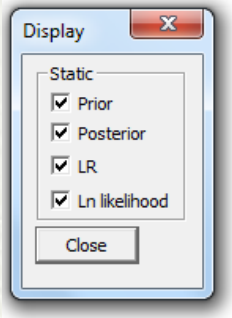


Familias 3.0 – At a glance

- Definition of pedigrees

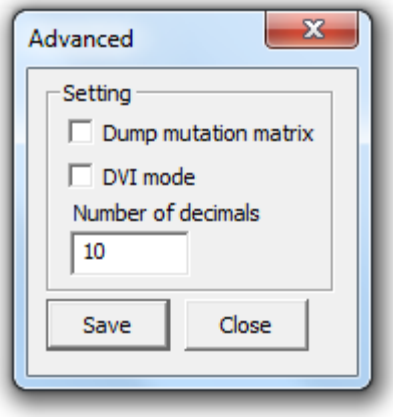






Familias 3.0 – At a glance

- Advanced settings



Familias 3.0 – New mutation model

- Theory
  - Steps
  - Microvariants
- Parameters
  - Mutation rate (μ)
  - Range (r)
  - Secondary mutation rate (α)

Familias 3.0 – New mutation model

- Mutation matrix

$$M = \begin{array}{c|cc} & 12 & 13 \\ 12 & 1-\mu & \mu \\ 13 & \mu & 1-\mu \end{array}$$

- Example

List of alleles
12.1
13
13.1
14
17

$$M = \begin{bmatrix} (1-\mu)(1-\alpha) & \alpha/3 & (1-\alpha)k_1 r^1 & \alpha/3 & \alpha/3 \\ \alpha/2 & (1-\mu)(1-\alpha) & \alpha/2 & (1-\alpha)k_2 r^1 & (1-\alpha)k_2 r^4 \\ (1-\alpha)k_3 r^1 & \alpha/3 & (1-\mu)(1-\alpha) & \alpha/3 & \alpha/3 \\ \alpha/2 & (1-\alpha)k_4 r^1 & \alpha/2 & (1-\mu)(1-\alpha) & (1-\alpha)k_1 r^3 \\ \alpha/2 & (1-\alpha)k_5 r^4 & \alpha/2 & (1-\alpha)k_5 r^3 & (1-\mu)(1-\alpha) \end{bmatrix}$$

For first row: $1 = (1-\mu)(1-\alpha) + \alpha/3 * 3 + (1-\alpha)k_1 r \Leftrightarrow k_1 = \frac{1-(1-\mu)(1-\alpha)-\alpha}{(1-\alpha)r} = \frac{\mu}{r}$

Familias 3.0 – Simulations

- What is a simulation?
  - Based on a model
- Why do we want to simulate?
  - Find distributions of LRs
  - What can we expect
- The utility of simulations
  - Assisting tool in decision prior to obtaining a case
  - Assisting tool after a case is complete
- Further uses
  - Use raw data to perform statistical calculations
  - Use simulated data in other settings

Familias 3.0 – Simulated example

- Steps in a simulation

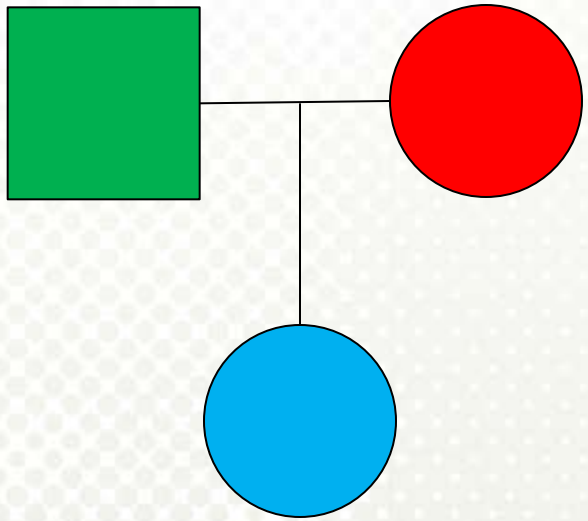
1. Define allele frequencies and mutation models
2. Define all persons we need
3. Define hypotheses (i.e. pedigrees)
4. Start simulations

- Algorithm

1. Define founders and non-founders
2. Sample founder alleles from allele frequencies (Account for kinship)
3. Randomly sample the non-founder alleles based on the parents alleles and the mutation models (i.e. allele may change/mutate given a mutation rate >0)
4. Repeat 2. such that all non-founders are sampled
5. Calculate likelihoods for the simulated data given all hypotheses
6. Repeat 1-4 for all hypotheses, i.e. we must simulate all pedigrees as true

Familias 3.0 – Simulated example

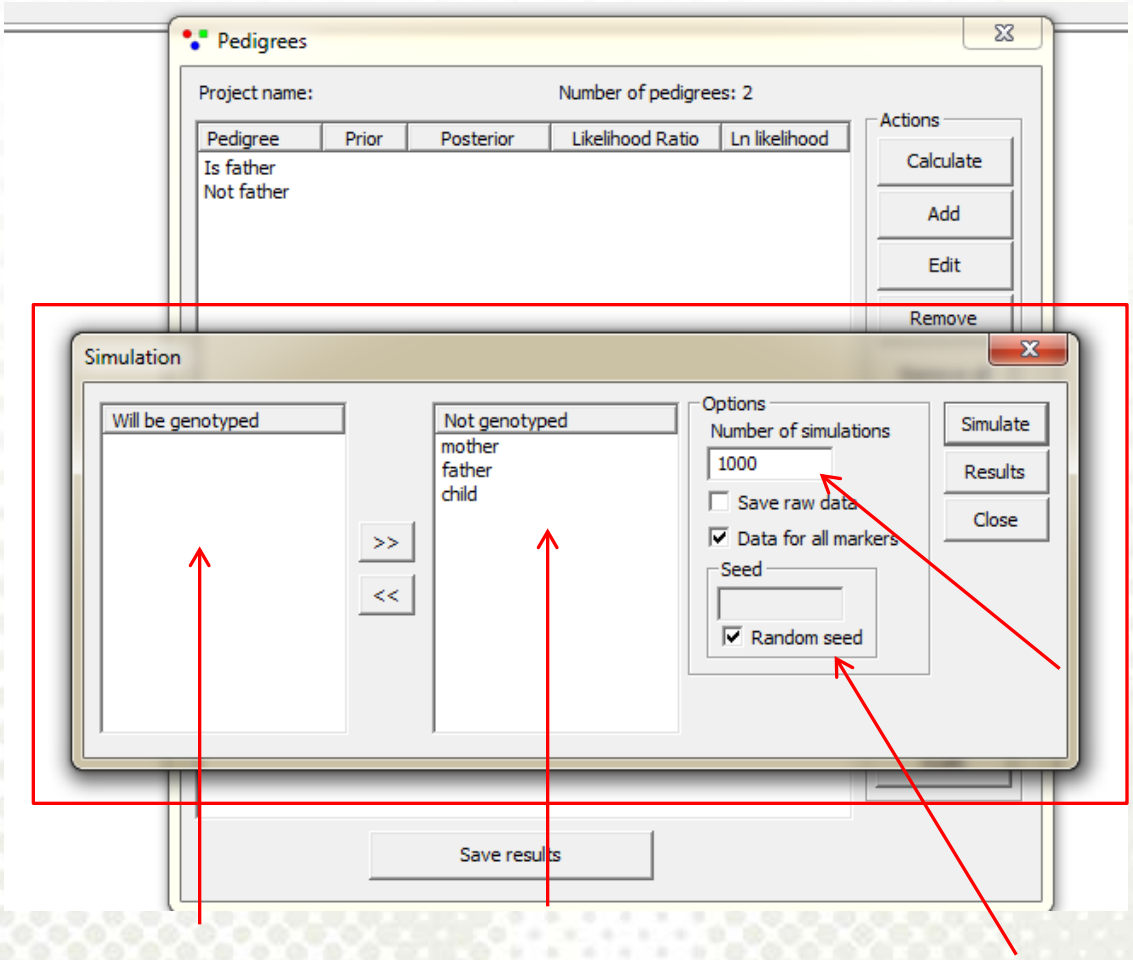
- Consider a paternity case



Familias 3.0 – Simulated example

- We define a marker S1
  - Alleles 12, 13, 14 and 15. (Uniform allele frequencies 0.25)
  - Simple mutation model with mutation rate 0.0
- We define the necessary persons
  - Mother, child and the father
  - We do NOT define any DNA data for the persons
- We define the pedigrees
  - Two different hypotheses

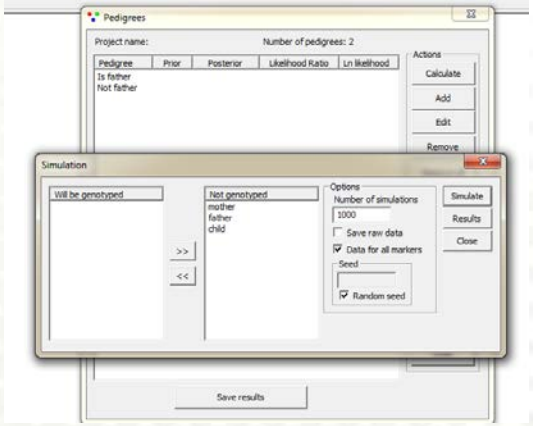
Familias 3.0 – Simulated example



Familias 3.0 – Simulated example

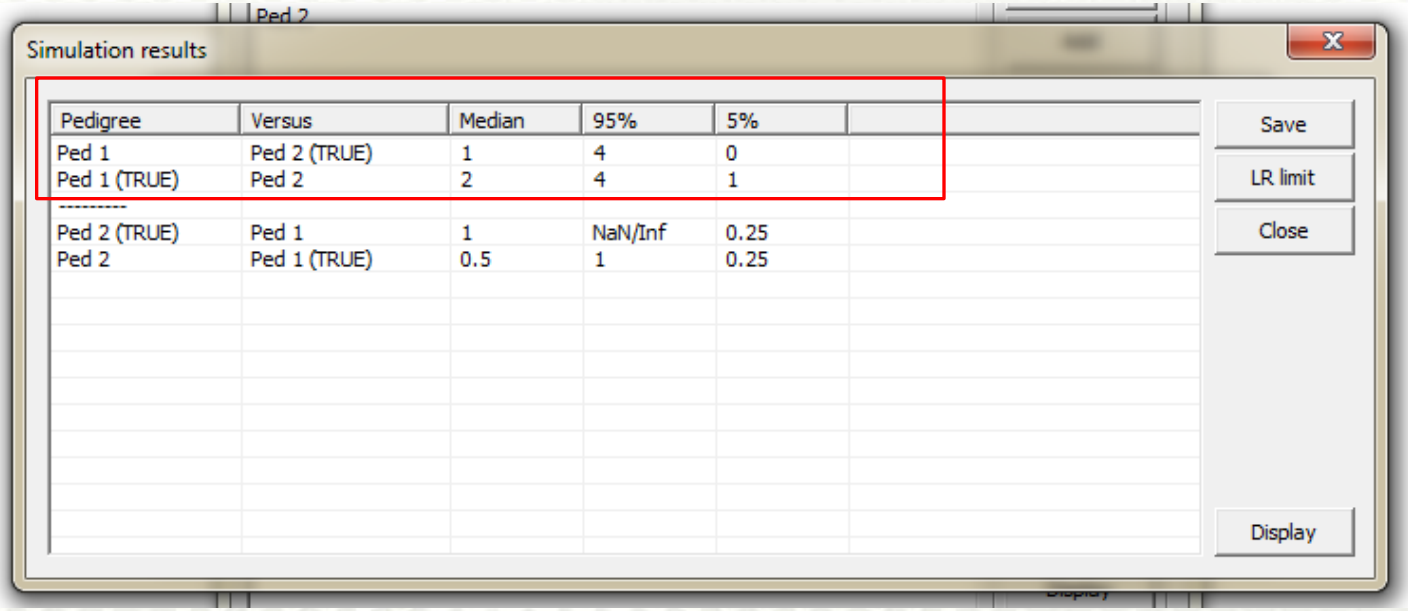
Options

- Number of simulations
  - Greater is often better
- Save the raw data
  - Use in other studies
  - Calculate your own statistics
- Random seed or non-random seed
- Select genotyped persons



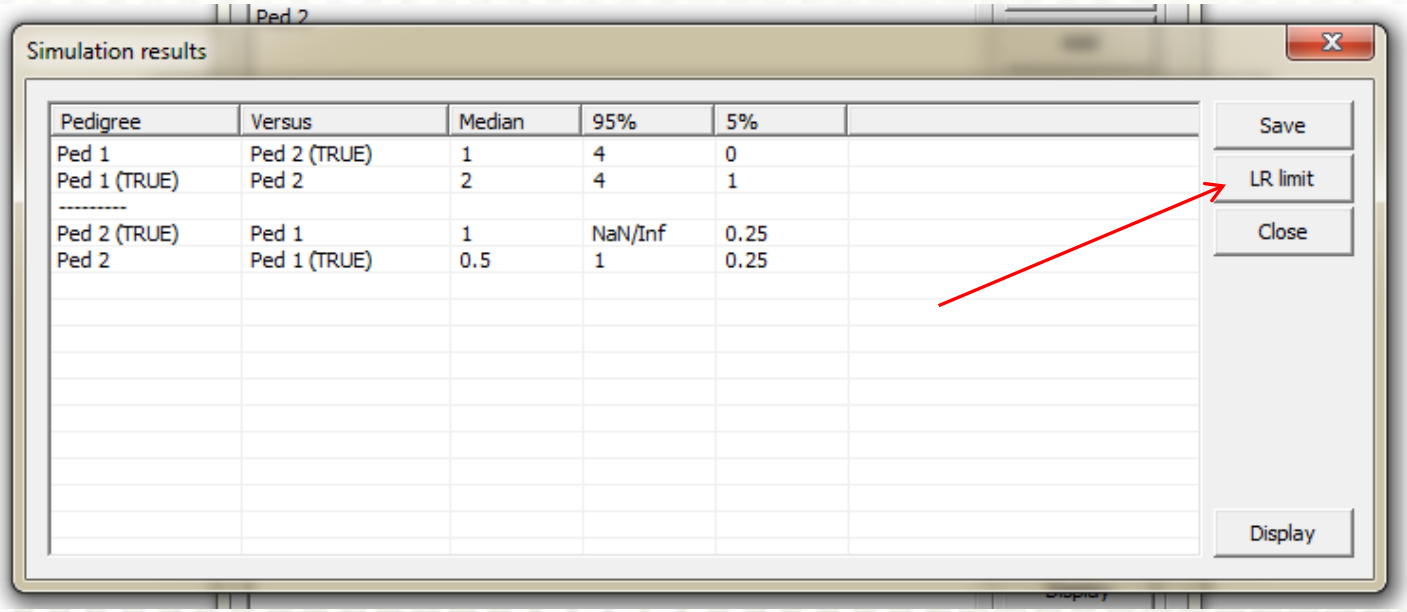
Familias 3.0 – Simulated example

- Simulate data
  - Simulate data given both hypotheses
  - In reality 1000*2 simulations are performed
  - Calculate likelihood for both hypotheses



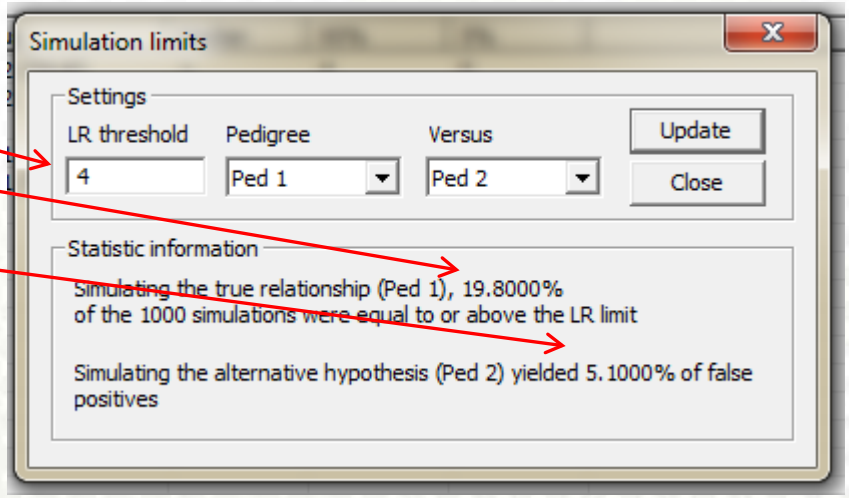
Familias 3.0 – Simulated example

- Save results as a report
- Find threshold



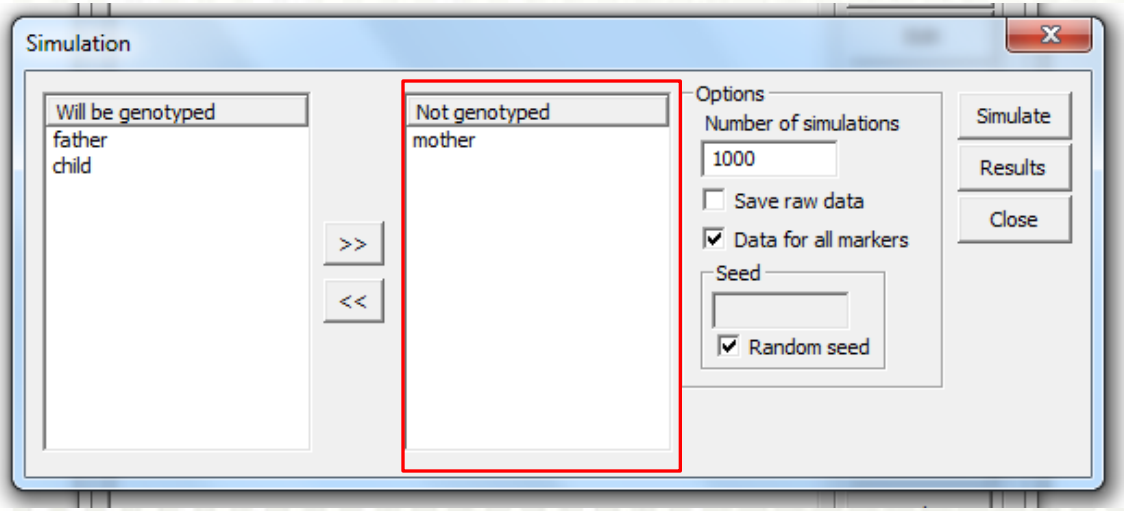
Familias 3.0 – Simulated example

- Threshold
- False negative rate
- False positive rate



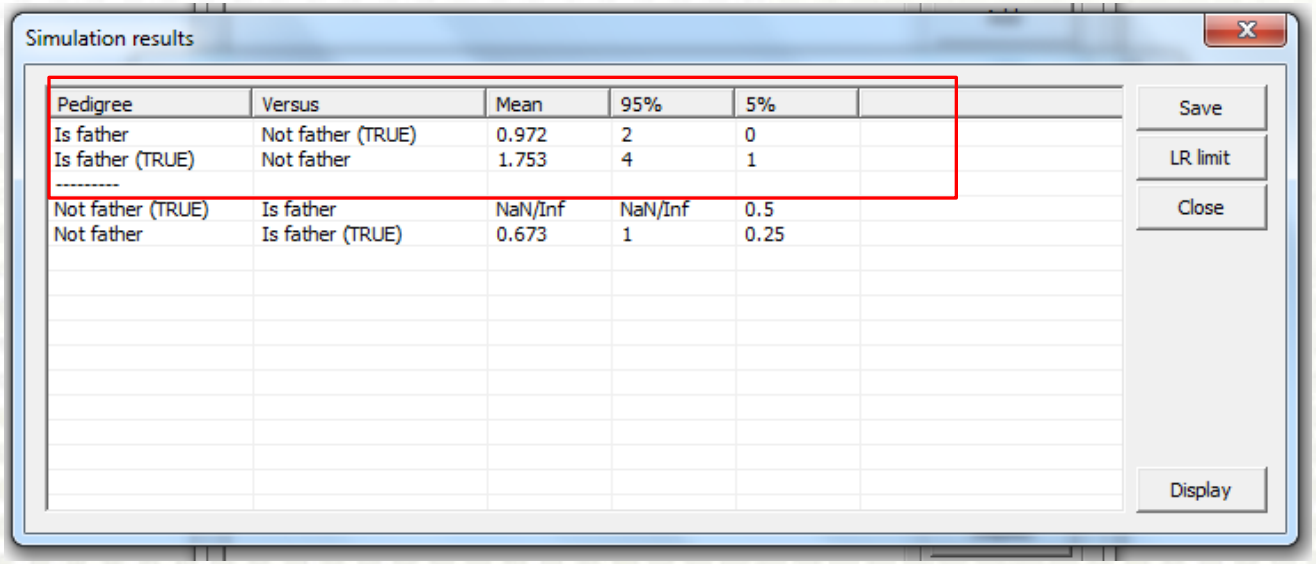
Familias 3.0 – Simulated example

- Return, the mother is not genotyped!



Familias 3.0 – Simulated example

- Simulate data



Familias 3.0 – Simulation summary

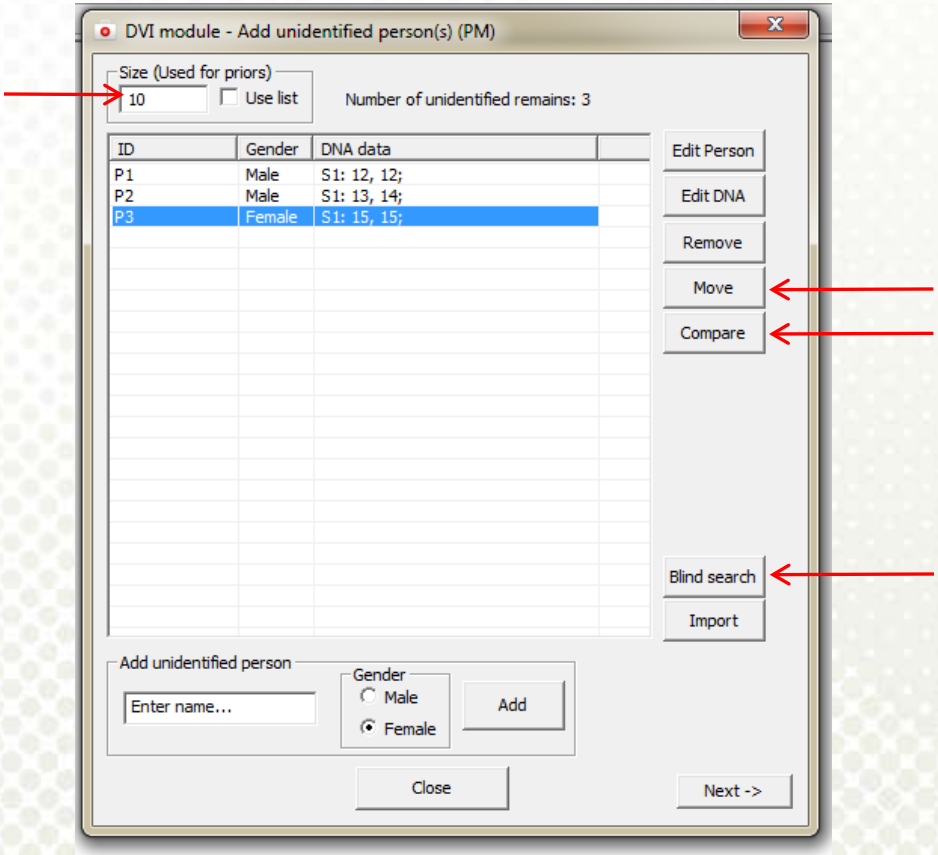
- Find good thresholds
  - False positive/negative rate
- Investigate what we can expect
- Include
  - Kinship
  - Mutations
  - Silent alleles
  - Multiple pedigrees
- Investigate the number of persons we must genotype
- Investigate the number of markers we must include

Familias 3.0 – Disaster Victim Identification (DVI)

- What is a DVI operation?
- What is the purpose of developing a DVI module?
- “Alternative” softwares
- Brief overview
- Blind search module

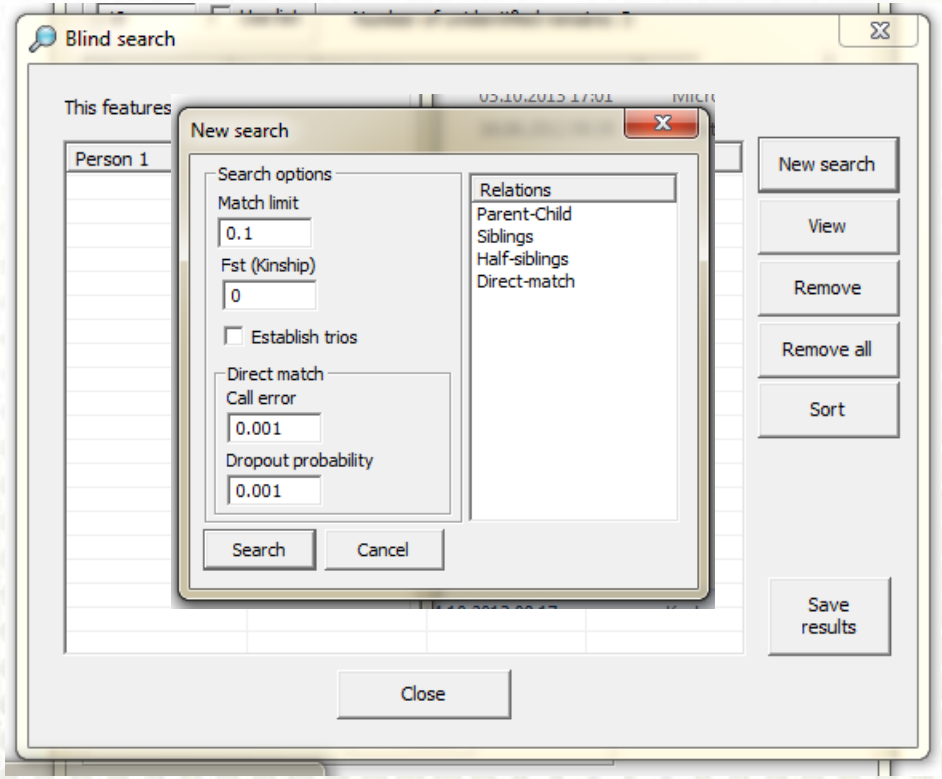
Familias 3.0 – DVI example

- Add Unidentified persons



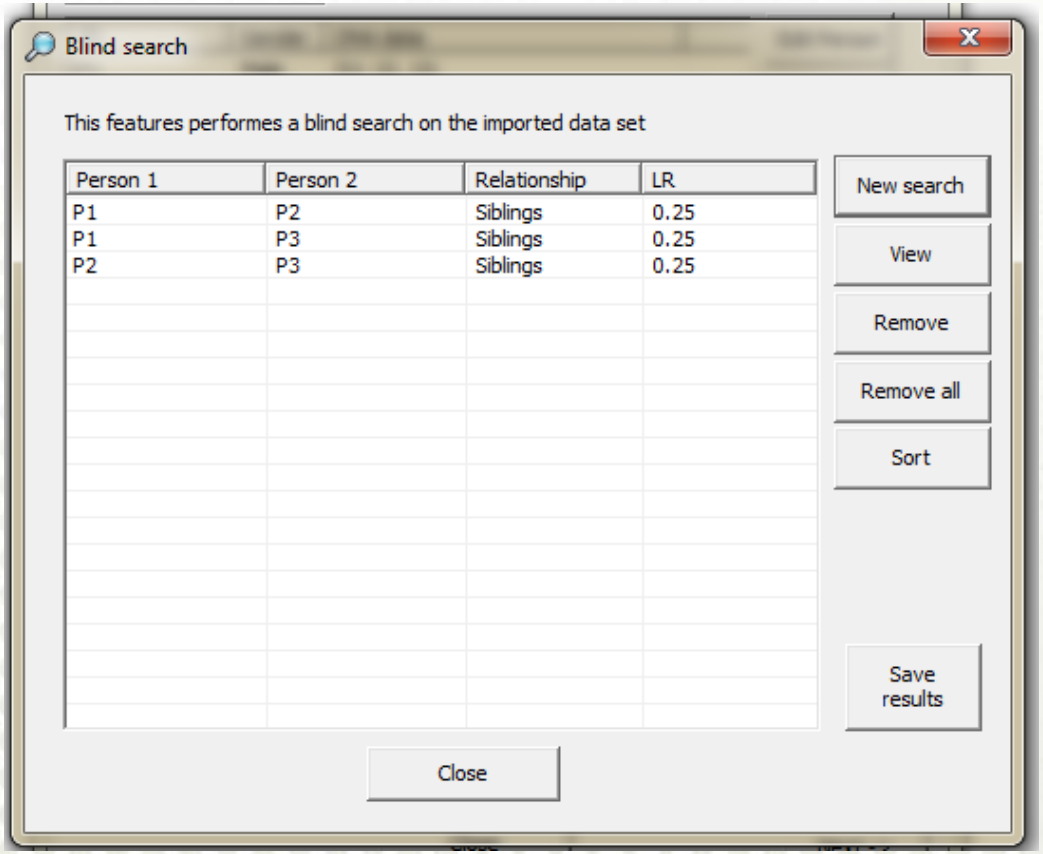
Familias 3.0 – DVI example

- Blind search



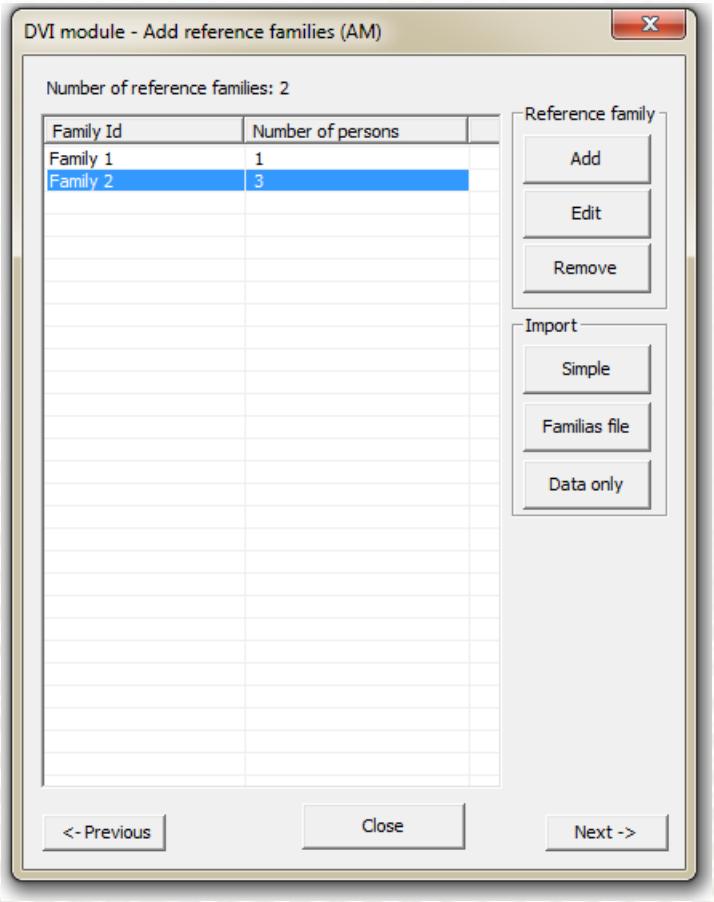
Familias 3.0 – DVI example

- Blind search



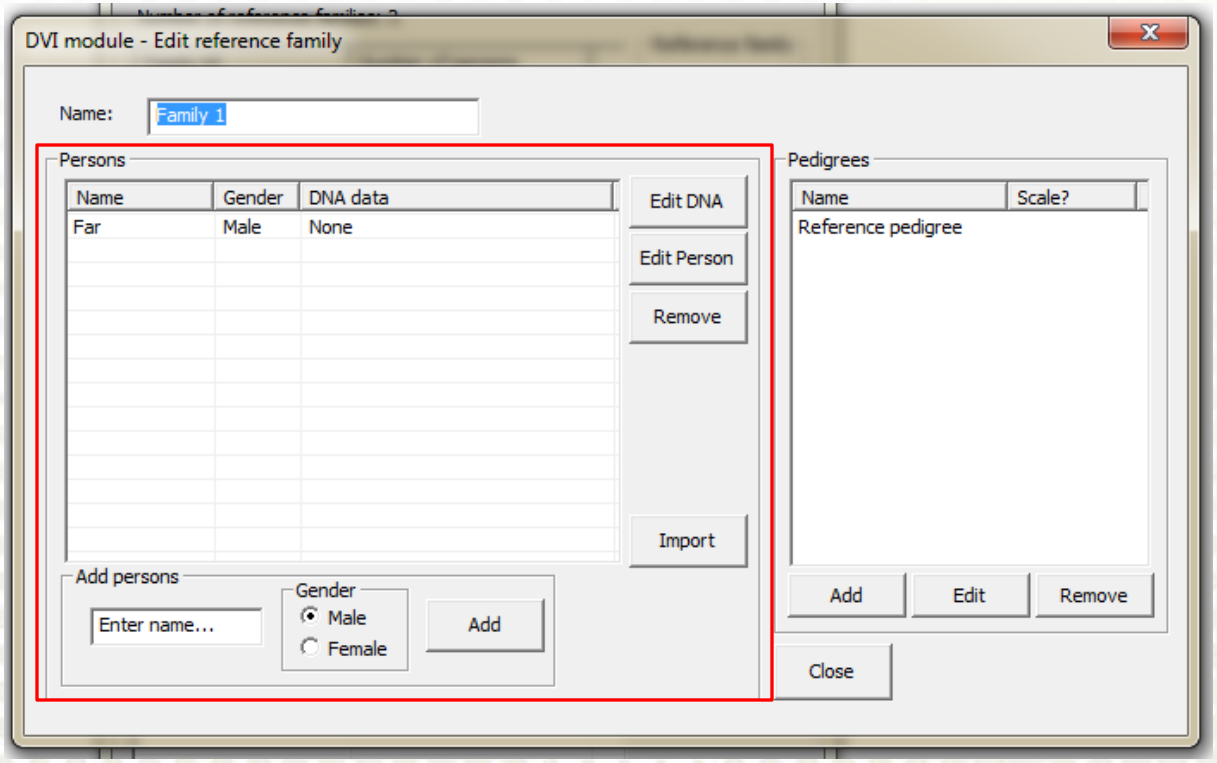
Familias 3.0 – DVI example

- Add Reference families



Familias 3.0 – DVI example

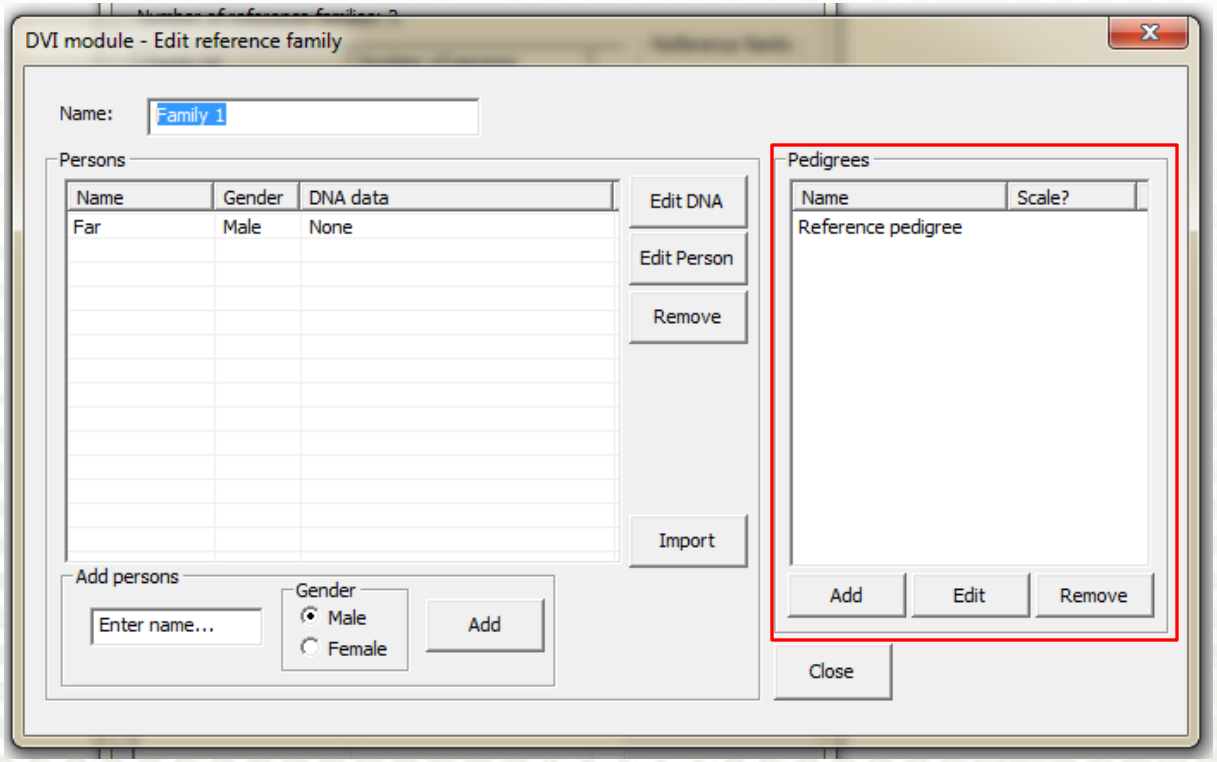
- Edit reference family



- Add persons

Familias 3.0 – DVI example

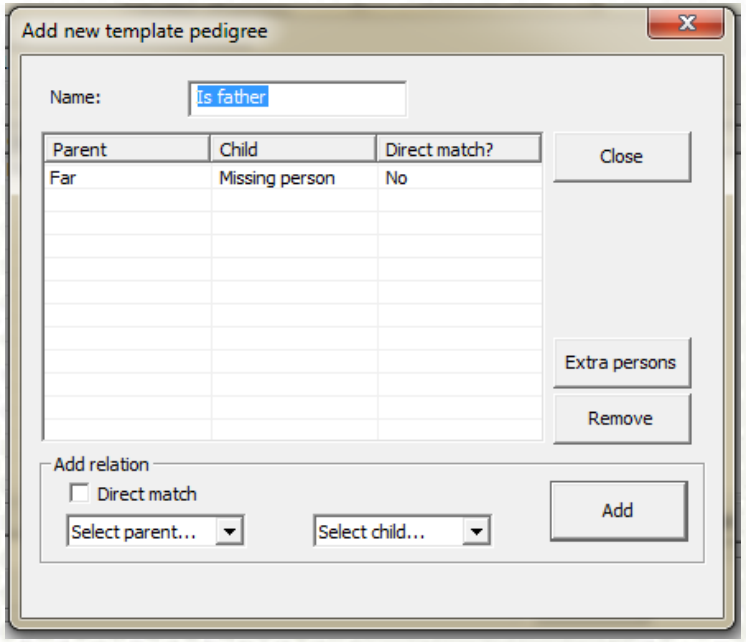
- Edit reference family



- Add pedigrees

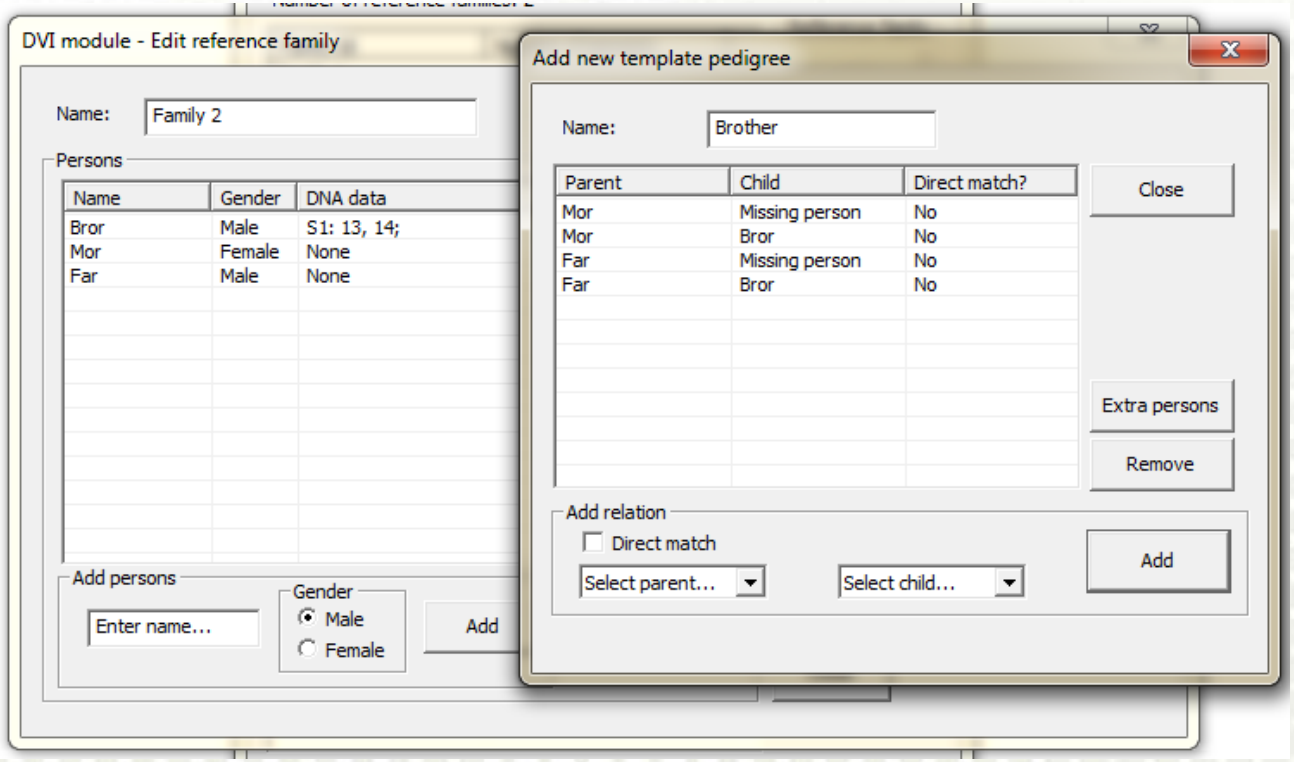
Familias 3.0 – DVI example

- Add pedigree



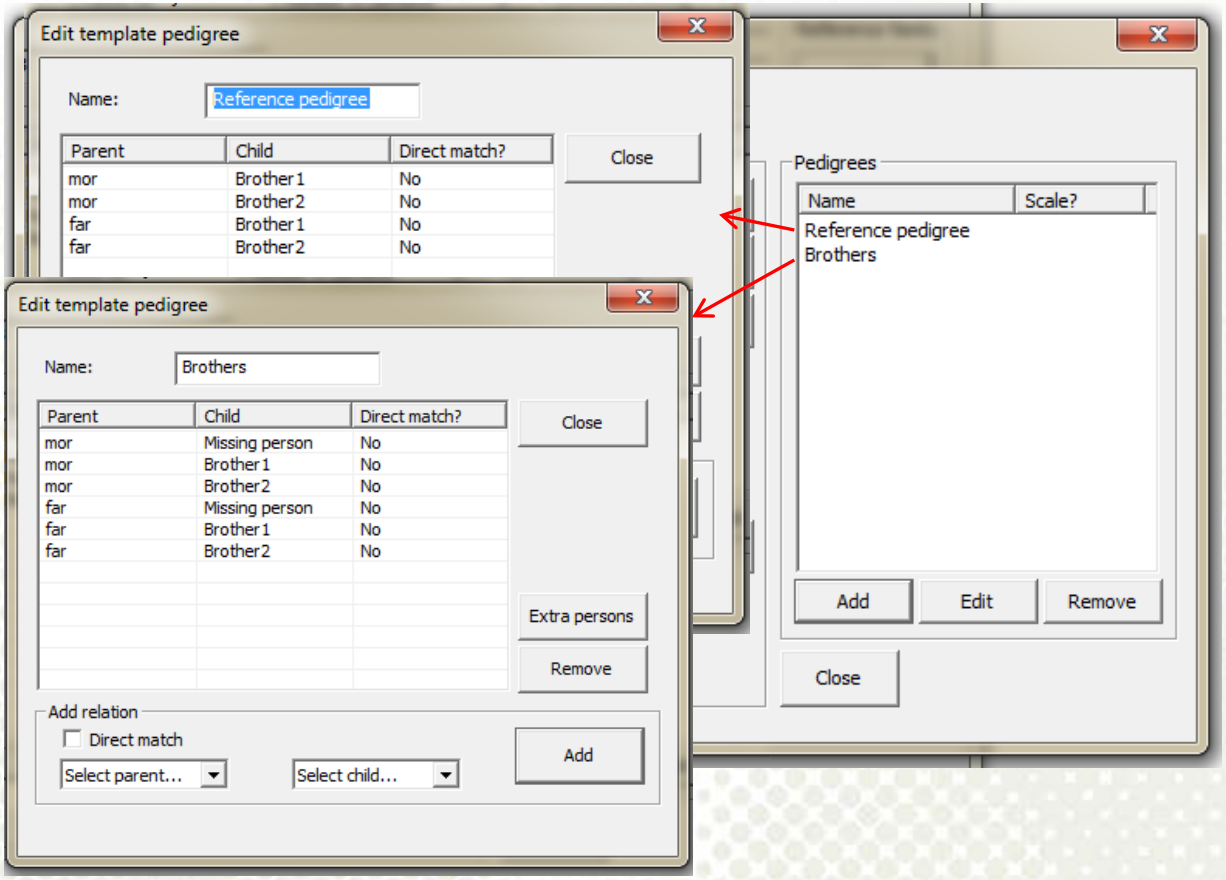
Familias 3.0 – DVI example

- What if we have more complicated relationships?
- Brother



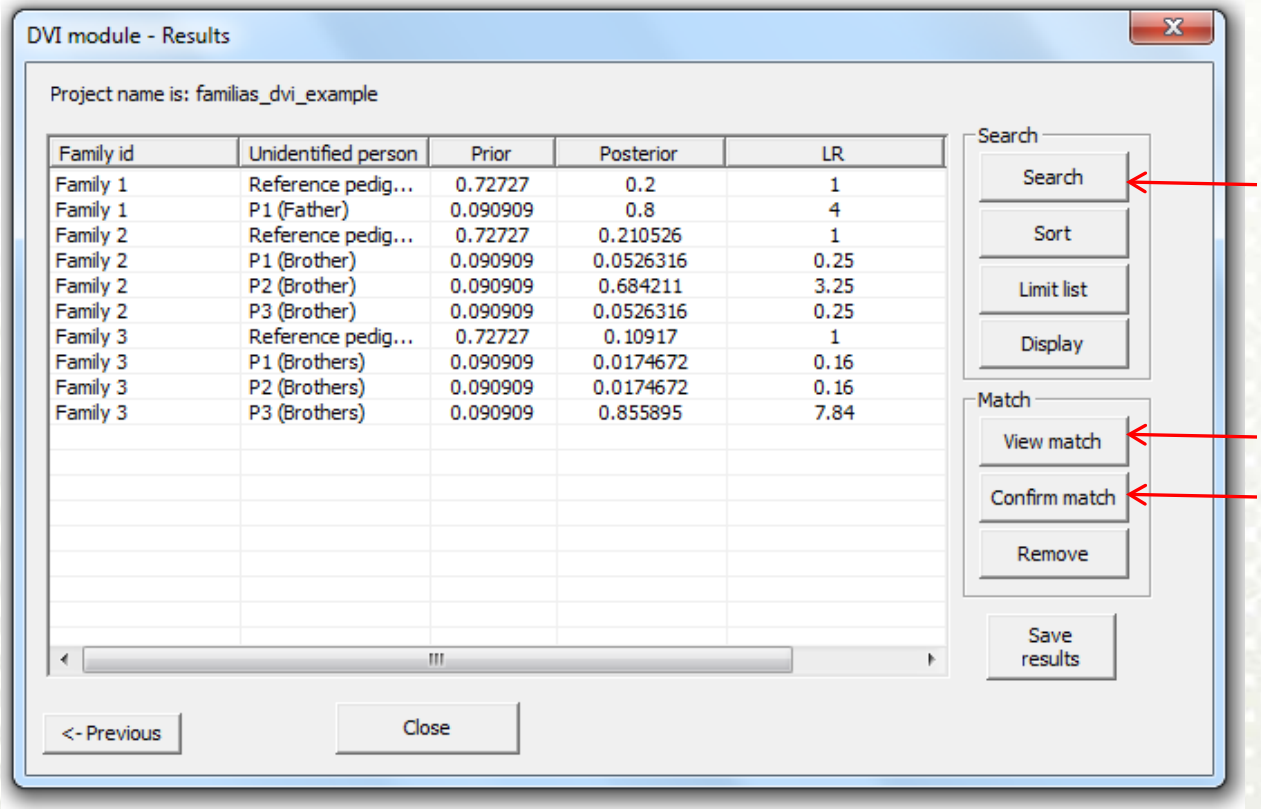
Familias 3.0 – DVI example

- Two brothers as reference persons!?



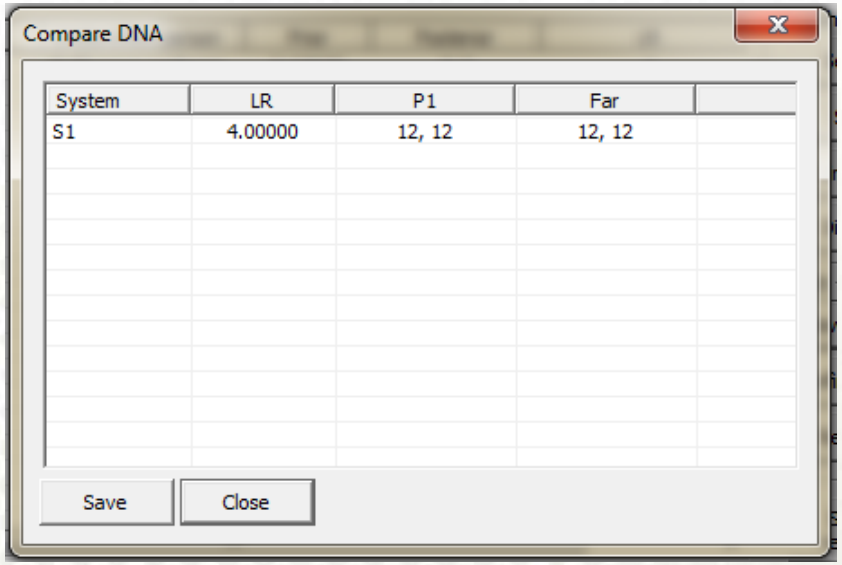
Familias 3.0 – DVI example

- Matching



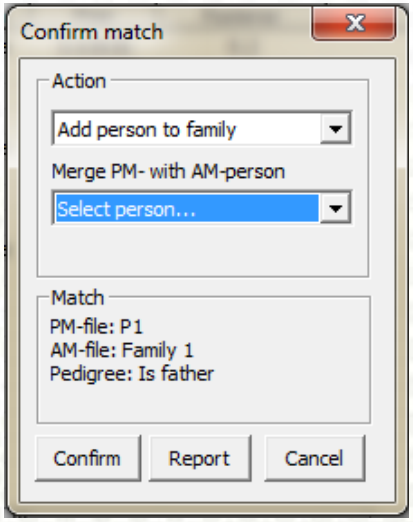
Familias 3.0 – DVI example

- View match



Familias 3.0 – DVI example

- Confirm match // Write report



Familias 3.0

- Questions?
- Exercises
- New file format

Familias 3.0 – Exploring new features

Daniel Kling
Norwegian Institute of Public Health *and* Norwegian University of Life Sciences
daniel.kling@fhi.no

Copenhagen, Oct 8-9, 2013