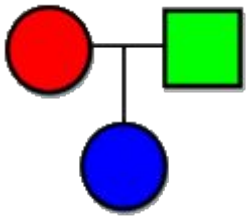


Familias - Tutorial

By: Daniel Kling, Lourdes Prieto and Thore Egeland

Updated 2022-11-15

Preface

This tutorial describes the basic features of Familias. The current version of the tutorial is based on version 3.3, but should work reasonably well for versions 3.2.2 and upwards.

For a more comprehensive description and theory we refer to the manual (not updated) available at http://www.familias.no or publications listed at the indicated link.

For Spanish translation of the current document please see http://familias.name/tutorial/familias_tutorial_spanish.pdf

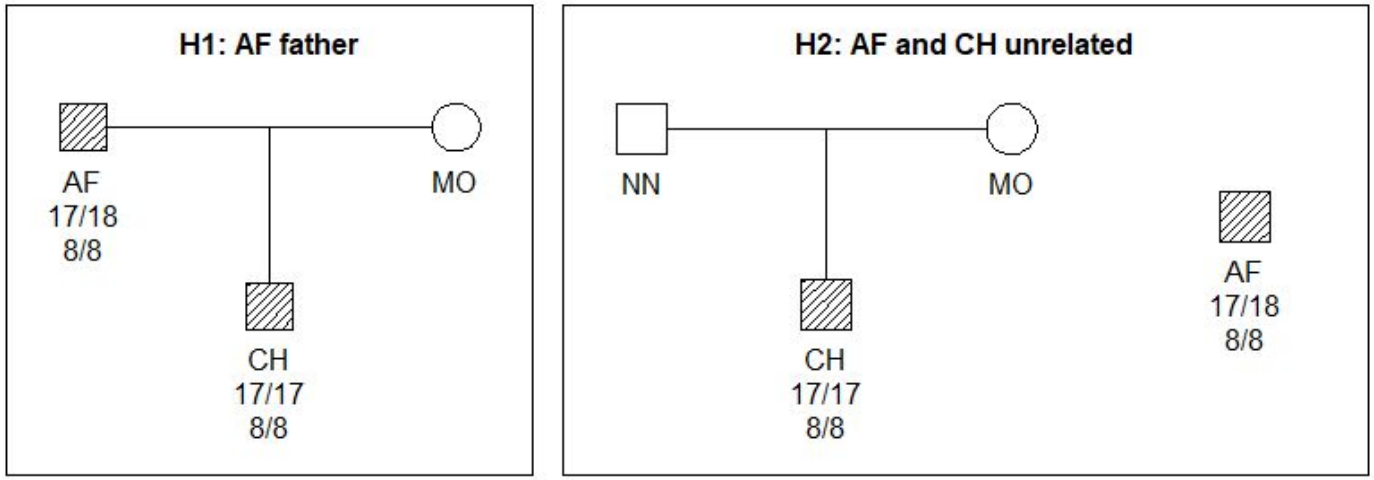
For videos including solutions to exercises in Egeland, Kling, Mostad (2015), see https://familias.name/VideosBook.pdf

Contents

1. Basics. A paternity case in four steps

Input file: tutorial-Ch1.fam

Example used to introduce Familias



$$LR = \text{Likelihood ratio} = \frac{P(\text{data} \mid \text{H1})}{P(\text{data} \mid \text{H2})}$$

$$LR_1 = \frac{2p_{17}p_{18} \cdot \frac{1}{2}p_{17}}{2p_{17}p_{18} \cdot p_{17}^2} = \frac{1}{2p_{17}} = \frac{1}{2 \cdot 0.204} = 2.45,$$

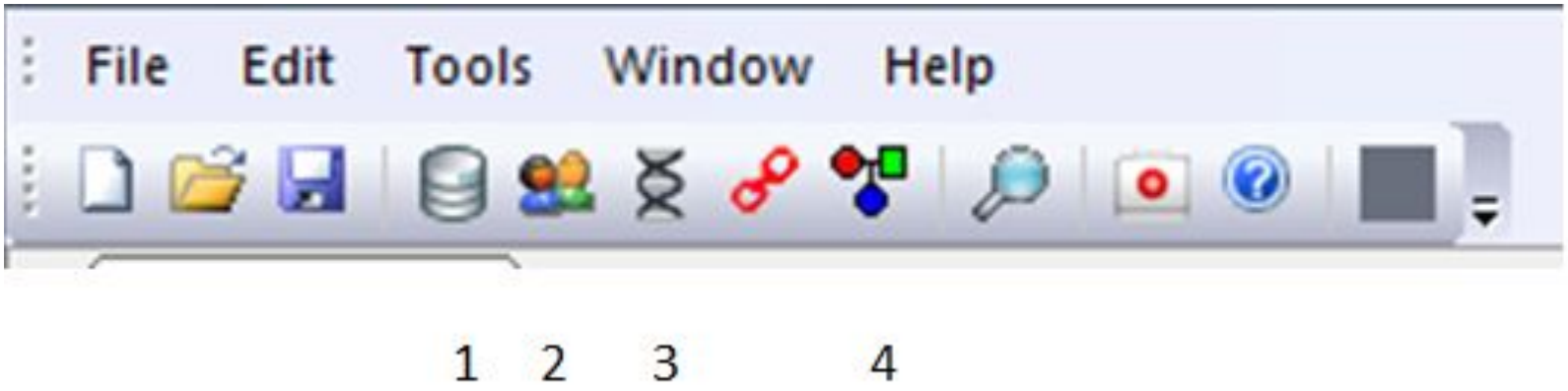
$$LR_2 = \frac{p_8^2 \cdot 1 \cdot p_8}{p_8^2 \cdot p_8^2} = \frac{1}{p_8} = \frac{1}{0.554} = 1.81,$$

$$LR = LR_1 \cdot LR_2 = 2.45 \cdot 1.81 = 4.4.$$

Interpretation: The data is 4.4 times more likely if H1 is true rather than H2.

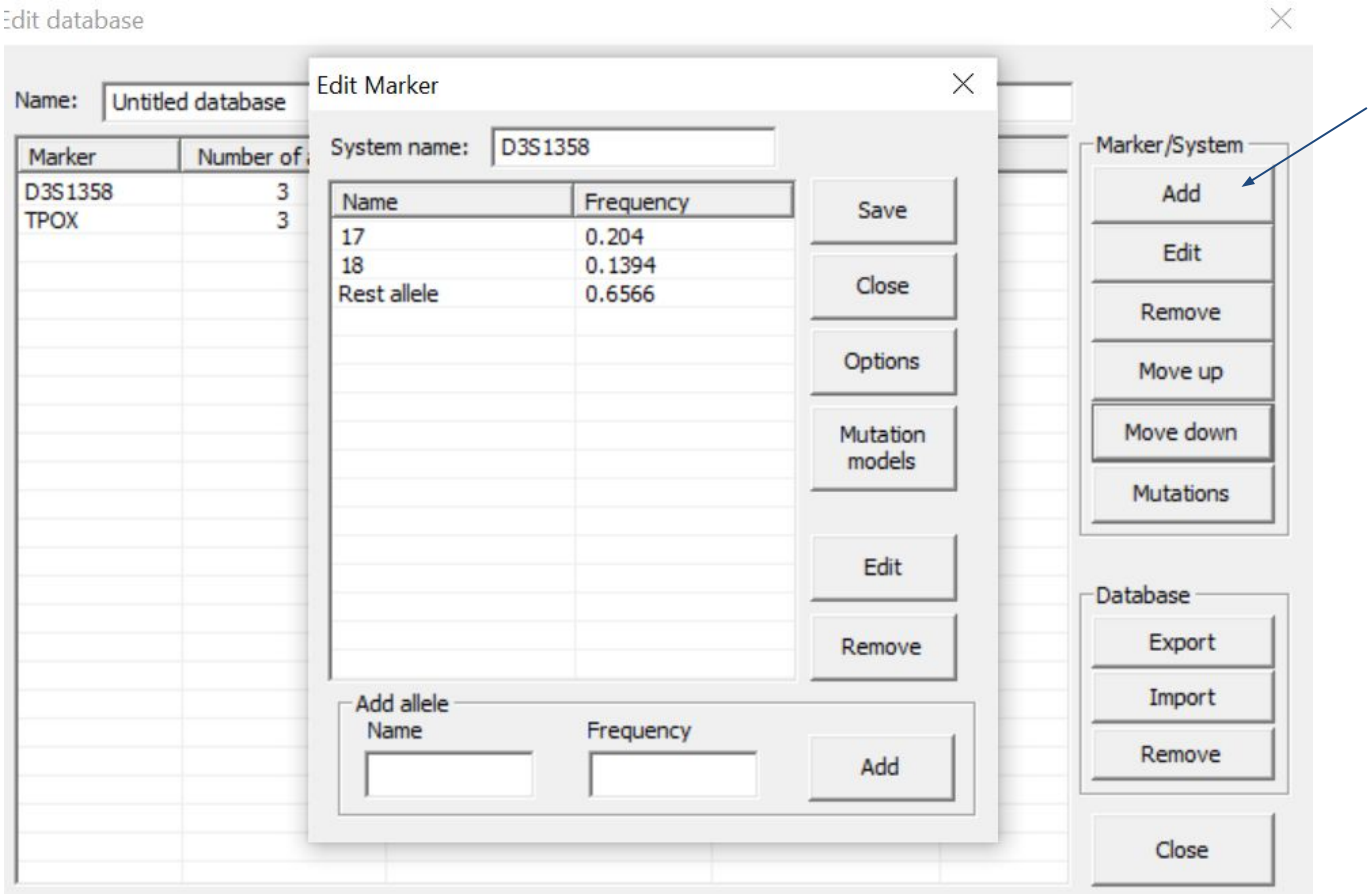
We next explain the steps needed to verify the calculations using Familias.

Four basic steps



1. **General DNA data**. Input of database, i.e., allele frequencies, etc.
2. **Persons**. Individual needed to define pedigrees.
3. **Case DNA data**. Marker data.
4. **Pedigrees**. Define hypotheses and do calculations.

Step1: General DNA data

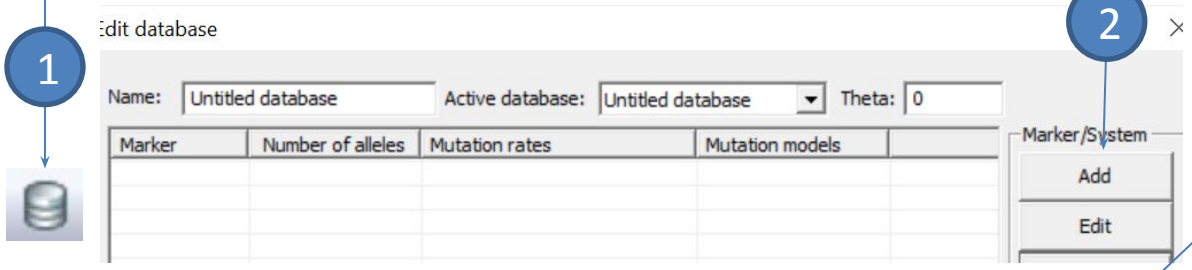


- Click ‘Add’ to enter a marker. In the new window, enter as shown.
- Similarly for next marker TPOX with alleles 8 and 9, frequencies 0.554 and 0.104.
- The order of markers can be modified by using “Move up” or “Move down” buttons.

Step 1 in detail: General DNA data

Click

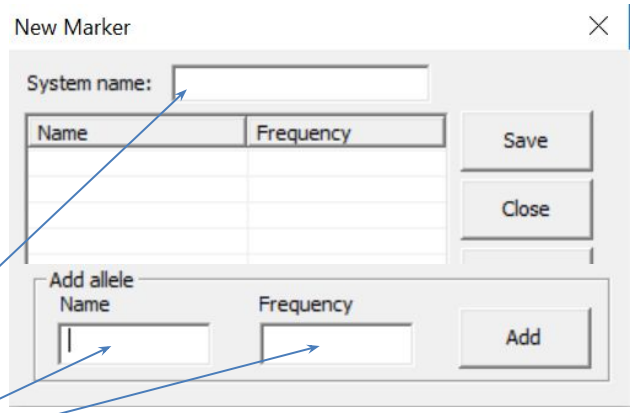
1



Click

2

The "New marker" window opens

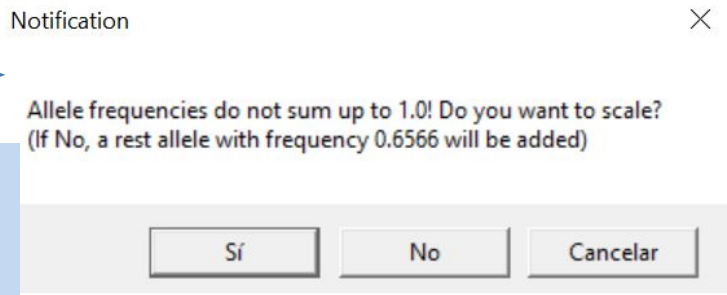


3 Add the name of the first marker (D3S1358 in the example)

4 Add the first allele (17), frequency (0.204) and press 'Add'

5 Do the same for the second allele (18, freq. 0.1394)

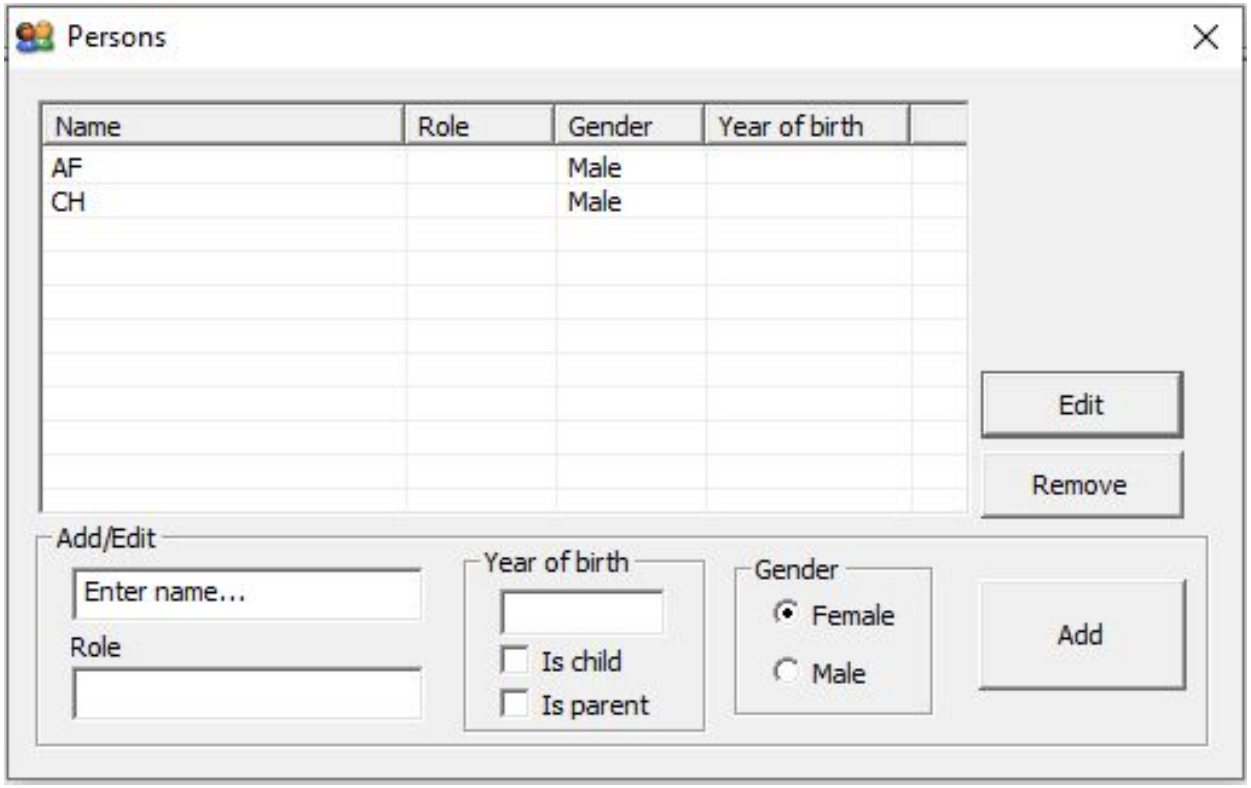
6 Click 'Save' and it will appear:



7 For "Scale", click yes (the freq. of both alleles will change to sum 1)
For "add a rest allele", click No (a rest allele with freq. 0.6566 will be added)

8 Repeat the process for the second marker (TPOX, in the example)

Step 2. Persons



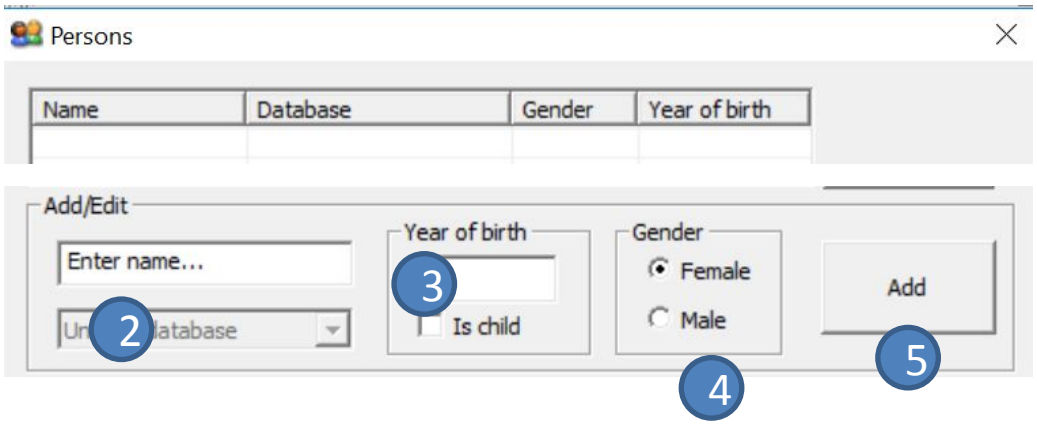
- Enter the persons: AF (alleged father), and CH (child) as shown.
- Information on ‘Role’, ’Year of birth’, ‘Is child’ and ‘Is parent’ are normally not needed nor used. Is parent is currently not used in any routine.

Step 2 in detail: persons

Click

1

The “Persons” window opens



2 Enter the name of the first person (AF)

3 Optional: enter the year of birth

4 Enter the gender

5 Click ‘Add’

6 Do the same with the following person (CH)

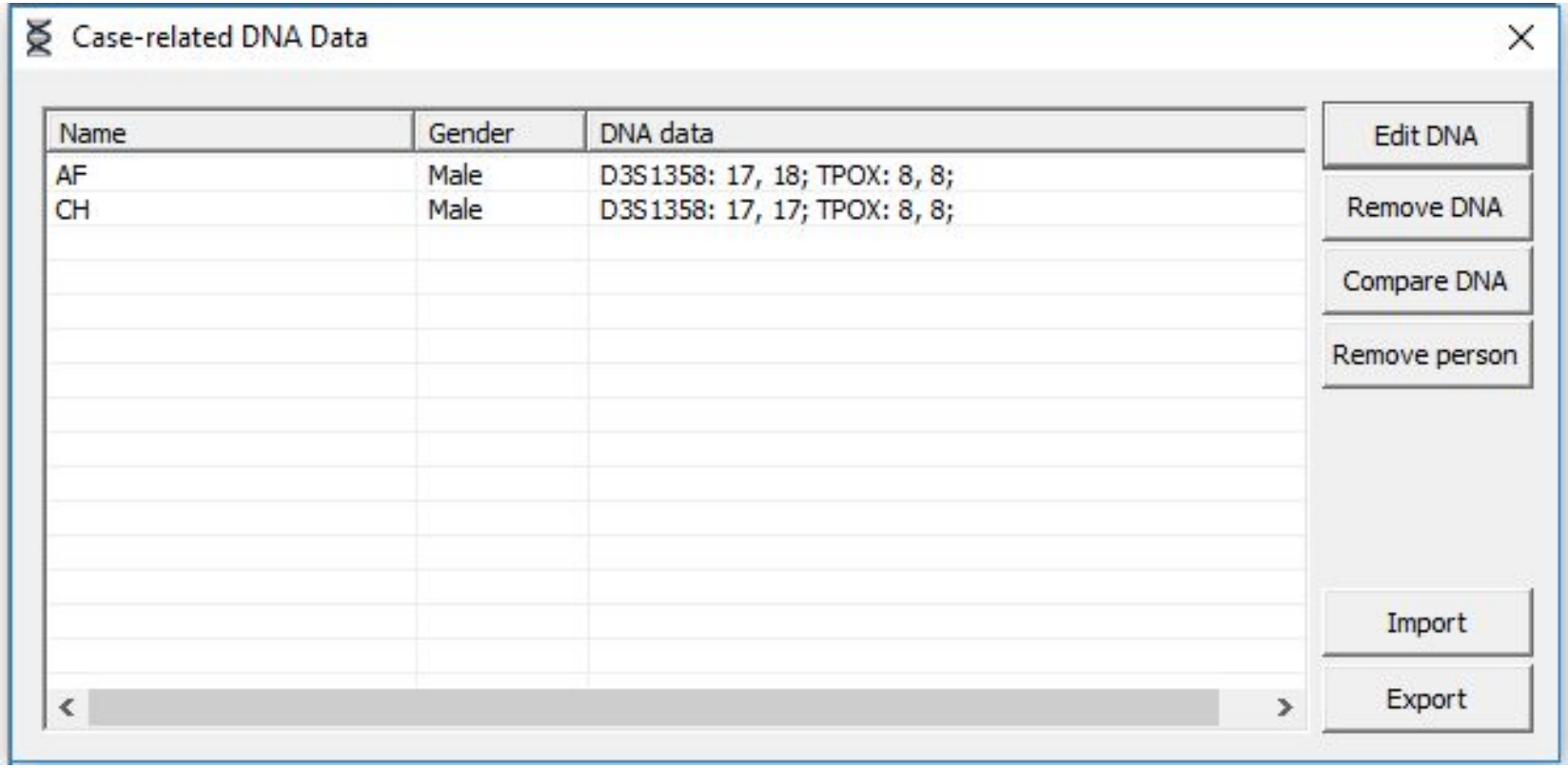
Comment 1

If you enter the year of birth, you make sure that a younger person can not be the parent of an older person (useful to avoid errors in the pedigrees and impossible pedigrees)

Comment 2

If you click the “Is child” box, you make sure that this person cannot have children (useful for the same reasons as before)

Step 3. Case data



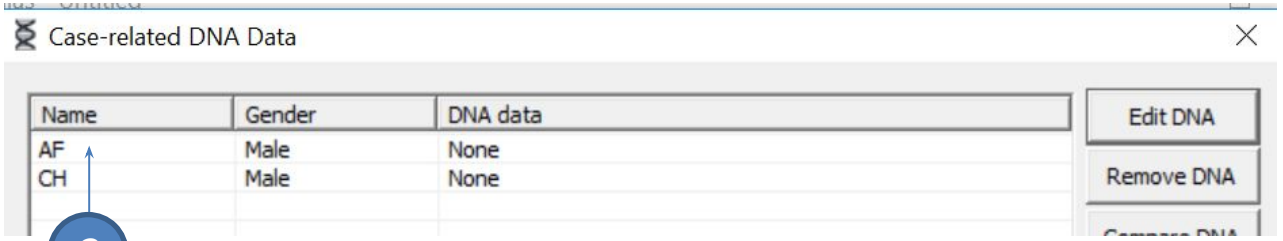
- Double click each person. In the new window select the marker in the pull down menu, press ‘Add’ and ‘OK’, to enter the data as shown.

Step 3 in detail: Case data

Click

1

The “Case-related DNA data” window opens

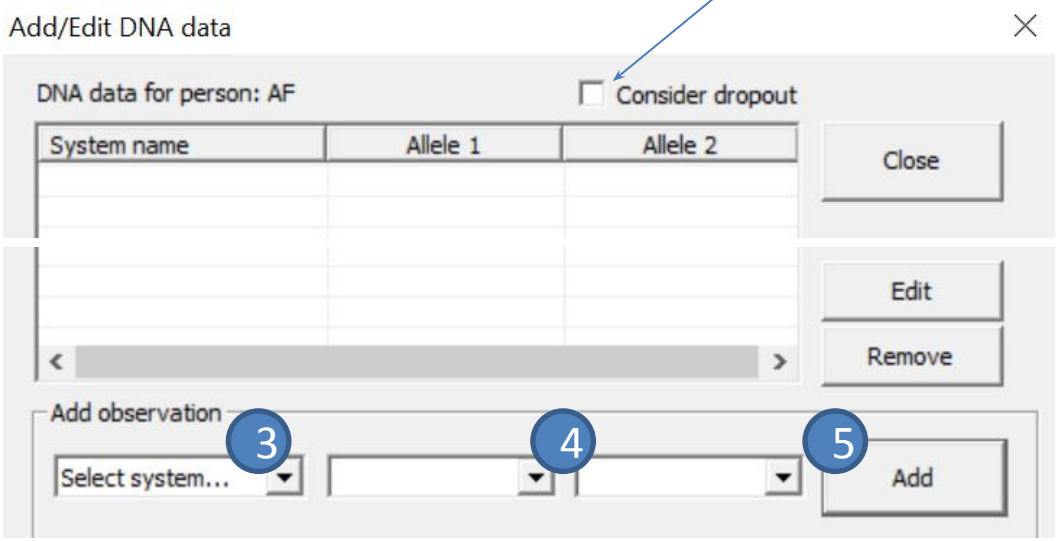


2

Double Click (AF)

Comment
If you want to consider dropout in one specific individual, click the box. See slide 18 to enter the probability of dropout

The “Add/Edit DNA data” window opens



3 Select the marker (D3S1358)

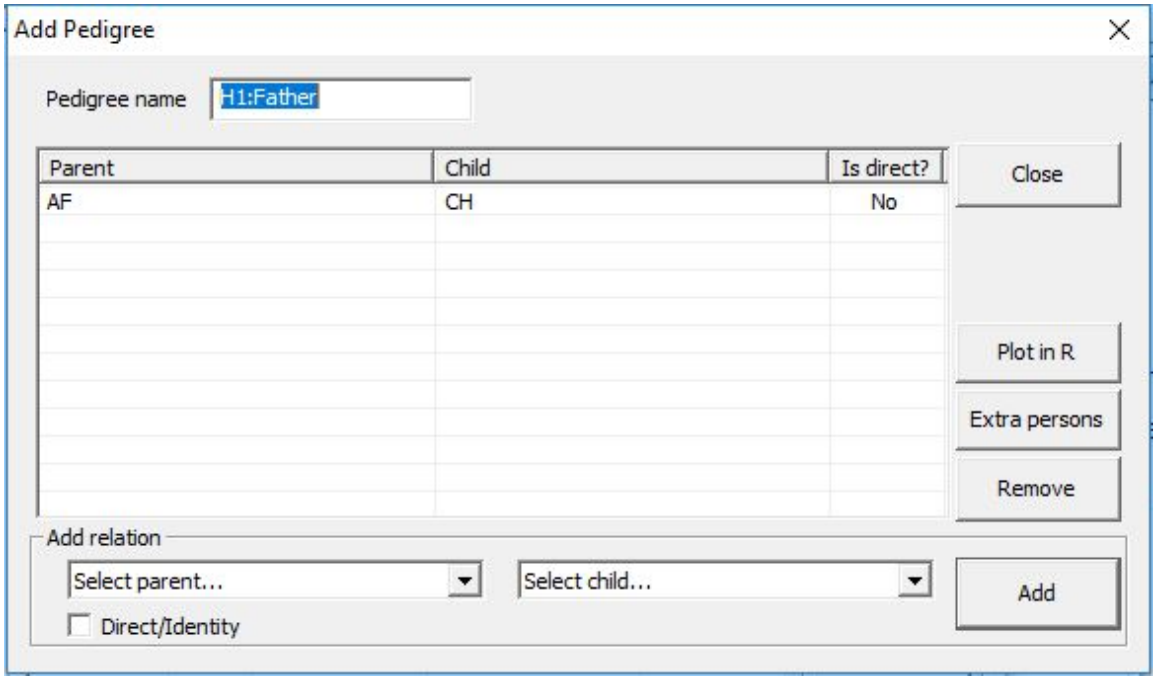
4 Enter the genotype (17-18)

5 Click ‘Add’

6 Do the same for TPOX and then close

7 Repeat the process from 2 for CH

Step 4. Pedigrees



Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
H1:Father	0.5	0.815639396	4.424152332	-6.92145
H2: Unrela...	0.5	0.184360604	1	-8.408529

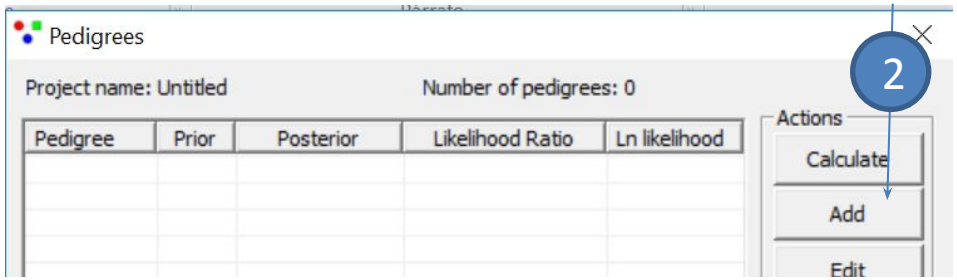
- Click ‘Add’ to enter the hypothesis “H1:Father” as shown (upper panel).
- Enter the pedigree “H2:unrelated”. In this case no relations are added.
- Click ‘Calculate’ to get the output in the lower panel.

Step 4 in detail: Pedigrees

Click

1

The “Pedigrees” window opens



Click

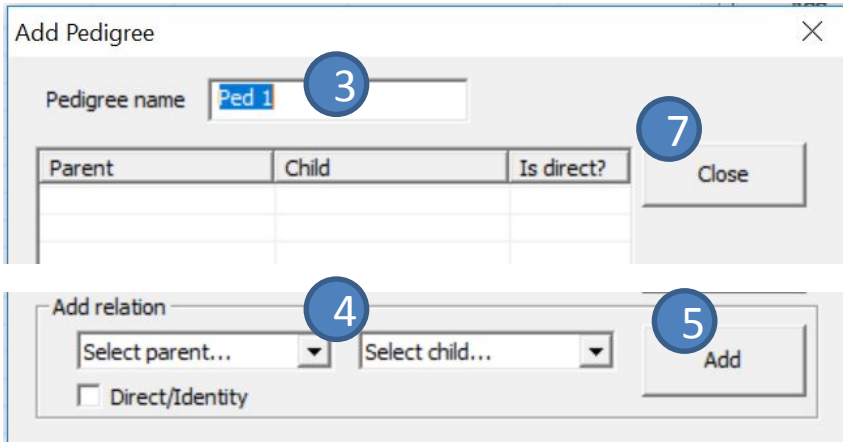
2

Comment

Familias only allows you to define parent-child relationships. If you want to define the relation “2 brothers”, you have to enter a mother and a father and establish that both are sons of them.

In order to define a disputed direct match, e.g. twins, use the Direct/Identity option

The “Add Pedigree” window opens



3 Name the pedigree (e.g., H1: Father)

4 Define the relation (AF father of CH)

5 Click ‘Add’

6 Name the second pedigree (H2: unrelated). You do not have to enter any relation in this case

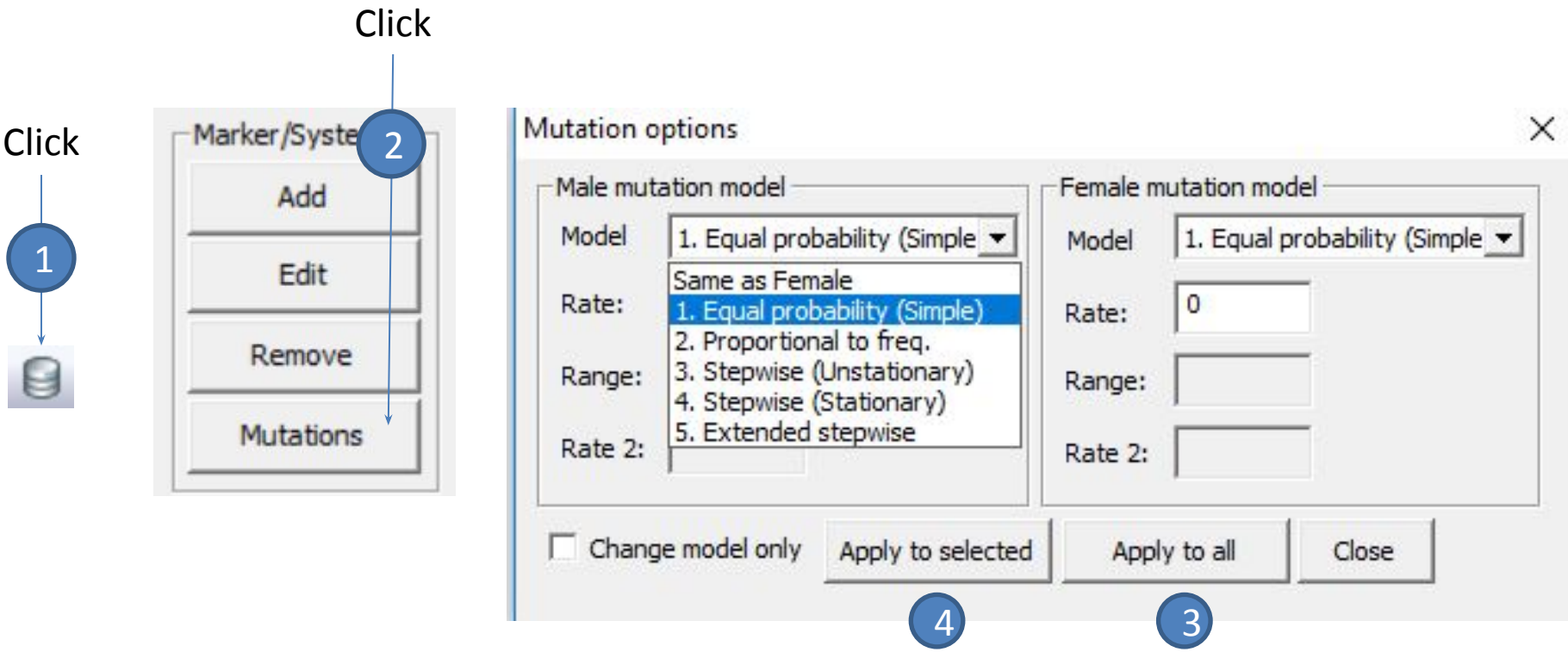
7 Click ‘Close’

8 Then click ‘Calculate’ in the Pedigrees window to obtain:

Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
H1:Father	0.5	0.815639396	4.424152332	-6.92145
H2: Unrela...	0.5	0.184360604	1	-8.408529

2. Complications

Mutations

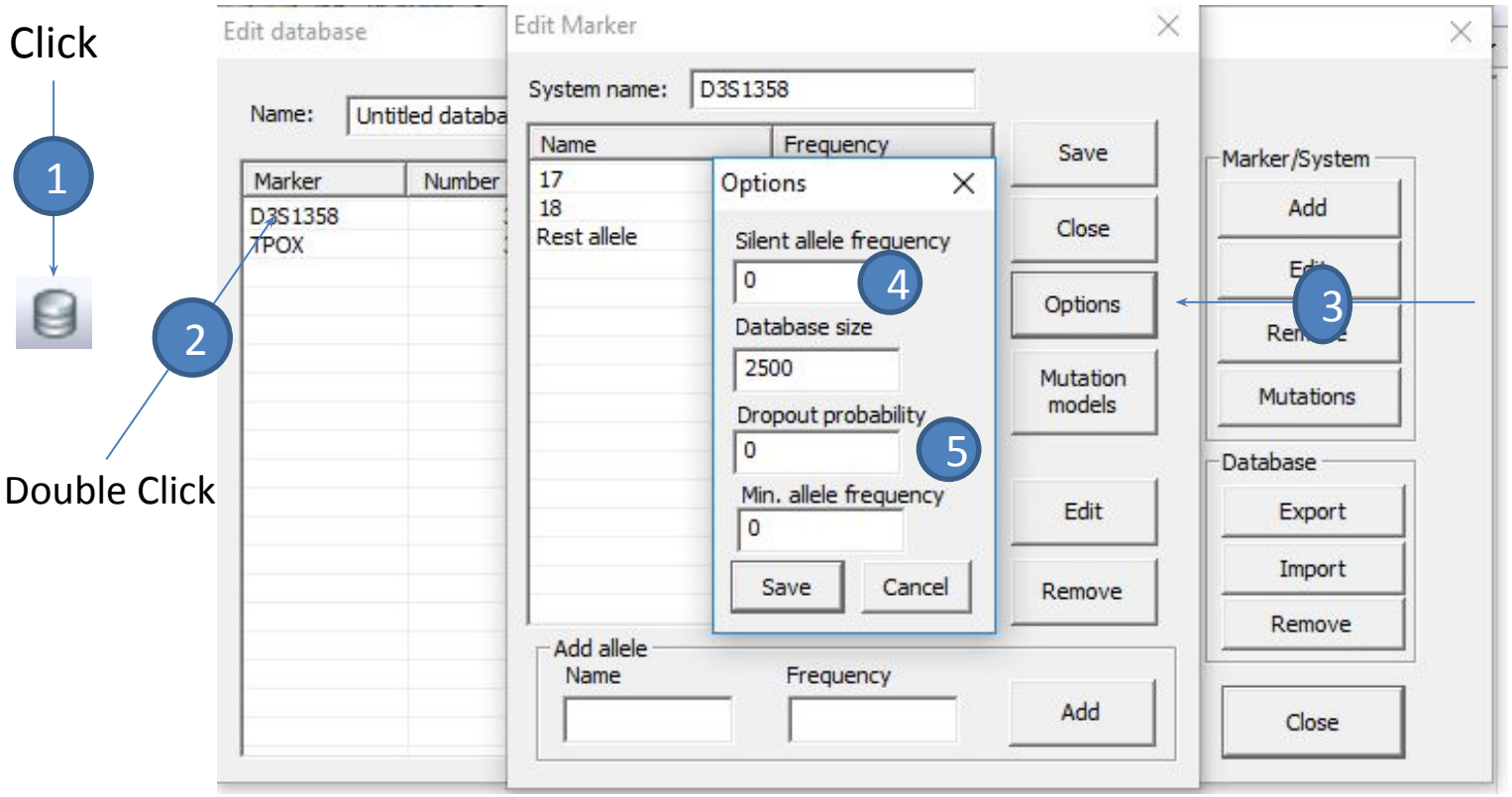


- Enter Step 1 (1), the Database window and press 'Mutations' (2) to get the above window.
- There are five models as shown (see the following slide for more info)
- One can assign the model to all markers (3) or only the ones selected (4) as shown.

Mutation models in Familias

- **Simple model**
  Each mutation has an equal probability of occurring. Appropriate for fast computations (e.g. complicated pedigrees and DVI) as well as for SNP markers.
- **Proportional to frequencies**
  Each mutation has a probability which is proportional to the frequency of the allele we are transitioning to. A low frequency will yield a low mutation probability. Appropriate for testing of statistical properties of calculations.
- **Stepwise model**
  The traditional stepwise model where the probability depends on the number of steps from the original allele to the mutated allele, decreasing probability for longer repeat mutations. Appropriate for STR markers without microvariants.
- **Stepwise stable model**
  The traditional stepwise model where the mutation probabilities have been adjusted to create a stable mutation matrix. Appropriate for testing of statistical properties of calculations.
- **Extended stepwise model (RECOMMENDED)**
  The most complete stepwise model, where both exact repeats as well as microvariants are accounted for. Appropriate for all STR markers.

Silent alleles and Dropout

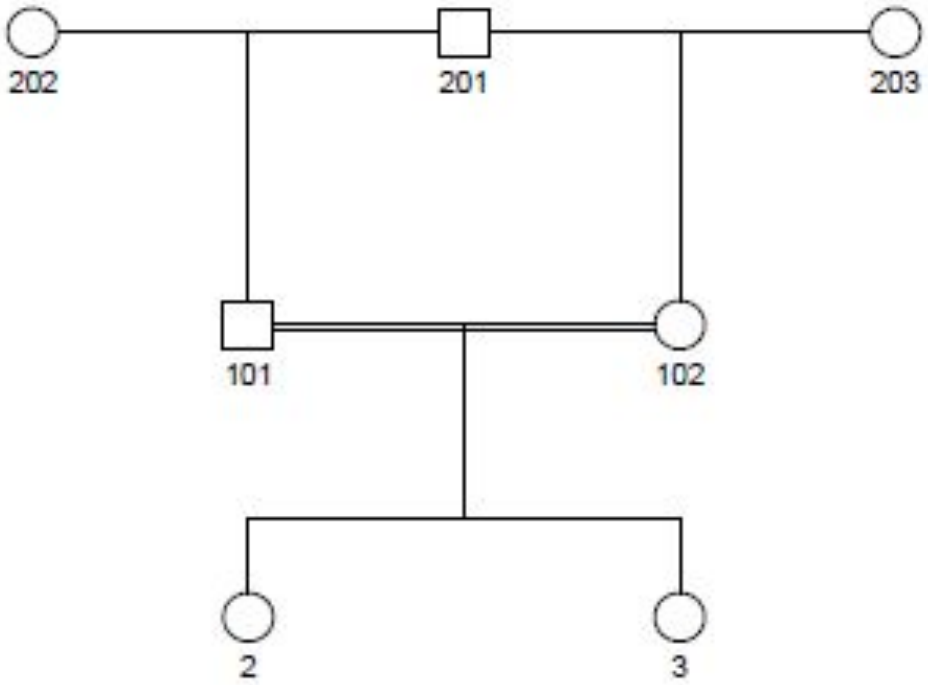


Comment 1
Note that changing the Database size won't affect the results.
Comment 2
Defining a non-zero Min. allele frequency only takes effect once Advanced option is checked

- Enter the Database window (1), double click the marker to edit (2).
- Clicking Options (3) gives the above window where the silent allele frequency (4) and/or Dropout probability (5) can be entered.
- For Dropout , further input is required in the Case DNA to specify the individuals susceptible to dropout (see slide 12).

Inbreeding: Complex pedigrees

- Draw pedigree
- Identify extra persons needed to define the pedigree.
- Enter required persons (Step 2) and pedigrees (Step 4).



Pedigree name : Sisters

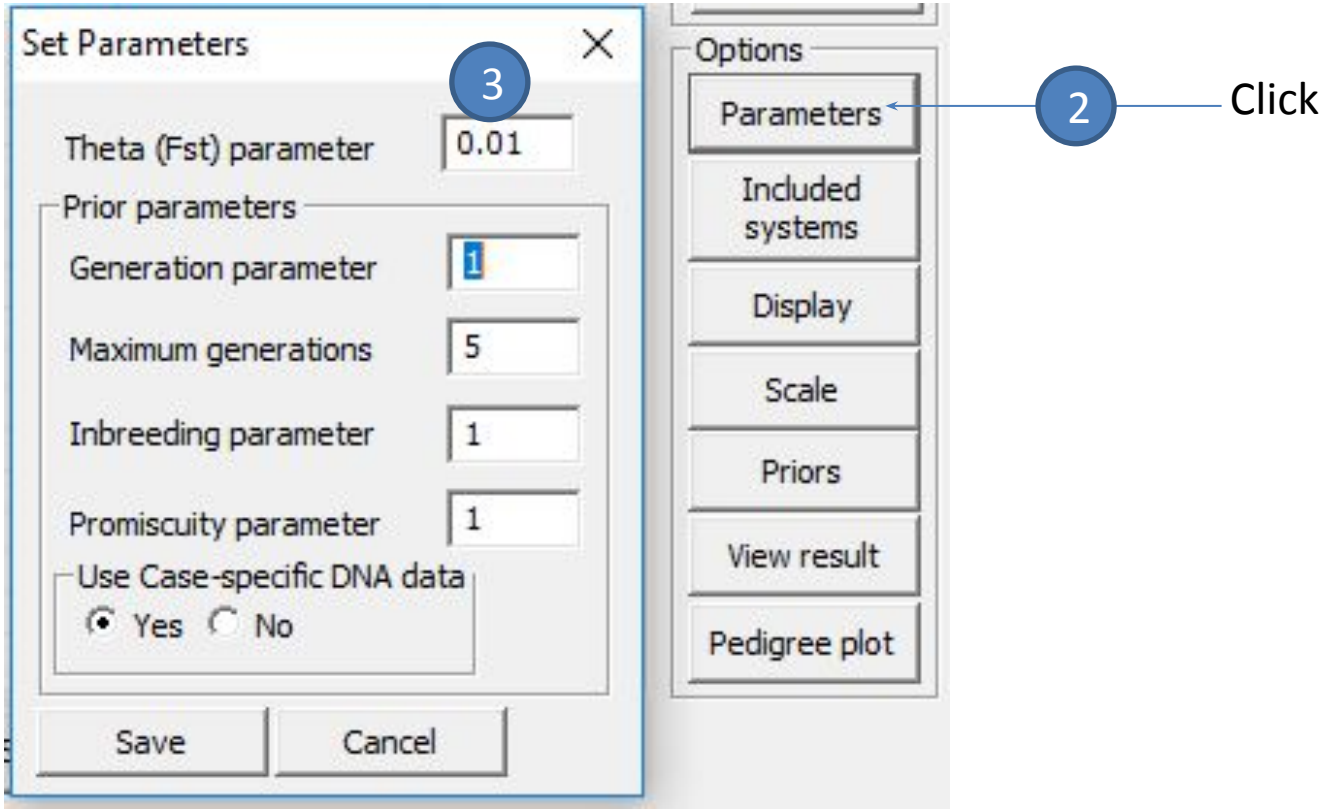
Parent	Child
202	101
201	101
201	102
203	102
101	2
101	3
102	2
102	3

Comment: Persons 202 and 203 above are not needed, but plotting functions will introduce them. The double lines connecting 101 and 102 indicate inbreeding, i.e., parents are related within the pedigree.

Click

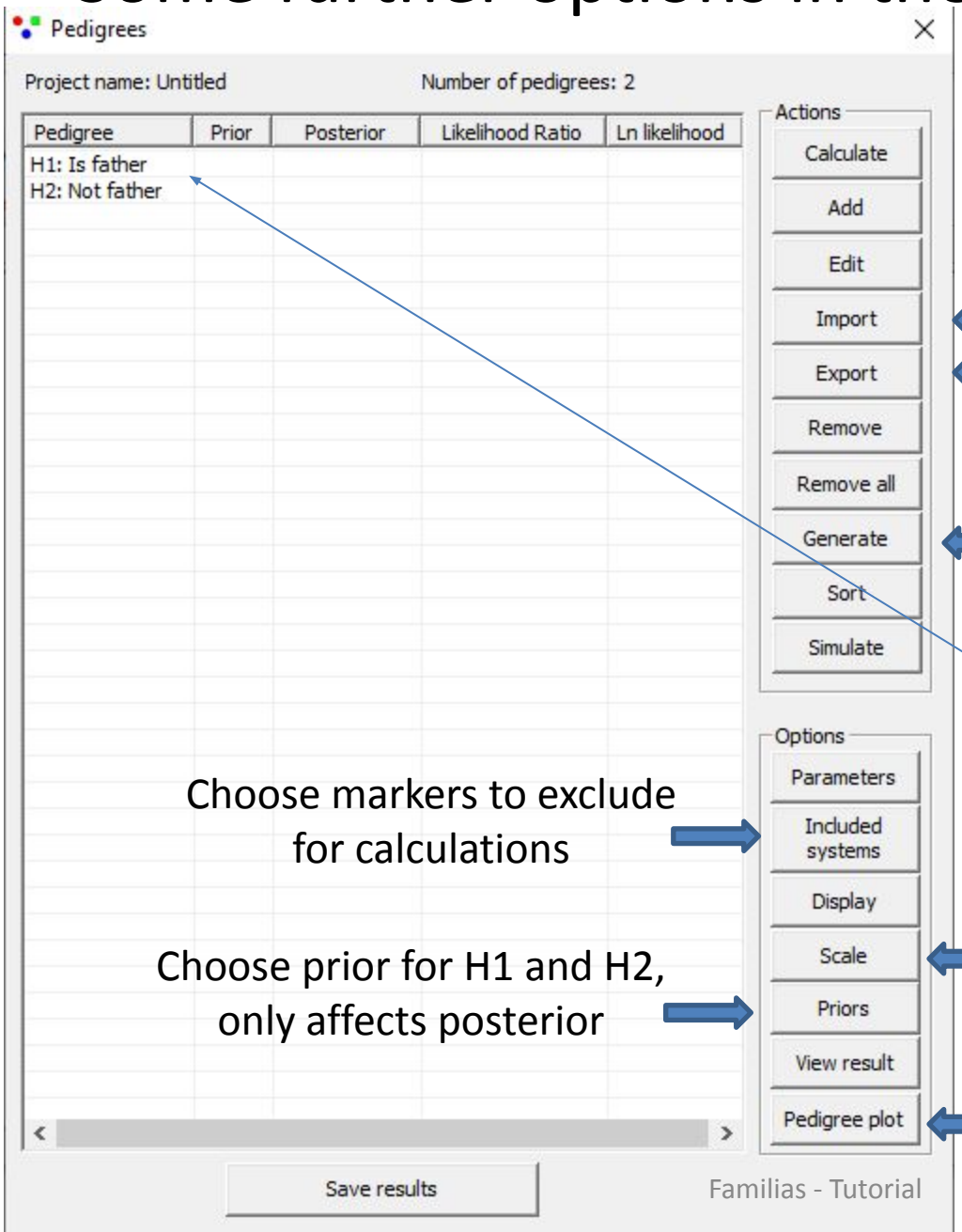
1

Theta correction



- Click ‘Parameters’ in the Pedigrees window (2) . Above the value 0.01 is entered (3).
- ‘Prior parameters’ are virtually never changed: they have no impact on LR, only on the prior (and hence the posterior).

Some further options in the Pedigree window



Imports pedigrees, see next page

Exports pedigree file

Generate pedigrees, see next slides

Choose markers to exclude for calculations

Choose prior for H1 and H2, only affects posterior

Select denominator of LR: marking H1 and pressing scale gives LR for H2 vs.H1

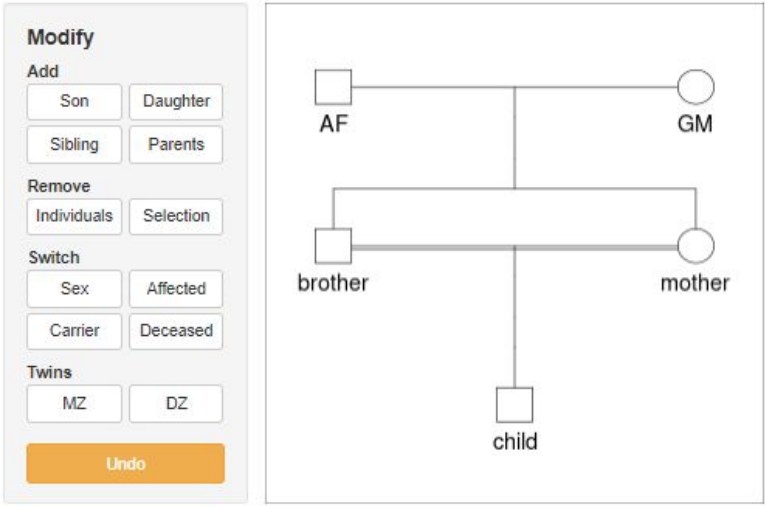
Plots pedigree, see Section 3

Creating pedigrees in QuickPed, import to Familias

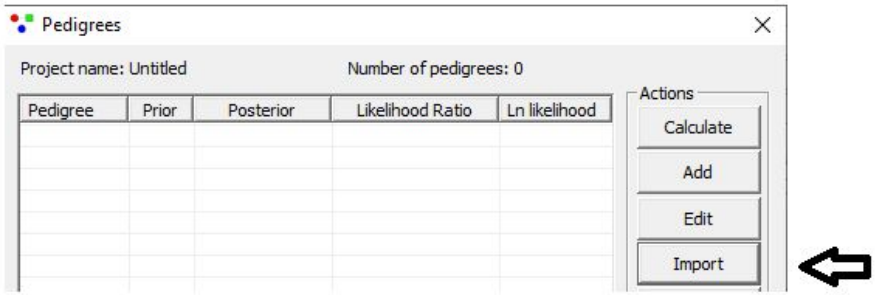
- The next slide shows how the app QuickPed (Vigeland, 2022), an online tool for drawing pedigrees and analysing relatedness, can be used to create pedigrees.
- The pedigrees can be imported to Familias as explained next.

QuickPed: An Interactive Pedigree Creator

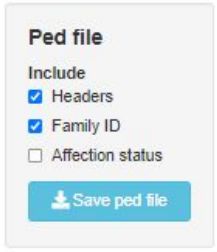
1



4



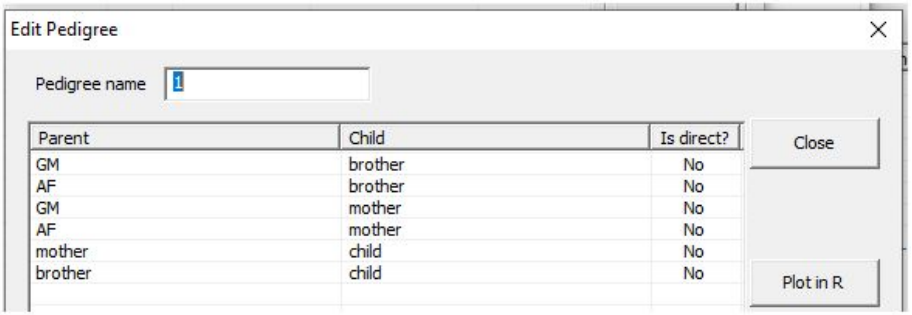
2



3

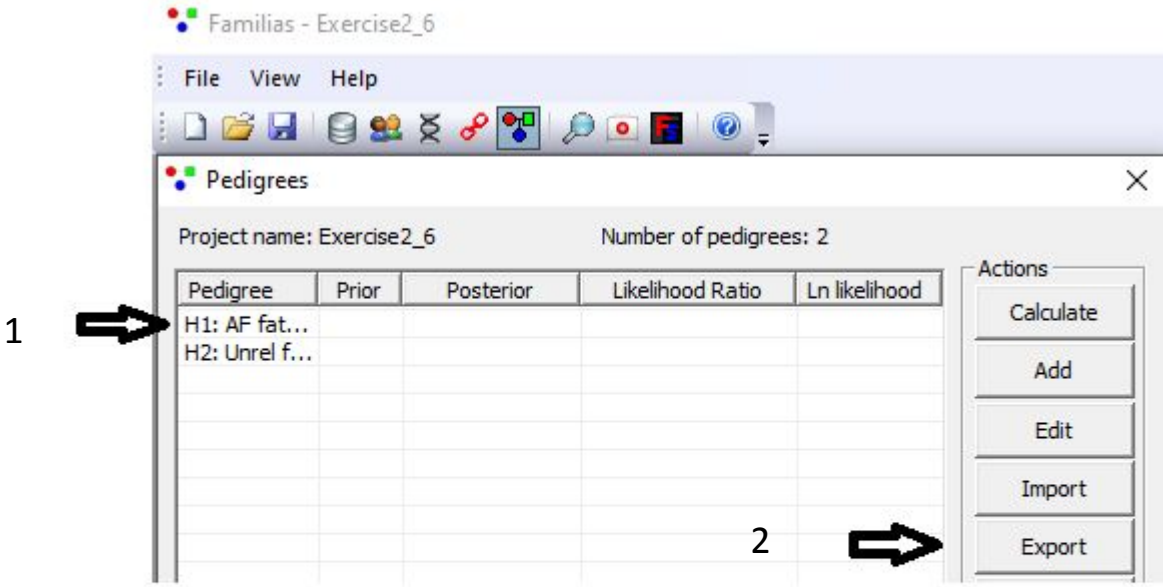
```
famid id fid mid sex
1 AF 0 0 1
1 GM 0 0 2
1 brother AF GM 1
1 mother AF GM 2
1 child brother mother 1
```

5



1. Create pedigree in e.g. QuickPed https://magnusdv.shinyapps.io/quickped/
2. Save ped file from QuickPed
3. Optional look at ped file tutorial.ped
4. Import tutorial.ped to Familias
5. Ped file in Familias

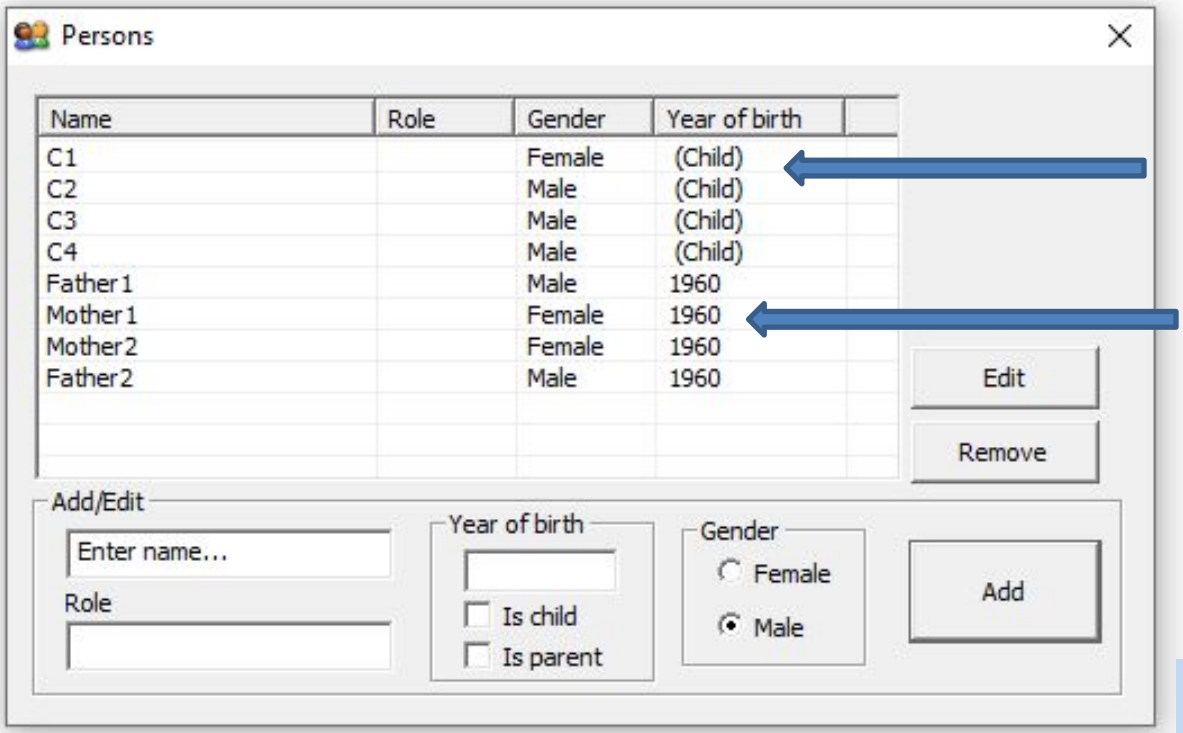
Export from Familias to QuickPed and plot



1. Highlight pedigree. Pedigree must be connected, i.e., none or both åparents must be present
2. Hit `Export`
3. `Load file` in QuickPed

Generate pedigrees

Question: What is the relationship between four children?
Full siblings, half siblings or unrelated?



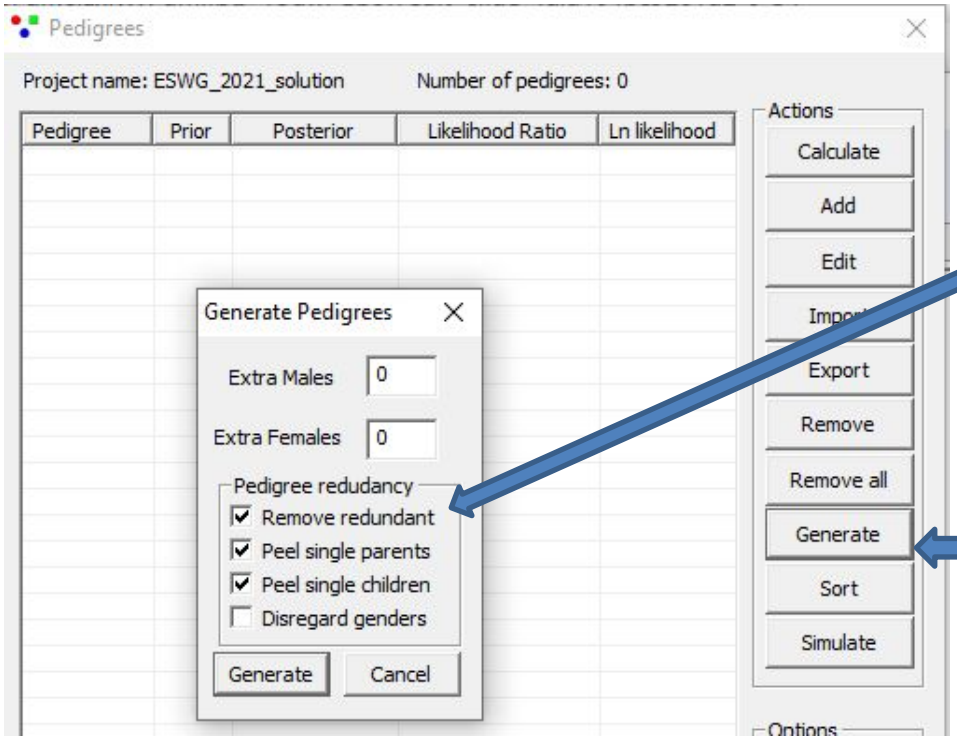
Define children (cannot be parents)

Define parents (born same year so cannot be parents of each other, avoid multiple generations)

Comment
See ESWG proficiency test for 2021: exercise with video and presentation

Generate pedigrees

Question: What is the relationship between four children?
Full siblings, half siblings or unrelated?



Generate options.

a) Include extra males/females (not recommended)
b) Remove redundant pedigrees (recommended)
c) Peel single parents (recommended)
d) Peel single children (recommended)
e) Disregard genders (optional)

Generate pedigrees

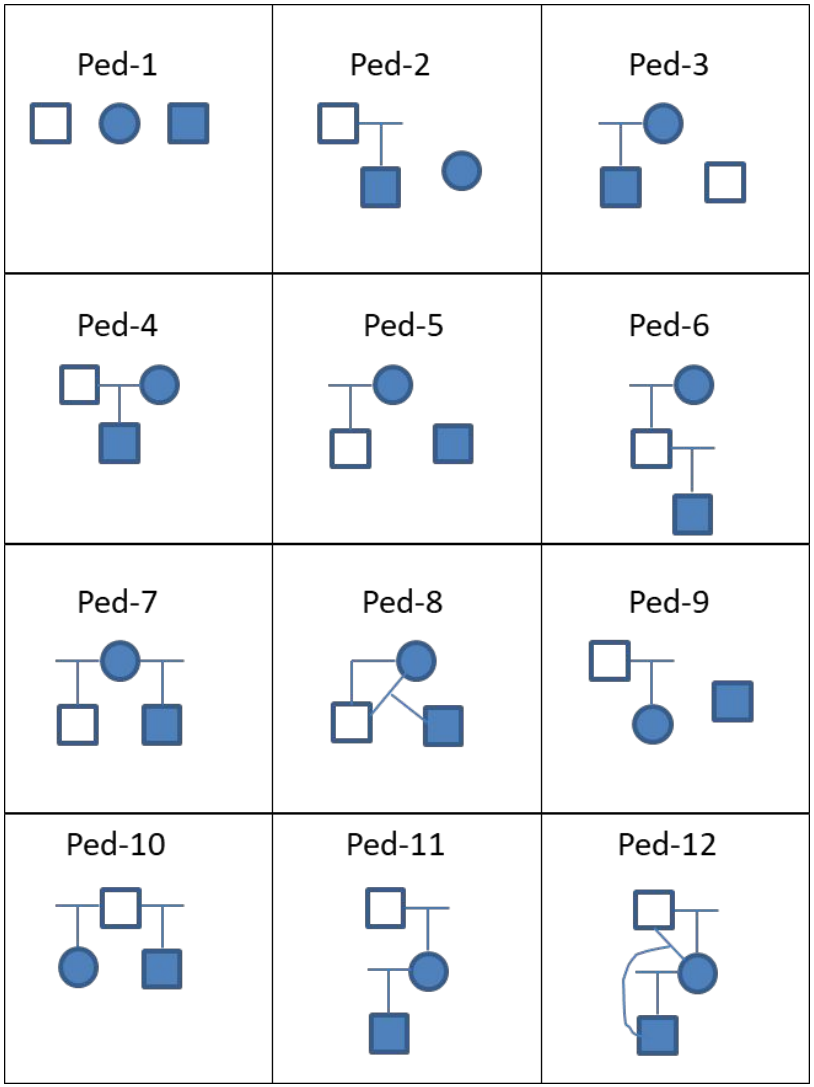
Generate pedigrees

Question: What is the relationship between four children? Full siblings, half siblings or unrelated?

Familias will generate all possible pedigrees with restrictions imposed by the defined persons and options. It will also, for example, generate pedigrees where an individual has a single parent, both not genotyped. This is removed with the two peel options given. Disregard genders option will consider maternal and paternal half siblings as identical and will thus consider such events as duplicates.

You can restrict the number of pedigrees generated if some relations are known in the 'Known relationship' window:

Generate pedigrees with restrictions: example



Freq 1 = 0.1; Freq 2 = 0.9; no mutations

The “Generate” function by default (without restrictions) gives the 12 pedigrees to the left (input file peel.fam)

- Recommended restrictions
  - Remove redundant
  - Peel single parents
  - Peel single children

  give the pedigrees 1, 3, 6, 8, 10 ,12
- Inbreeding parameter 0 gives 0 likelihood for pedigrees 8 and 12.

3. R Familias, paramlink and plotting

What is  ?

- A framework for statistical and numerical computing
  - calculator
  - flexible plotting
  - large core of functions for data handling and numerical analysis
  - programming language
  - external packages
    - anyone can make one
    - thousands!

- About R:
  - it's freely available from https://cran.r-project.org/
  - it's widely used
  - it *can* do anything (but it may not be easy)

Installing and loading the R library Familias

To access the functions of an external library, exemplified with Familias below, you must:

- install the package, this done only once. On most platforms the following will work

  (further details are here):

  - install.packages("https://familias.name/RFamilias/Familias_2.5.zip", repos = NULL, type = "win.binary")
    #(internet connection needed)

- load it into R every new session:
  - **library(`Familias`)**

To obtain help, enter

- **help(Familias)**

Sample session from help ("FamiliasPosterior"): copy and paste to R!

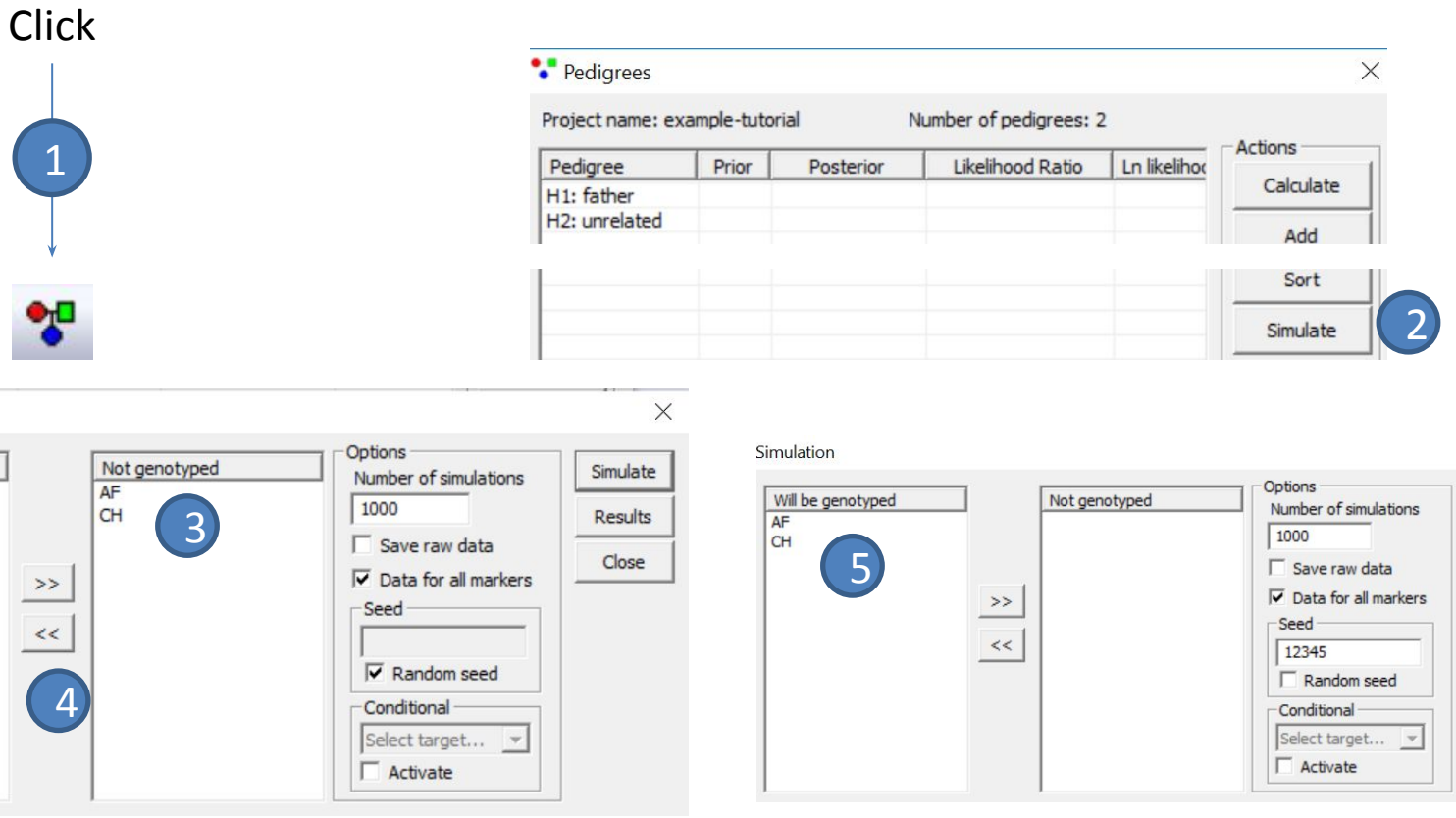
```
library(Familias)
persons <- c("mother", "daughter", "AF")
ped1 <- FamiliasPedigree(id=persons, dadid=c(NA, "AF", NA), momid=c(NA, "mother", NA),
                  sex=c("female", "female", "male"))
ped2 <- FamiliasPedigree(id=c(persons, "TF"), dadid=c(NA, "TF", NA, NA),
                  momid=c(NA, "mother", NA, NA),
                  sex=c("female", "female", "male", "male"))
ped3 <- FamiliasPedigree(id=c(persons, "TF", "gf", "gm"), dadid = c(NA, "TF", "gf", "gf", NA, NA),
                  momid=c(NA, "mother", "gm", "gm", NA, NA),
                  sex=c("female", "female", "male", "male", "male", "female"))
X11()
plot(ped1);title("ped1, i.e., AF is father")
mypedigrees <- list(isFather = ped1, unrelated=ped2, isUncle = ped3)
locus1 <- FamiliasLocus(frequencies=c(0.1, 0.2, 0.3, 0.4),
                 allelenames= c("A", "B", "C", "D"), name="locus1")
locus2 <- FamiliasLocus(c(0.2, 0.3, 0.5), c(17, 18, 19), "loc2", femaleMutationRate = 0.05)
myloci <- list(locus1, locus2)
datamatrix <- data.frame(locus1.1=c("A", "A", "A"), locus1.2=c("B", "B", "C"),
                  locus2.1=c(17, 19, 19), locus2.2=c(18, 18, 18))
rownames(datamatrix) <- persons
result = FamiliasPosterior(mypedigrees, myloci, datamatrix,ref=2)
result
```

R code generated from Windows Familias

- Familias exports to **R** for
  - ✔ Plotting
  - ✔ Conditional simulation
  - ✔ Probability of exclusion
  - ✔ …
- Export is done from within Windows Familias;
  - ✔ File > Export to R-Familias (complete export)
  - ✔ 'Plot in R' in Edit Pedigree window (only plot)
  - ✔ Conditional simulation in DVI module

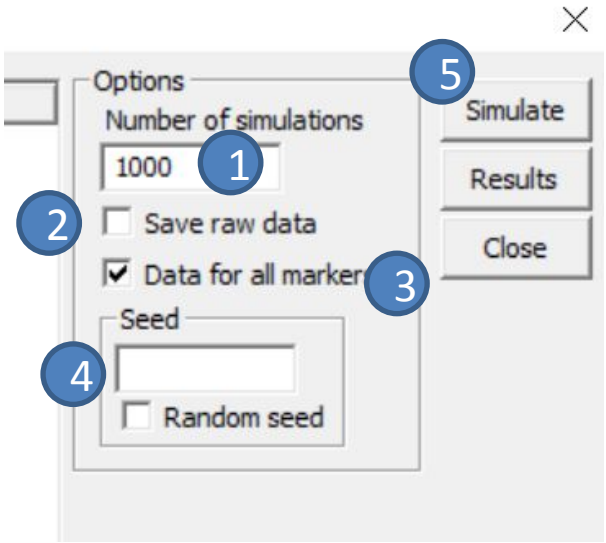
4. Simulations

Simulations

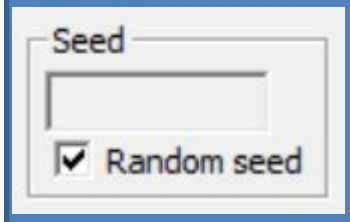


- In the Pedigrees window (1) , once you have defined the hypotheses you want to simulate, click 'Simulate' (2). The Simulation window opens.
- Select the individuals that will be genotyped (3) by clicking on them and then on the arrows (4). The individuals will be moved to the "will be genotyped" box (5).

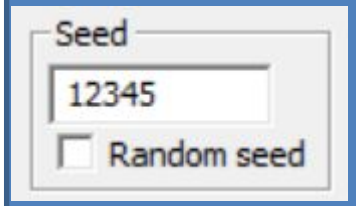
Simulations - Options



1. You can choose the number of simulations (1000 by default)
2. You can save the raw data (genotypes and/or LRs) in a .txt file
3. You can perform simulations with either all markers in the frequency database or only the ones selected in the Pedigrees dialog (Included systems)
4. You can specify where the simulations will start (Seed).



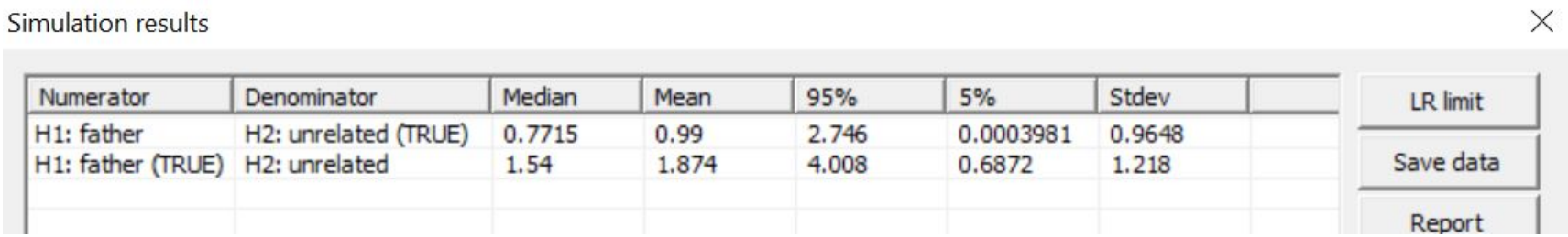
If "Random seed" is used, different simulations will be obtained each time you simulate.



If "Random seed" is unclicked, you have to enter a number, and you will obtain exactly the same results each time (given that all other values/parameters remain unchanged and using always the same seed number).

5. Click 'Simulate'

Simulations - Results


Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	0.7715	0.99	2.746	0.0003981	0.9648
H1: father (TRUE)	H2: unrelated	1.54	1.874	4.008	0.6872	1.218

LR limit
Save data
Report

Meaning of first line in the figure (H2: "AF and CH are unrelated" IS TRUE):
LR values = 0 are not expected as a mutation model is used (stepwise stationary, rate 0.001 and range 0.1) in the example. However, low LRs are generally expected as the true hypothesis is unrelated.
The mean of the obtained LR values is 0.99
50% of the simulations are below (and above) the median 0.7715
95% of the simulations are below 2.746
5% of the simulations are below 0.0003981

Meaning of line two in the figure (H1: "AF is the father of CH" IS TRUE):
The mean of the obtained LR values is 1.874
50% of the simulations are below the median 1.54
95% of the simulations are below 4.008
5% of the simulations are below 0.6872

You can use this file to reproduce these results

Comment
The LR values under both situations (H1 true and H2 true) overlaps. So, we will not be able to reach any conclusion about paternity if only these 2 markers (with 3 and 2 alleles) are analysed.

Simulation – A “more real” example

Same hypotheses, same simulation options (1000 simulations, all markers, Seed 12345), but 16 aSTR markers with real allele frequencies (PP ESX17)

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	4.051e-031	1.018e-009	8.159e-017	2.145e-047	3.045e-008
H1: father (TRUE)	H2: unrelated	4.418e+006	6.317e+008	1.101e+009	3.319e+004	5.737e+...

LR limit

Save data

Meaning of first line in the figure (H2: “AF and CH are unrelated” IS TRUE):
LR values = 0 are not expected as a stepwise stationary mutation model with mutation rate 0.001 and range 0.1 was used in the example. However, very low LRs are expected given that unrelated is the true hypothesis.
The mean of the obtained LR values is 1.018e-09
50% of the simulations are below the median 4.051e-31
95% of the simulations are below 8.159e-17
5% of the simulations are below 2.145e-47

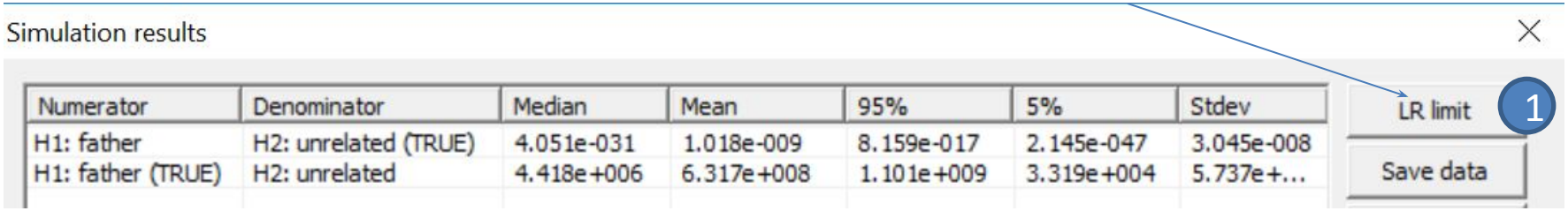
Comment
The LR values under both situations (H1 true and H2 true) DO NOT overlap. So, we will be able to reach a conclusion about paternity

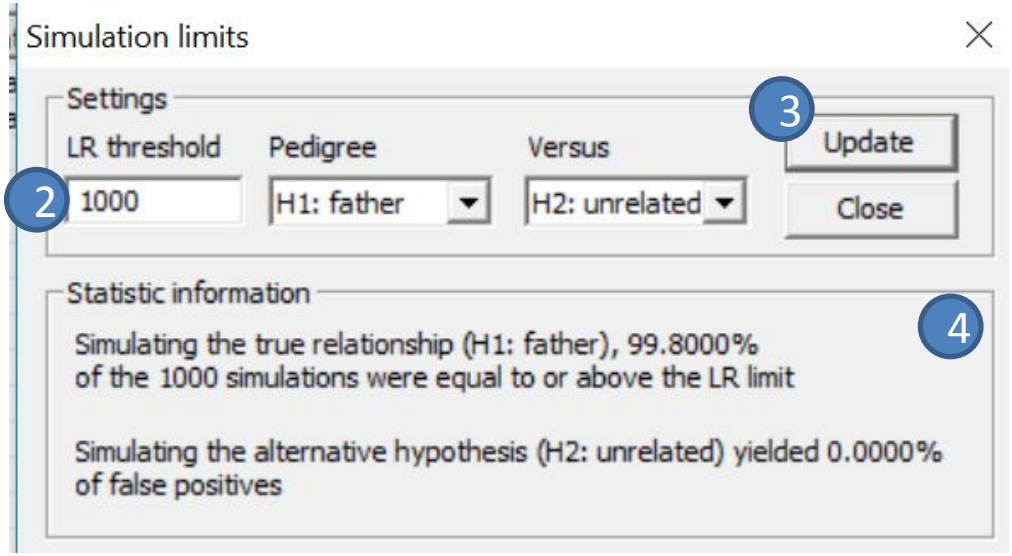
Meaning of line two in the figure (H1: “AF is the father of CH” IS TRUE):
The mean of the obtained LR values is 6.317e+08
50% of the simulations are below the median 4.418e+06
95% of the simulations are below 1.101e+09
5% of the simulations are below 3.319e+04

You can use this file to reproduce these results

Simulations – LR limit

Useful to display results in a way that may be easier to understand
The LR limit button is used to find the fraction of simulations exceeding an LR threshold





- Click on LR limit (1). The "Simulation limits" window opens.
- Select your LR threshold (2), 1000 for instance, and click "Update" (3).
- You will obtain the info in 4 .

Meaning
If your LR threshold = 1000, no false positives are expected if you analyze 16 markers. False negatives are very low (since 99.8% of simulations gave LRs ≥ 1000)

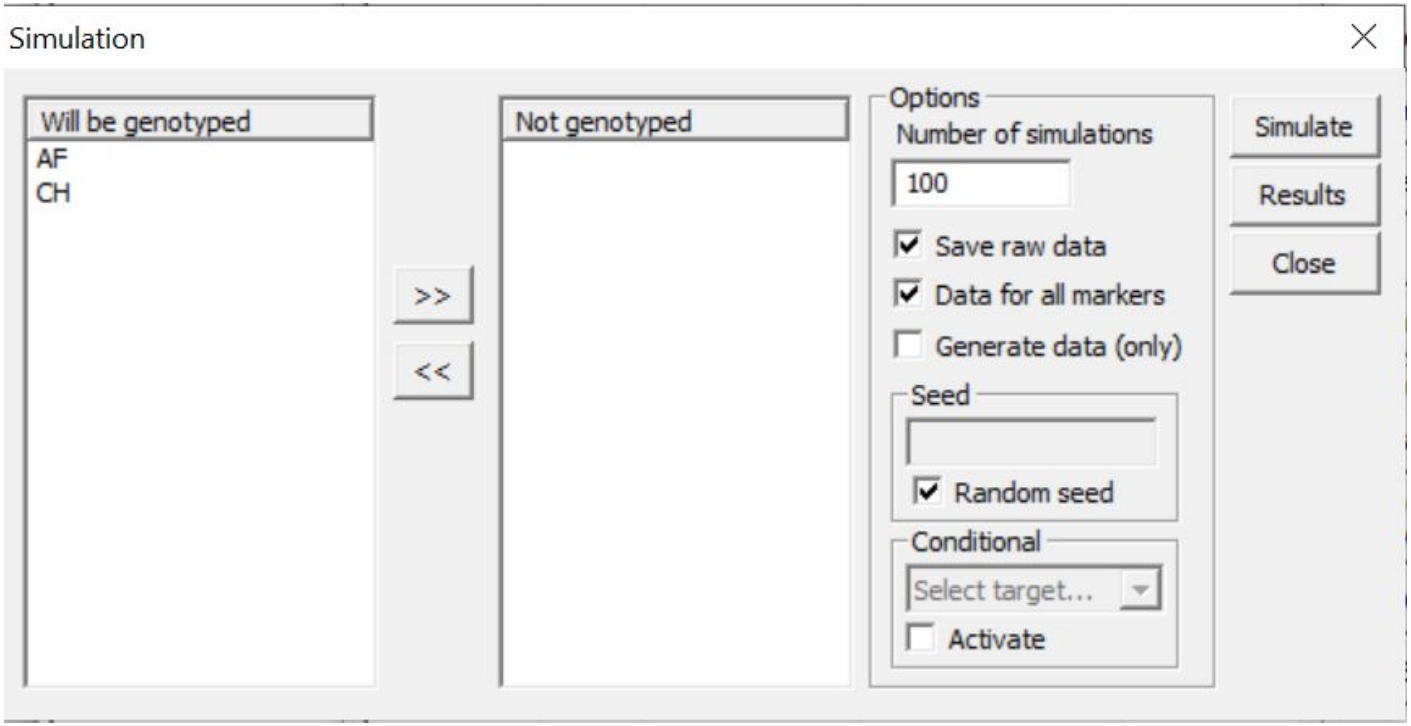
False positive means LR ≥ 1000 if H2 is true
False negative means LR ≤ 1000 if H1 is true

Simulations - Further

- We recommend at least 1000 simulation to obtain an idea of the LR distribution. However, to explore more extreme possibilities, a much greater number of simulations should be performed, for instance 100,000.
- We recommend to use random seed in the general simulation scenario whereas a fixed seed (say 12345) should be used when results need to be reproducible, for instance in connection with a publication.
- Hint: The raw exported genotype data may be used as input in Familias. For instance, if simulating data for a DVI test scenario (see Section 6), data for families may be simulated and used as a starting point.
- For plotting purposes (or other uses) the raw LRs from the simulation can be exported from the Results window. We recommend to use the logarithm of the LR as the resulting distribution is approximately Normal. See next slides for some examples.

Plotting simulations I. Commands in Familias

- Download the file in
  http://familias.name/tutorial/Basics-sim.fam
  and perform 100 simulations using the following options



Plotting simulations II. Excel

- The output file can be explored in Excel (and plotted, if you are an expert in Excel)

Simulation raw data, number of simulations: 100

#Simu	Marker	TRUE PED	logLikelihc	logLikelihc	AF	Allele 1	Allele2	CH	Allele 1	Allele2
1	D3S1358	H1:Father	-2.16555	-2.63434		17	99		17	99
1	D3S1358	H2: Unrela	-3.94669	-3.01511		99	18		99	17
2	D3S1358	H1:Father	-2.63076	-3.39588		99	18		99	18
2	D3S1358	H2: Unrela	-3.94669	-3.01511		18	99		99	17

- Data may be arranged in Excel (the Data > filter option may be useful)

Plotting simulations III. Data organised for R.

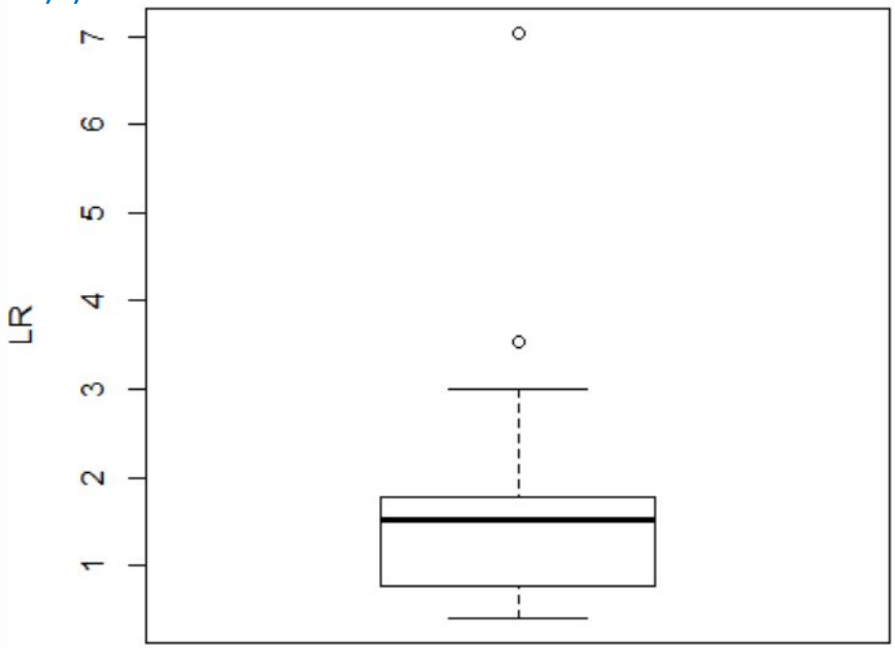
1	Sim	Marker	TRUEPED	logH1	logH2
2	1	D3S1358	H1	-2.17	-2.63
3	1	D3S1358	H2	-3.95	-3.02
4	2	D3S1358	H1	-2.63	-3.4
5	2	D3S1358	H2	-3.95	-3.02

- Plotting and further analysis can be done in Excel (not explained here).
- We rather use R and save the above file as a tab delimited file called Sim.txt
  - Variable names need to comply with R requirements
    (for instance, 'TRUEPED', not 'TRUE PED')
  - Value of variables also need to comply with R requirements:
    We can use e.g. 'H1.father' but not say 'H1:father' or 'H1 Father'.

Plotting simulations IV. Using R

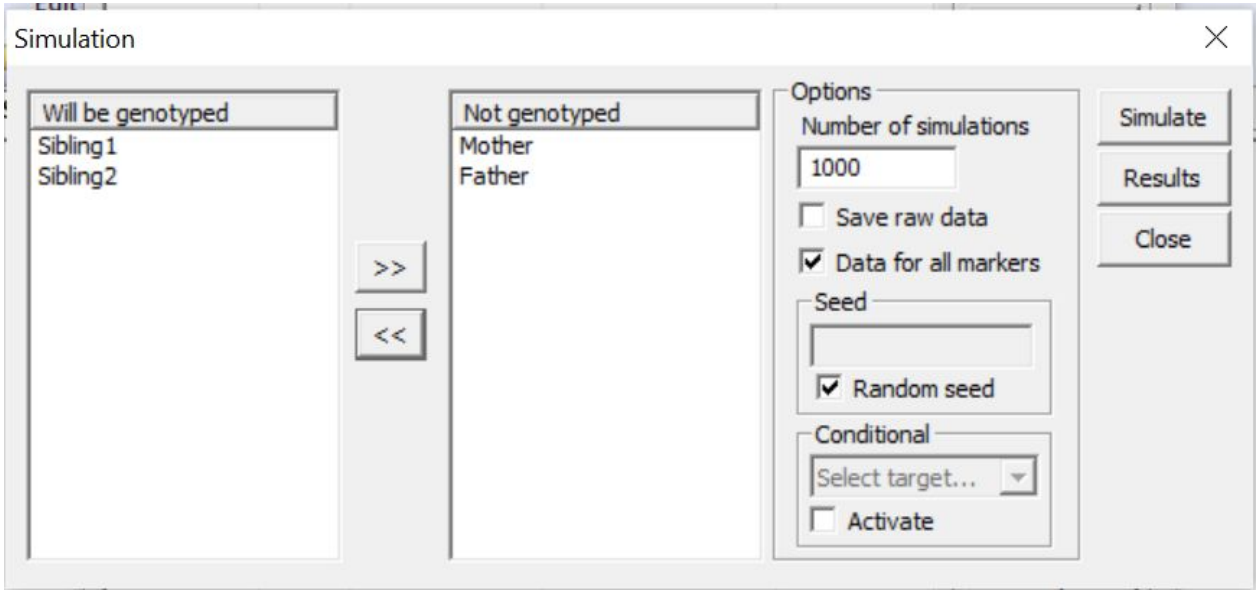
- We provide an example for plotting in R, with the file already in it's correct format.
- You can copy and paste the following commands in R, but internet connection is required since the file of interest is in the link of the first command line:

```
dat = read.table("http://familias.name/tutorial/Sim.txt", header = TRUE)
summary(dat) # Check data
dat = dat[dat$TRUEPED == "H1", ] #Extracts data for hypothesis H1
summary(dat)
logLR = dat$logH1 - dat$logH2 # Calculates log (=ln , i.e., base=e) of LR
hist(logLR, xlab = "logLR", ylab = "count",  main ="logLR (lnLR)")
boxplot(logLR, ylab ="logLR")
LR = exp(logLR)
hist(LR, xlab = "LR", ylab = "count ", main ="LR")
boxplot(LR, ylab ="LR")
```



Plotting simulations V

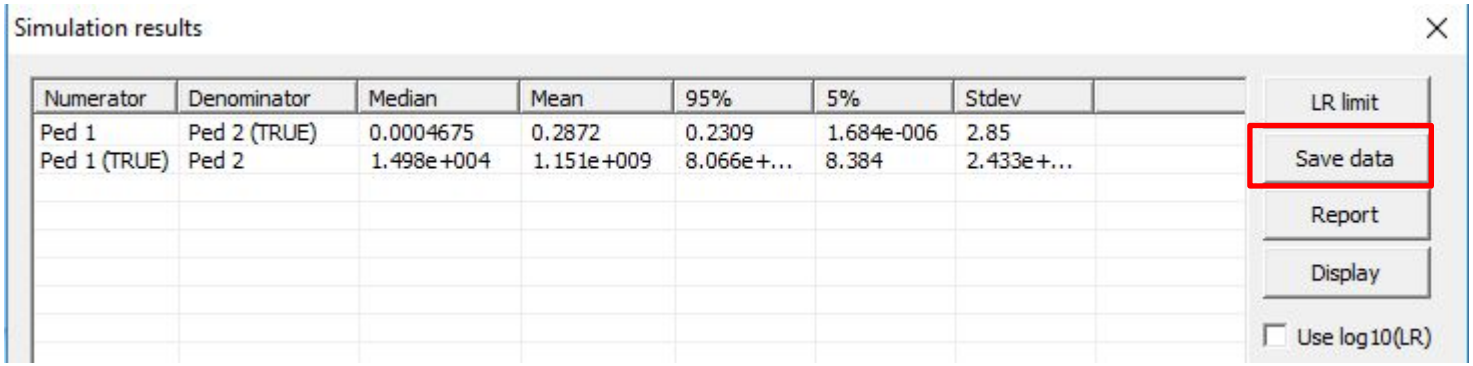
- 1000 simulations are saved based on analysis using http://familias.name/tutorial/Siblings-sim.fam



We do not tick the ”Save raw data” option in this window!

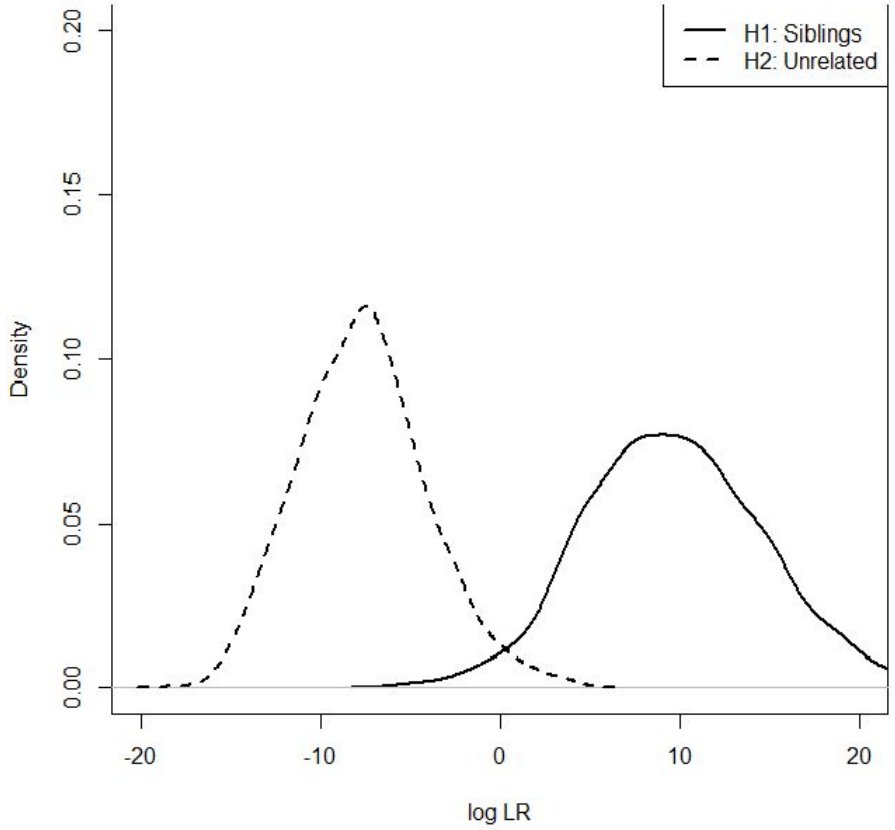
Plotting simulations VI

- Data saved to file Sim2.txt and processed as shown below




```
# First, density plot (log LR distributions), running this code requires an internet connection
dat = read.table("http://familias.name/tutorial/Sim2.txt", header = FALSE, skip=5)
plot(density(log10(dat$V1)), xlab= "log10 LR", ylab= "Density", col=1, lwd=2, main="", xlim=c(-20,20), ylim=c(0,0.2))
points(density(log10(dat$V2)), col=1, lty=2, lwd=2, type= "l")
legend("topright", c("H1: Siblings", "H2: Unrelated"), lwd=2,lty=1:2)
# Next exceedance plots, i.e. the probability to exceed each threshold given each hypothesis
dev.new()
threshold <- exp(seq(log(0.01), log(20), length.out=1000))
eh1 = eh2 = threshold
for (i in 1:1000) { eh1[i] = sum(log10(dat$V1)>threshold[i])/1000; eh2[i]= sum(log10(dat$V2)>threshold[i])/1000; }
plot(x=threshold, y=eh1, lwd=2, ylab= "Exceedance probability" ,xlab= "log10 Threshold" , type= "l")
points(x=threshold, y=eh2, lwd=2, lty=2 , type= "l")
grid(); legend("topright", c("H1: Siblings", "H2: Unrelated"), lwd=2,lty=1:2)
```

Plotting simulations VII



Meaning
After 1000 simulations:
- the values of the LRs if H2 is true range from 10^{-20} to 10^{5} (more or less)
- the values of the LRs if H1 is true range from 10^{-5} to $>10^{20}$

The probability that LR(unrelated) is larger than LR(sibs) is fortunately small, and estimated in R to be 0.002 (2 of 1000 simulations):

```
sum(dat$V2 > dat$V1)/1000
```

Data for plotting: http://familias.name/tutorial/Sim2.txt

Conditional Simulations I

- The genotypes of the typed individuals are taken into account.
- Very useful for DVI since specific reference genotypes may determine how informative a pedigree is.
- Example: the reference is the father and the child is the missing person. If the genotype of the father is rare, simulated LRs will tend to be larger than if the father's genotype is common.
- Only genotypes compatible with the father will be simulated

	Father
D3S1338	14-16

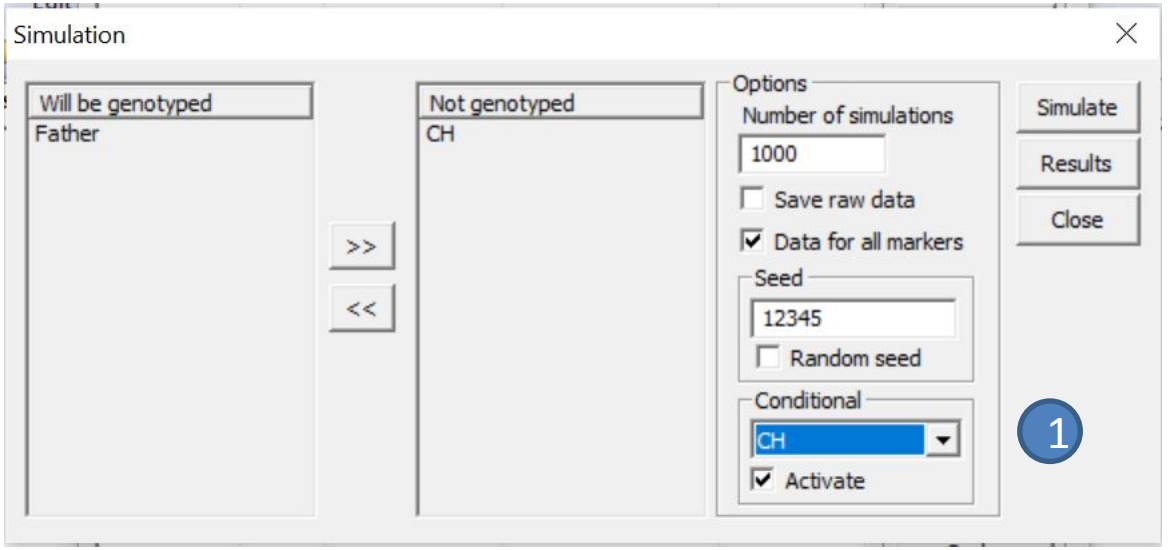
D3S1338	Simulated child
Simu1	14-14
Simu2	14-16
Simu2	16-17
…	…

Only genotypes with alleles 14, 16 or both will be simulated
The LR for each simulated genotype is then calculated

Conditional Simulations II

Use data in file http://familias.name/tutorial/cond-simu-father-child.fam

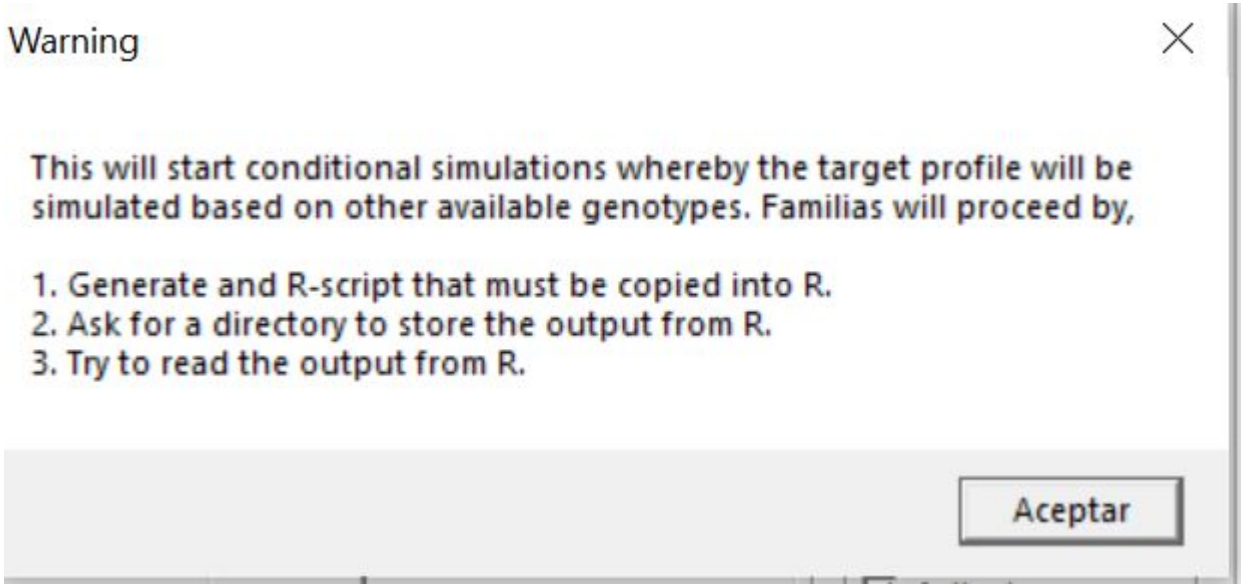
Open the Simulation window from the pedigrees window 1
Activate conditional and select the target (child CH in this case)
Press Simulate 2



Note: Familias will only simulate data for the target given the already defined genotypes! The option to include/exclude persons is unavailable (the Will be genotype/Not genotyped lists are grey)

Conditional Simulations III

An R-script will be generated.
Copy and paste the script in R

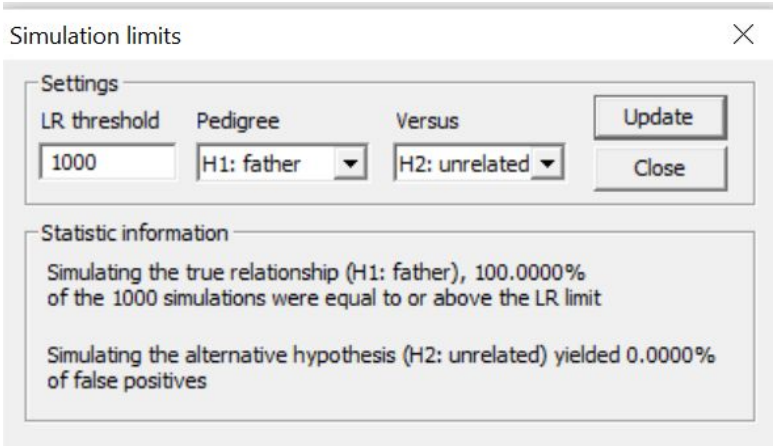


Conditional Simulations IV

The results will appear in the Familias screen

Simulation results

Numerator	Denominator	Median	Mean	95%	5%	Stdev
H1: father	H2: unrelated (TRUE)	0	0	0	0	0
H1: father (TRUE)	H2: unrelated	1.968e+008	1.4e+010	2.844e+010	1.058e+006	1.796e+...



5. Blind search

Input file: tutorial-Ch5.fam

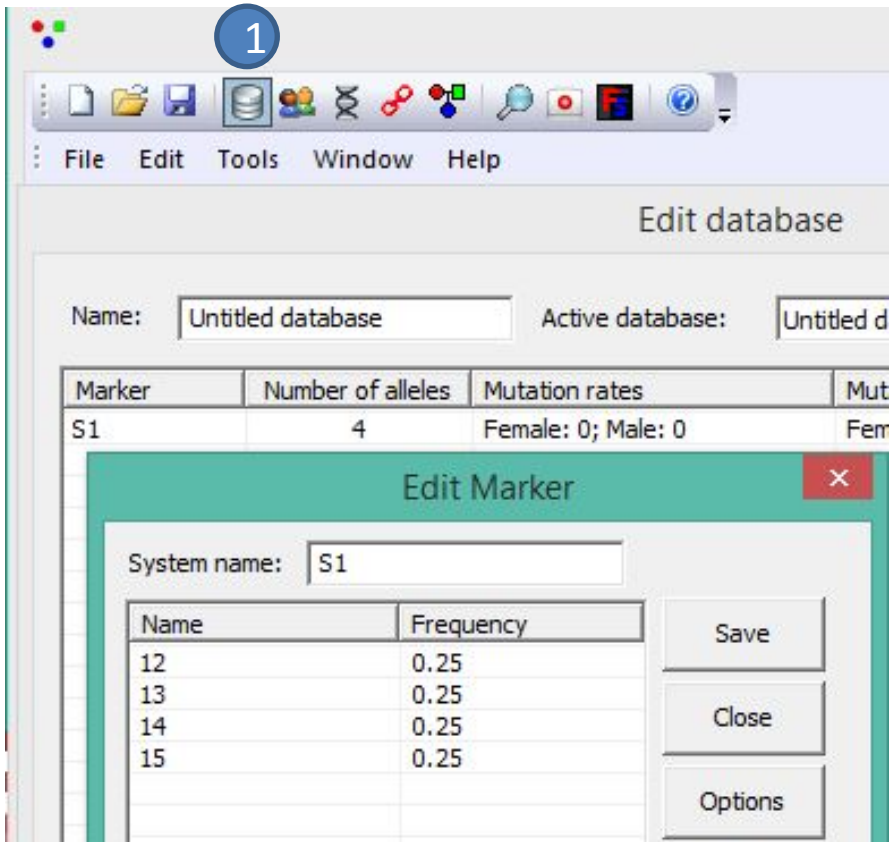
Purpose

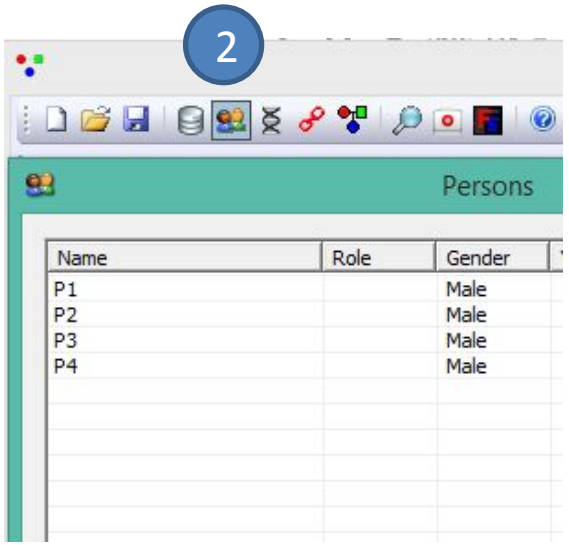
- Find the pairwise relationships between DNA profiles.
  - Simple case:

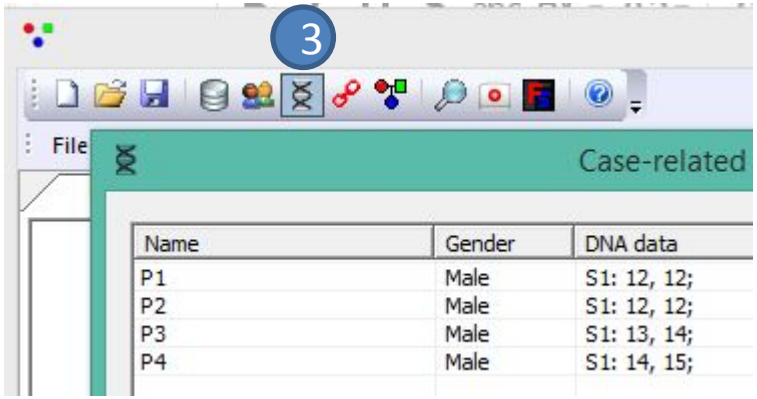
Name	Gender	DNA data
P1	Male	S1: 12, 12;
P2	Male	S1: 12, 12;
P3	Male	S1: 13, 14;
P4	Male	S1: 14, 15;

- With the above profiles, we can do the following six comparisons
  - P1-P2, P1-P3, P1-P4, P2-P3, P2-P4, P3-P4
- 'Blind' implies that some fixed relationships are searched (as opposed to explicitly specifying pedigrees) namely
  - Parent-Child, Siblings, Half-siblings, Direct-match, Cousins, 2nd cousins
- We could e.g. guess that P1-P2 is most likely a 'Direct-match' (from the same person). However, other relationships are also possible, but less likely.

How to do Blind Search - Preliminaries

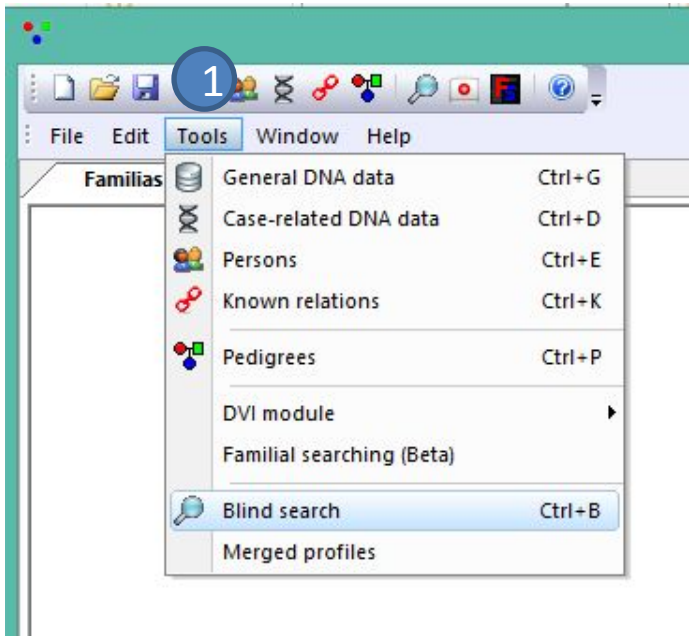


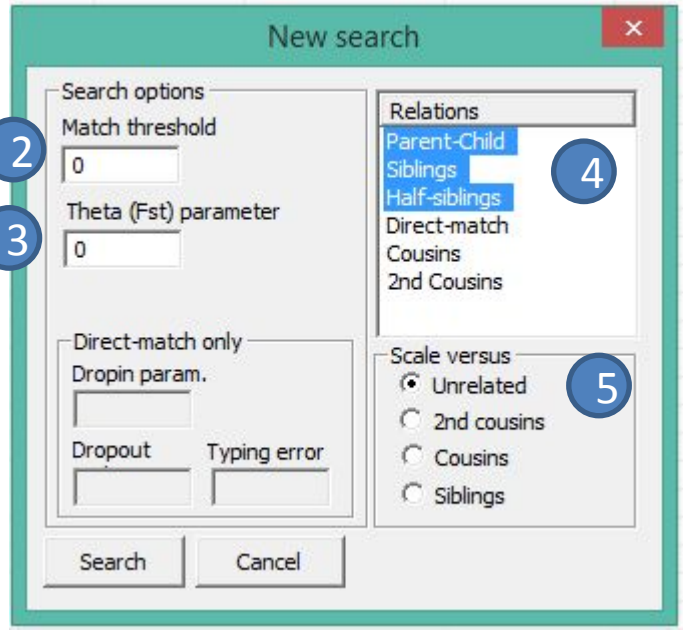




- Define General DNA data 1, Persons 2 and Case DNA data 3

The search





- The search is initiated 1 (alternatively Blind search can be done in the DVI and Familial Searching modules; then only step 1 of previous page is done before entering these modules)

- Parameter settings:
  - ✔ LR exceeding 0 will be reported (default is 1) 2
  - ✔ Theta, default 0 is used 3
  - ✔ Desired comparisons 4. Here: Parent-Child, Siblings, Half-siblings
  - ✔ Compare against unrelated (denominator of LR) 5

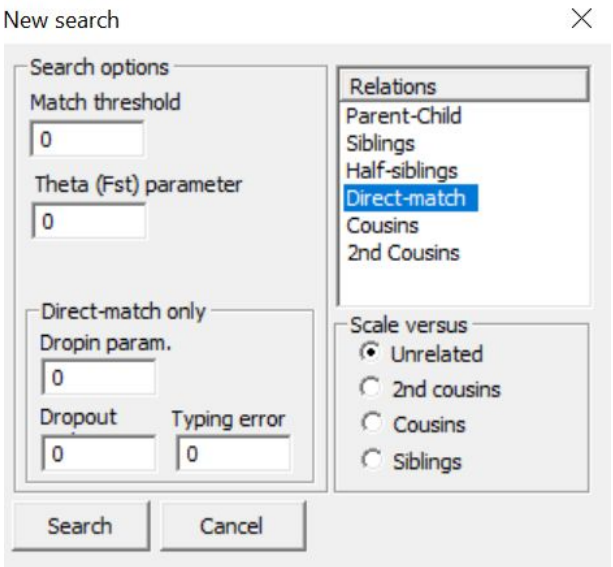
Basic output

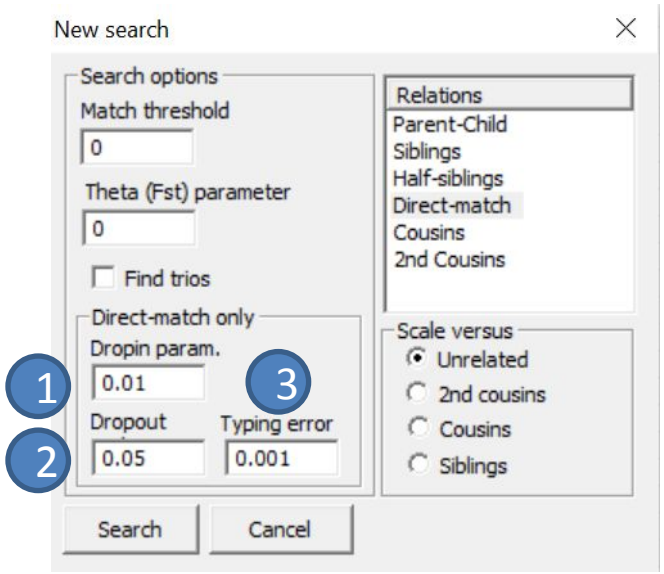
Person 1	Person 2	Gender ma...	Relationship	LR	Inconsistencies
P1	P2	-	Siblings	6.25	NA
P1	P2	-	Parent-Child	4	0
P1	P2	-	Half-siblings	2.5	NA
P3	P4	-	Half-siblings	1	NA
P3	P4	-	Parent-Child	1	0
P3	P4	-	Siblings	0.75	NA
P2	P4	-	Half-siblings	0.5	NA
P2	P3	-	Half-siblings	0.5	NA
P1	P4	-	Half-siblings	0.5	NA
P1	P3	-	Half-siblings	0.5	NA
P2	P4	-	Siblings	0.25	NA
P2	P3	-	Siblings	0.25	NA
P1	P4	-	Siblings	0.25	NA
P1	P3	-	Siblings	0.25	NA
P2	P4	-	Parent-Child	0	1
P2	P3	-	Parent-Child	0	1
P1	P4	-	Parent-Child	0	1
P1	P3	-	Parent-Child	0	1

- Three relationships are studied, and there are six pairs for each, so the list contains 3*6=18 comparisons.
- The list is sorted according to LR.
- Inconsistencies are only possible for ‘Parent-Child’ and these are indicated
- Mutations are accounted for in the Blind search in the DVI module only for ‘Parent-Child’

Direct match

- For Direct match, additional parameters can be set
  - Dropin param. (1), Dropout (2) and Typing error (3). Output below for parameters equal 0.
  - The P1- P2 LR is now reduced to 9.6 while the other comparisons now give LR > 0.





Person 1	Perso...	Gender ma...	Relationship	LR	Inconsist...
P1	P2	Yes	Direct-match	16	0
P3	P4	Yes	Direct-match	0	1
P2	P4	Yes	Direct-match	0	1
P2	P3	Yes	Direct-match	0	1
P1	P4	Yes	Direct-match	0	1
P1	P3	Yes	Direct-match	0	1

Person 1	Perso...	Gender ma...	Relationship	LR	Inconsist...
P1	P2	Yes	Direct-match	16	0
P3	P4	Yes	Direct-match	0	1
P2	P4	Yes	Direct-match	0	1
P2	P3	Yes	Direct-match	0	1
P1	P4	Yes	Direct-match	0	1
P1	P3	Yes	Direct-match	0	1

Advanced options and output

- There are several further options, most of them are intuitive and not needed for simple applications.
- There is also more advanced output like information on IBS (Identical By State) alleles and IBD (Identical By Descent) alleles. This output is not relevant for standard applications.

6. DVI (Disaster victim identification)

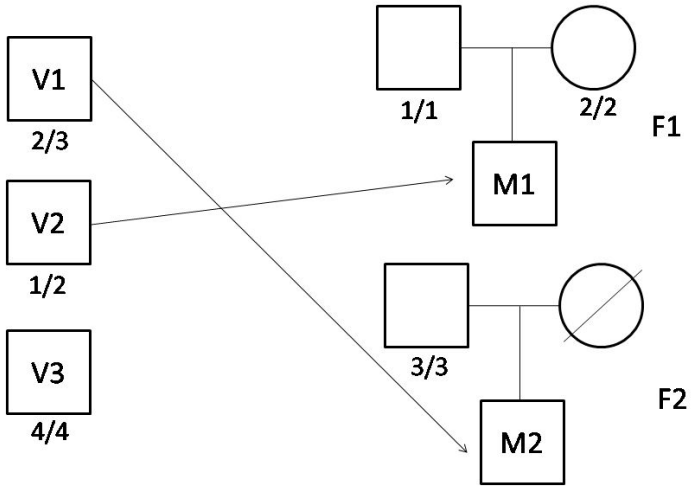
Some more advanced features of DVI, including simulation and working with separate databases of allele frequencies, are explained in here.

Input file for most examples that follow: Exercise.3.1.fam. Some screen shots and output differ slightly (insignificantly) from most recent version of Familias.

Basic steps

1. **General DNA data**. Input of database, i.e., allele frequencies, etc. Described in detail in Section 1.
2. **Unidentified samples**. Samples to identify, in a DVI operation this is a set of unidentified remains whereas in some missing person scenarios it may also be identified individuals where the aim is to reunite with biological family.
3. **Reference families**. Reference data used to identify and reunite families with missing persons.
4. **Search**. Perform and interpret the results.

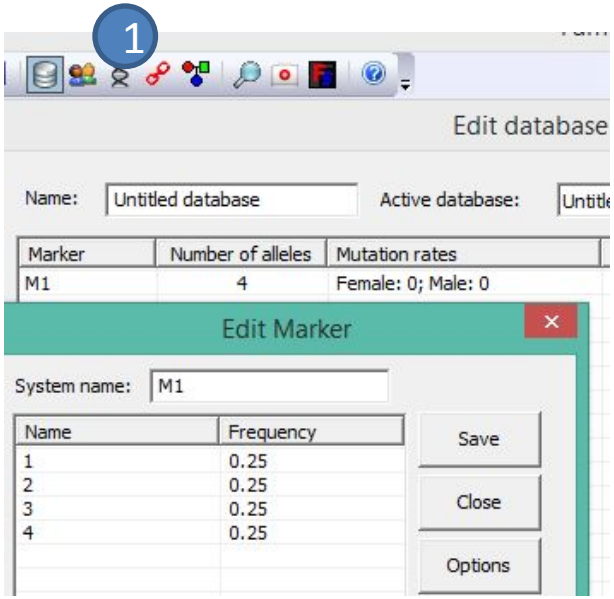
DVI illustrated



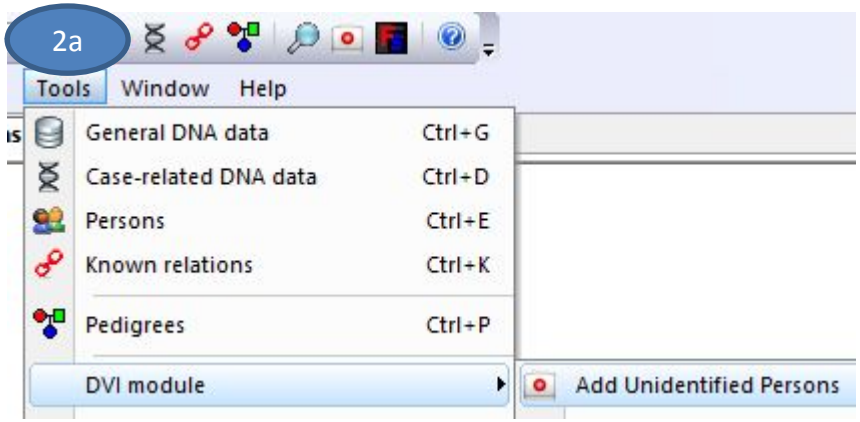
1. **General DNA data:**
   A marker with alleles 1 ,2, 3, 4 (all with freq = 0.25)
2. **Unidentified samples:**
   Three samples V1, V2 and V3 and their genotypes.
3. **Reference families:**
   Two reference families F1, F2 with missing individuals (sons here) M1 and M2.
   Marker data for available relatives (mother and father for F1, only father for F2.)

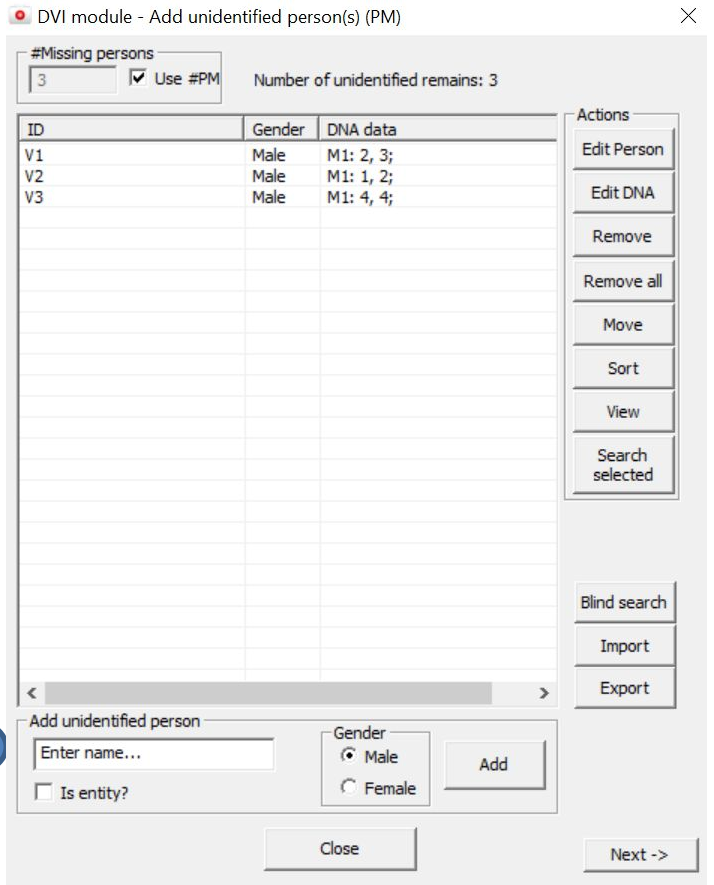
- We next illustrate manual input for this toy example with manual input; this is a useful exercise.
- For real cases, data will be imported from files and this is explained at the end of this section.

How to do DVI



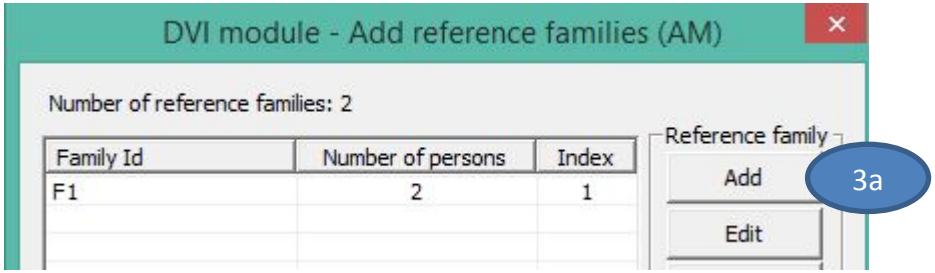
Comment
The "Is entity?" option is in Beta testing and allows the user to specify a family/pedigree instead of an unidentified person

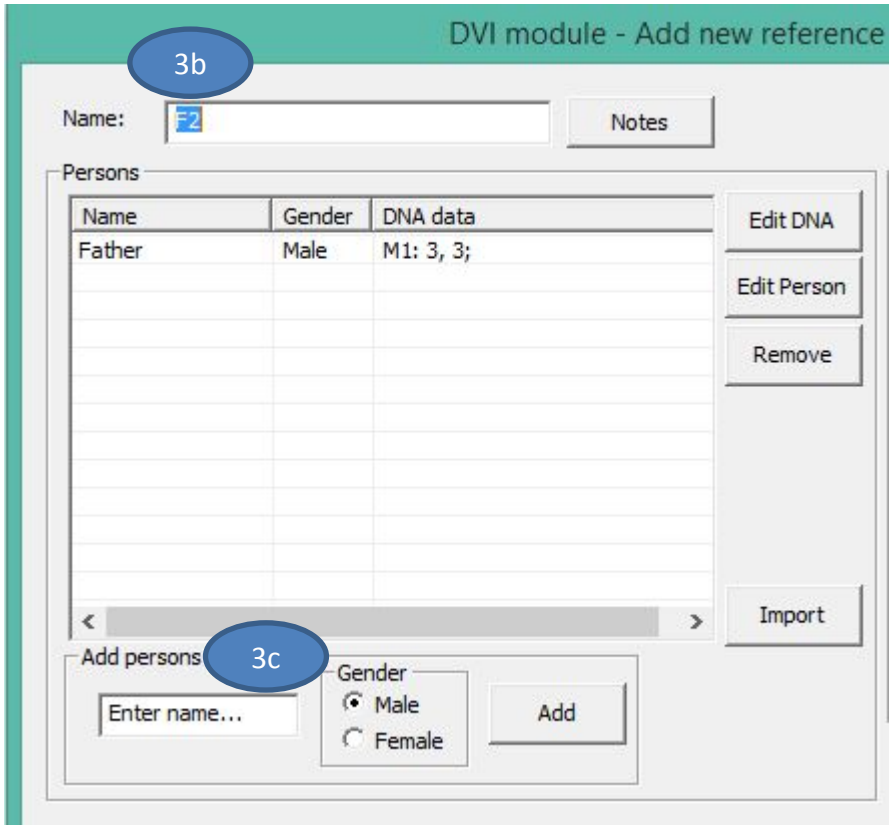




Define marker 1, enter DVI module > Add 2a and input unidentified samples 2b. Press Next 2c

Defining reference families



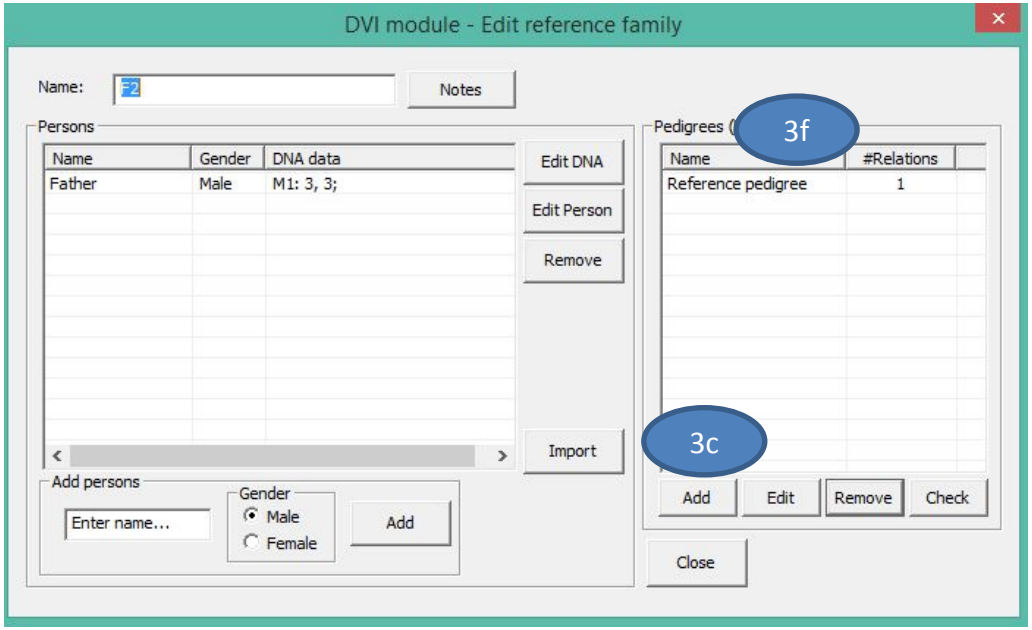


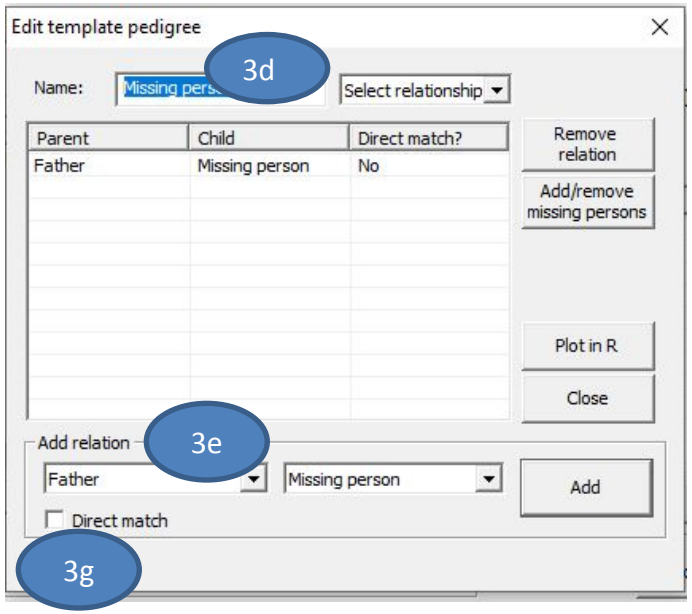
Comment
In these windows you have to define the name of the family, the members and the DNA data (relationships between them and the missing person are not defined here)

- The first family has been defined, we press 'Add' 3a to enter a new family.

- We name the family 3b
- We enter the persons needed to define the pedigree and the genotypes if available 3c

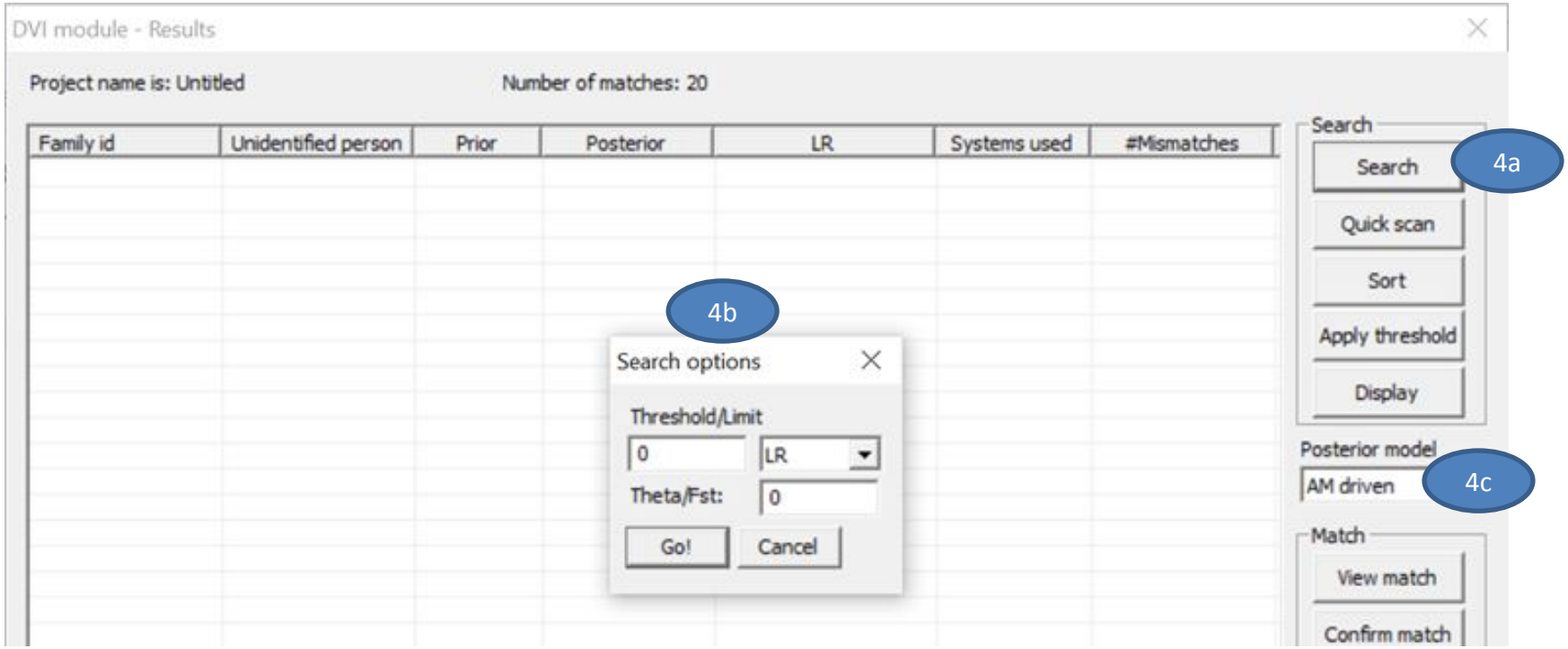
Defining pedigrees, i.e., how MP is related





- We go Add 3c to define the pedigree
- We name the pedigree 3d
- We define how MP is related with this family 3e
- We don't touch the 'Reference pedigree', 3f, it's there since a MP may not belong to any reference family.
- When input is finished, we press 'Next' to do the search ….
- Note: In order to define a direct reference samples (e.g. personal belonging), press Direct match 3g when adding relation

The DVI search



- We press search 4a
- Choose options 4b Here Threshold/limit is 0, so we see all results
- Posterior model: 4c
  - AM driven: each family treated separately
  - PM driven: each person treated separately
  - One-to-one: only two hypotheses

see next slide

- Note: priors and posteriors are not applicable when the Quick scan is performed but will still be displayed.

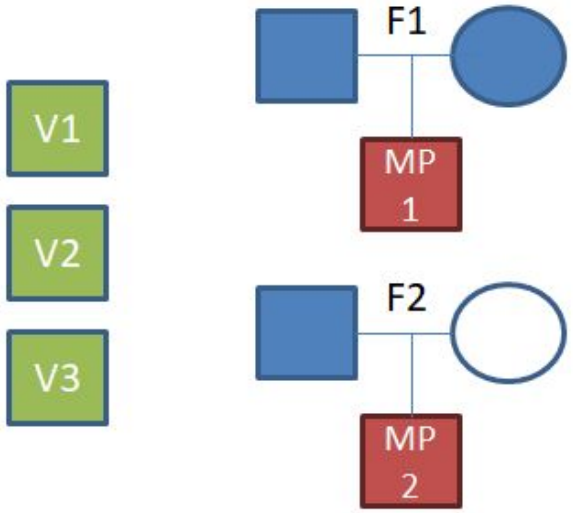
Posterior models in the example

AM driven

H_1 = family F1 is related to V1
H_2 = family F1 is related to V2
H_3 = family F1 is related to V3
H_4 = family F1 is not related to any V
H_5 = family F2 is related to V1
H_6 = family F2 is related to V2
H_7 = family F2 is related to V3
H_8 = family F2 is not related to any V

PM driven

H_1 = V1 is related to family F1
H_2 = V1 is related to family F2
H_3 = V1 is not related to any family
H_4 = V2 is related to family F1
H_5 = V2 is related to family F2
H_6 = V2 is not related to any family
H_7 = V3 is related to family F1
H_8 = V3 is related to family F2
H_9 = V3 is not related to any family



One to one

H_1 = Vn is related to Fm
H_2 = Vn is not related to Fm
Vn = any victim
Fm = any family

Output for AM driven posterior model

DVI module - Results

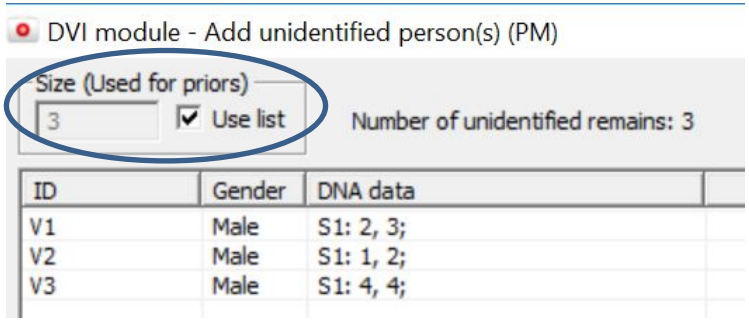
Project name is: Untitled

Family id	Unidentified person	Prior	Posterior	LR	Systems used	#Mismatches	Pedigree
F1	V1	0.25	0	0	1	1	Missing son1
F1	V2	0.25	0.888889	8	1	0	Missing son1
F1	V3	0.25	0	0	1	1	Missing son1
F2	V1	0.25	0.666667	2	1	0	Missing son2
F2	V2	0.25	0	0	1	1	Missing son2
F2	V3	0.25	0	0	1	1	Missing son2

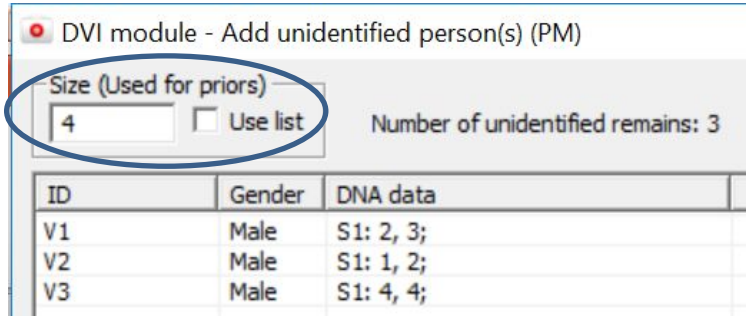
- The first family (here F1) is analysed first.
- The LR = P(marker data| MP belongs to F1)/ P(marker data| MP is not related to F1) is found for all unidentified samples.
  For V1 and V3, LR = 0, while for V2, LR = 1/(2*0.25*0.25) = 8.
- For the posterior we need a prior. Here we assume that *a priori* V1, V2 and V3 have a prior of 0.25 to belong to F1; the missing person may be someone not found with prior 0.25. The posterior probabilities are
  P(V1 from F1) = 0, P(V2 from F1) = 0.89, P(V3 from F1) = 0, P(Unknown from F1) = 0.11.
- Remaining families, here only F2, are treated similarly.

How Familias set priors

In the "Add unidentified person(s) (PM)" window



- By default (Use list), Familias takes into account the number of unidentified samples (3 in this case) and also adds the possibility that the missing person was not found: 1 / 4 = 0.25 (as was shown in previous slide). The text "Use list" is replaced by "Use #PM" in recent versions.



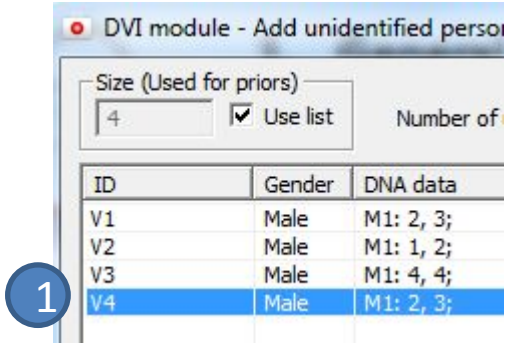
- You can also unclick "Use list" and add the size you want. For instance, if in this case we knew that at least there are 4 victims (although only 3 were found), we can add "4" in the size box. Then, prior will be 1/5= 0.20, for each victim.

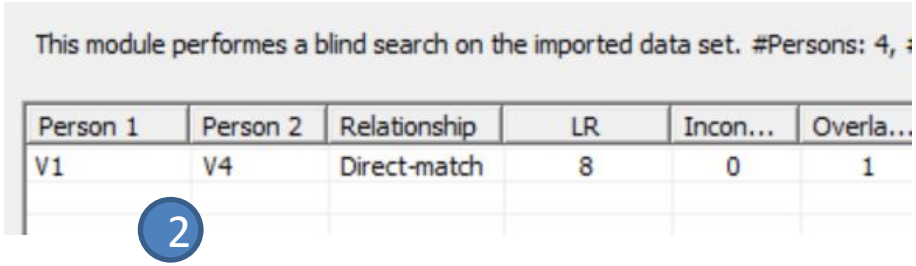
Further: unidentified persons – comparisons

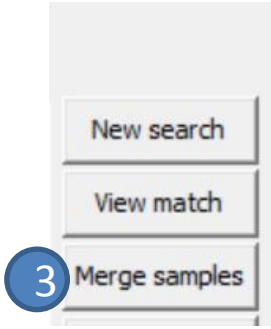
Very useful to re-associate human remains (from the same individual) and/or detect possible relationships among victims

Example:

Add a new human remain V4 with the same profile as V1 (M1 = 2/3) in the "Add unidentified persons" window 1 and click "Blind search". Press "New search" and select the "Direct match" relation vs. "unrelated" to obtain 2 . Select the samples and Click "Merge samples" 3



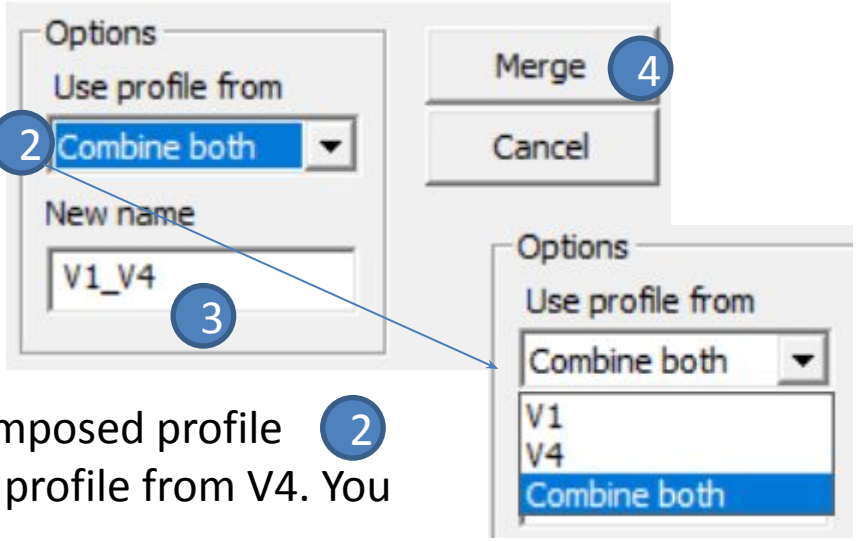




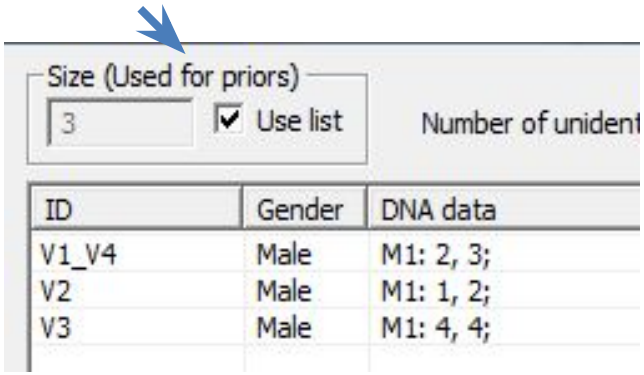
Merging samples

1

Marker	LR	V1	V4	Merged profile
M1	8	2, 3	2, 3	2, 3

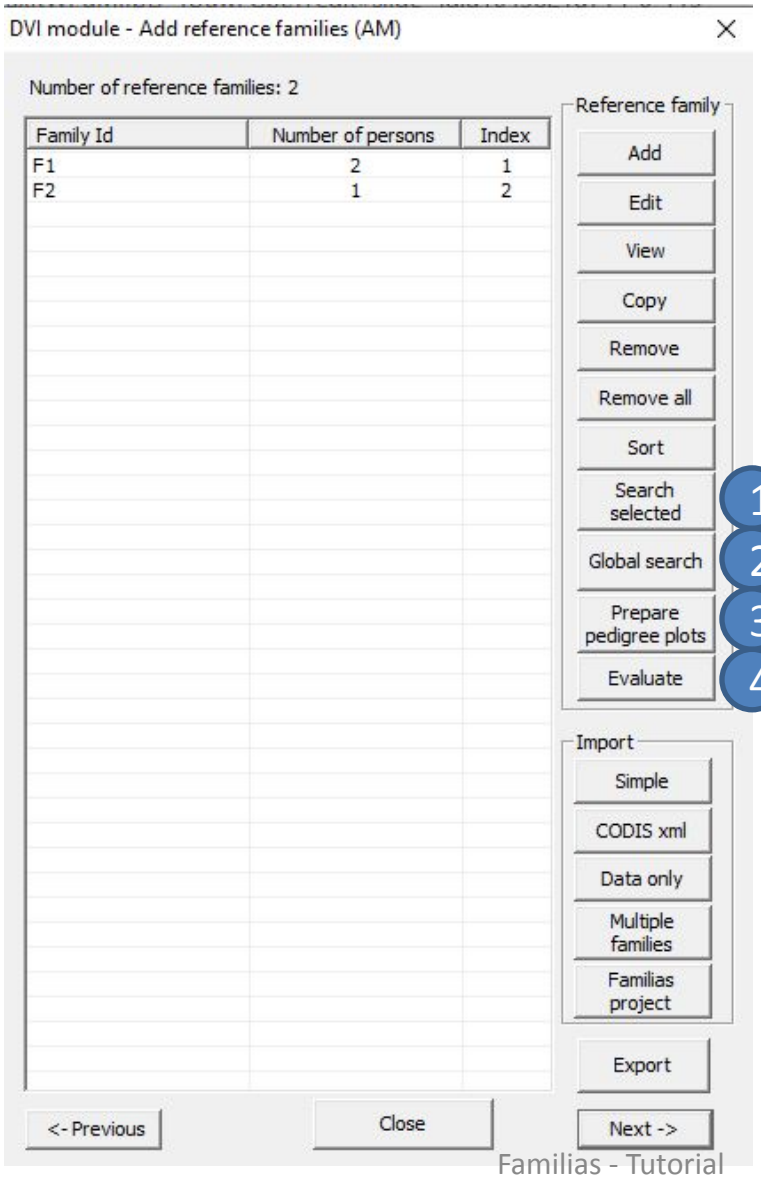


You will obtain a summary of the samples. 1
In the "Options" menu you can select: create a composed profile 2 (combine both option), use profile from V1 or use profile from V4. You can also rename this combined profile 3.
Click "Merge" 4 and the composed sample V1_V4 will appear now in the "Unidentified persons" window



Comment
In this example V1 and V4 have the same profile, but in real cases it is possible that one of the samples is incomplete. This is the reason why the option "Combine both" is available.

Further - Reference families



1 You can search only the selected families

2 Perform global search (not yet available)

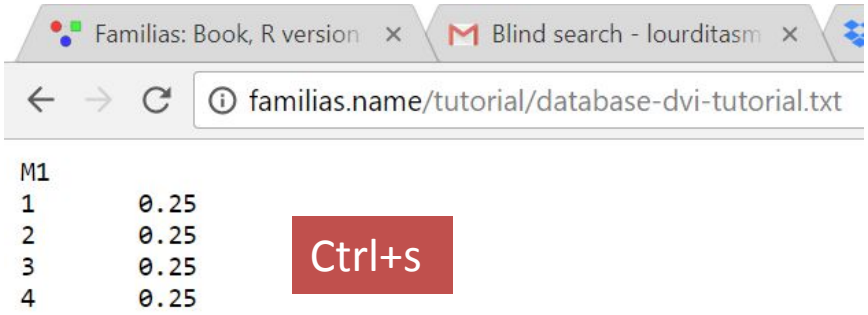
3 Prepare plots in R for selected families

4 Evaluate the reference families, simulations as well as other summary statistics

DVI with import from files

- Start Familias and import the frequency database:
  - **Tools > General DNA data** file: http://familias.name/tutorial/database-dvi-tutorial.txt
- Read in the unidentified samples:
  - **Tools > DVI module > Add Unide…** file: http://familias.name/tutorial/unidentified-dvi-tutorial.txt
- Option 1: Read reference persons with marker data and enter pedigrees manually
  - **Tools > DVI module > Add refe … Data only**
    file: http://familias.name/tutorial/ReferenceFamilies-dvi-tutorial.txt
  - Define pedigrees manually
- Option 2: Read all input for reference families
  - **Tools > DVI module > Add refe … Multiple families**
    file: http://familias.name/tutorial/pedigrees-dvi-tutorial.txt
- The search can be done next giving results as before. Further details on import files are in Section 8 below.

Note: you can copy and paste the links in your browser and then save the files by using Ctrl+s keys of your keyboard

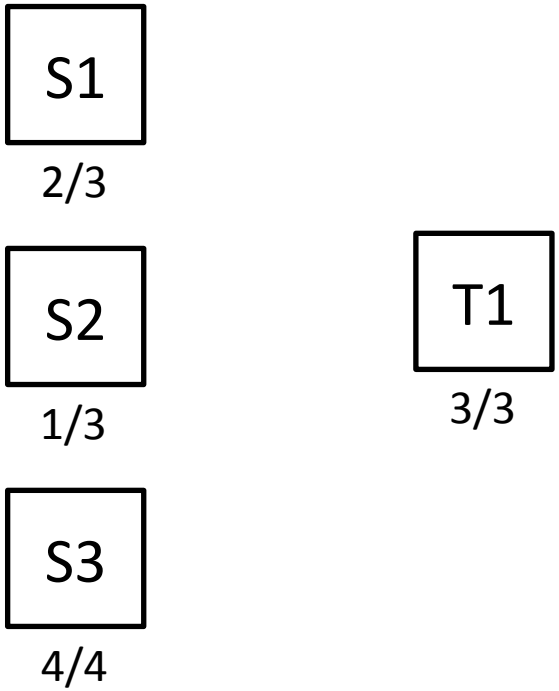


7. Familial searching

Basic steps

1. **General DNA data**. Input of database, i.e., allele frequencies, etc.
2. **Profile database**. Individuals/traces in the offender database.
3. **Traces**. Traces to perform the familial search on
4. **Search**. Perform and interpret the results from a familial search

Familial searching illustrated



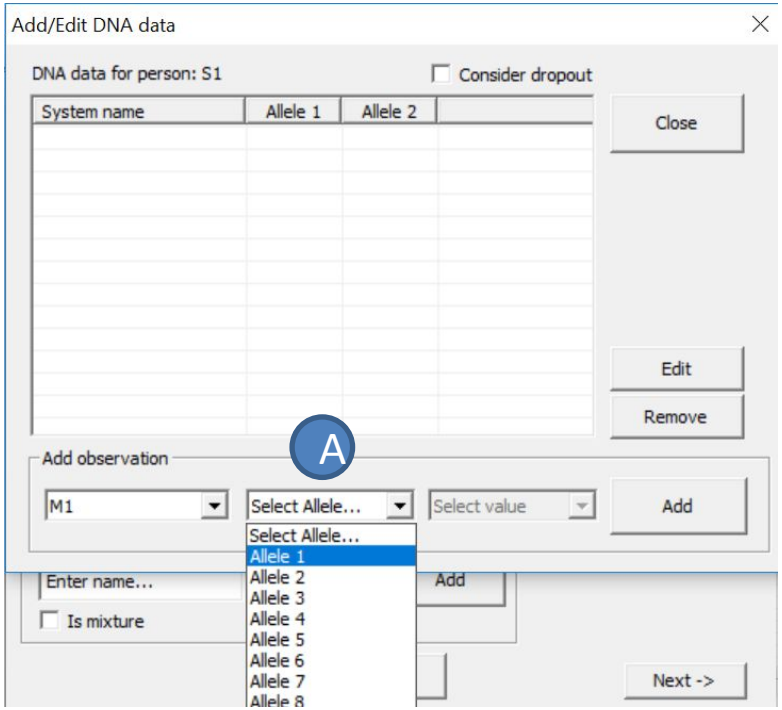
1. **General DNA data:**
   A marker with alleles 1 ,2, 3, 4 (all with Freq.=0.25).
2. **Offender database:**
   Three samples S1, S2 and S3 and their genotypes.
3. **Target profile:**
   A trace T1 with a genotype that does not fully match any of the offenders'.

- We next illustrate manual input for this toy example with manual input; this is a useful exercise.
- For real cases, data will be imported from files and this is explained thereafter. Familias can handle databases of > 300,000 samples.

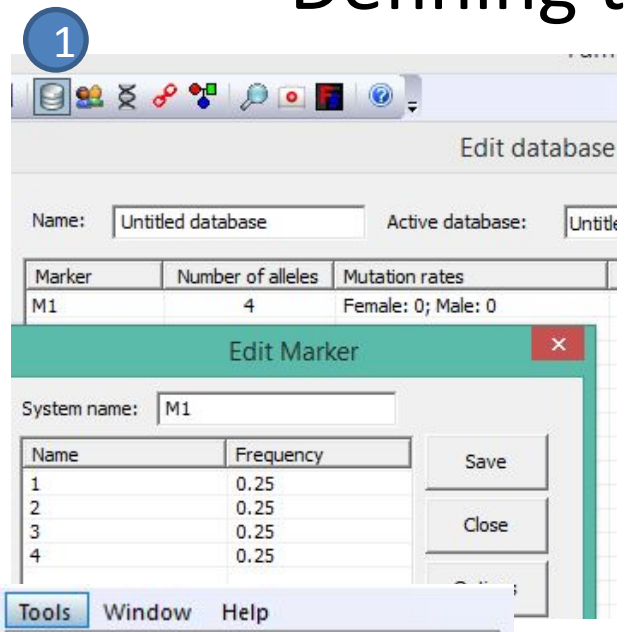
How to enter profiles in this module

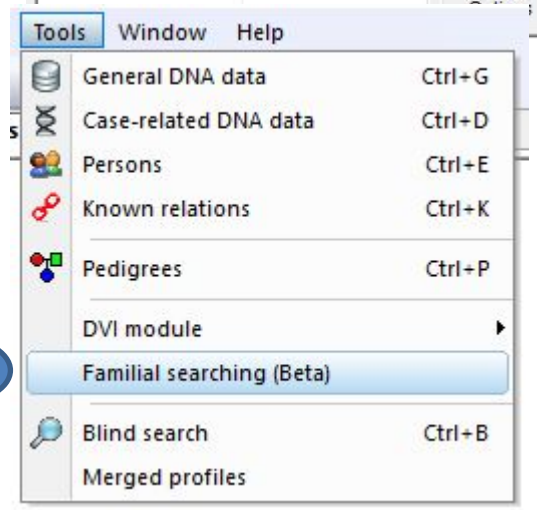
Select the first offender/trace and click «Edit DNA» to enter the profile
In Familial Searching module, **profiles are entered in a different way** (since you may need to add more than two alleles for DNA mixtures)
To enter S1=2/3, you have to:

- Select «Allele 1» in the dropdown menu (A) and then select the value (B) of this first allele (2, in this example). **Do not press Add just yet**
- Select «Allele 2» in the dropdown menu and then select the value of this second allele (3, in this example). Now, press Add to obtain (C).

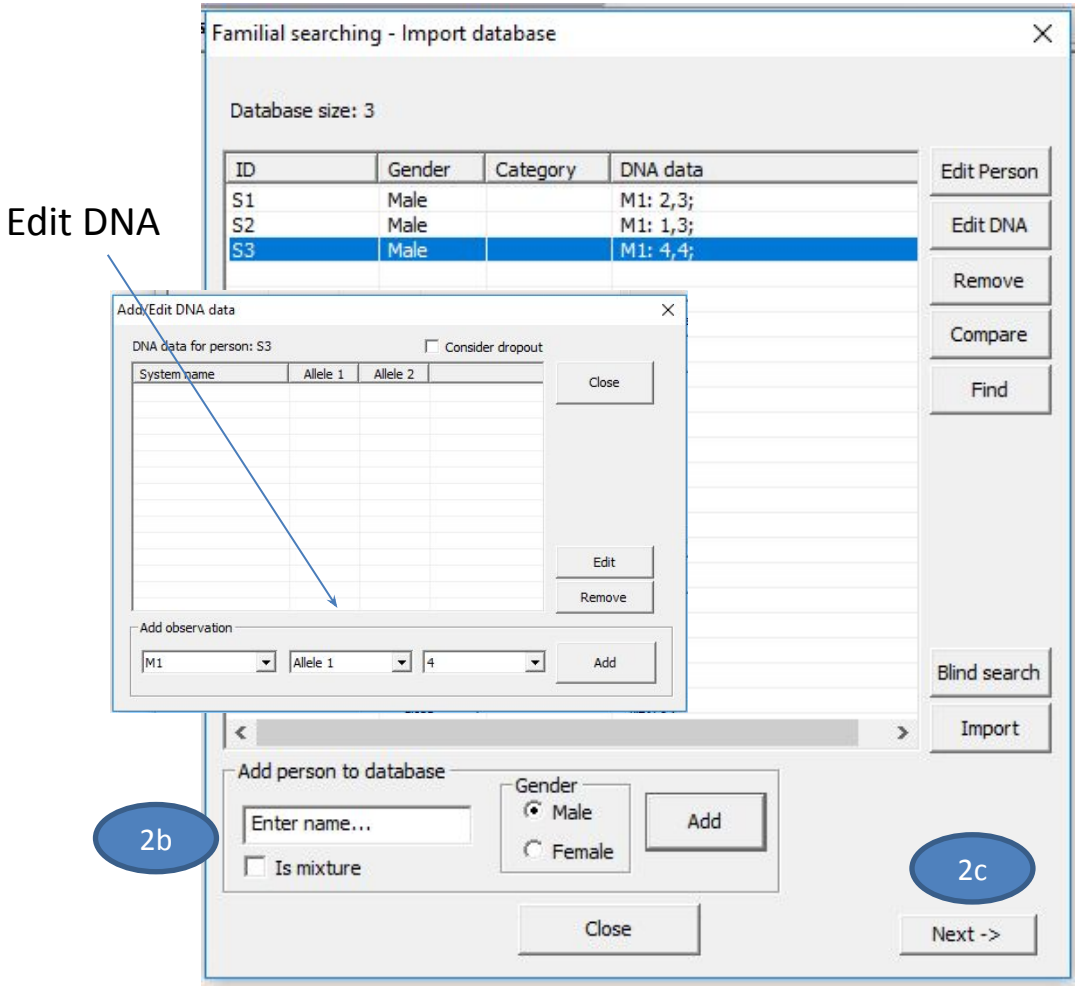


Defining the offender database



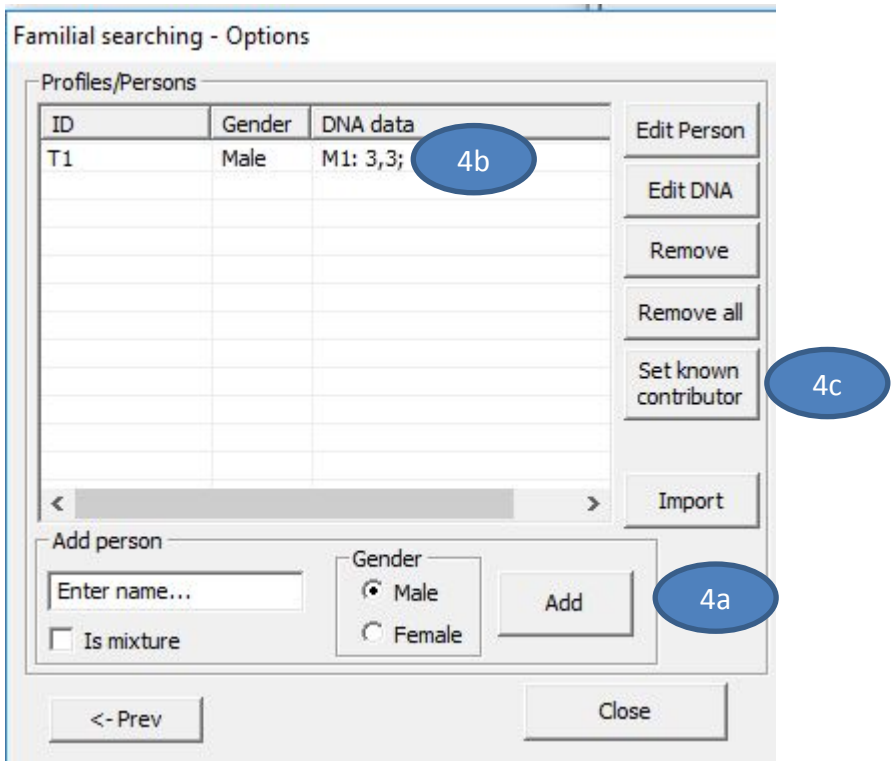


Edit DNA



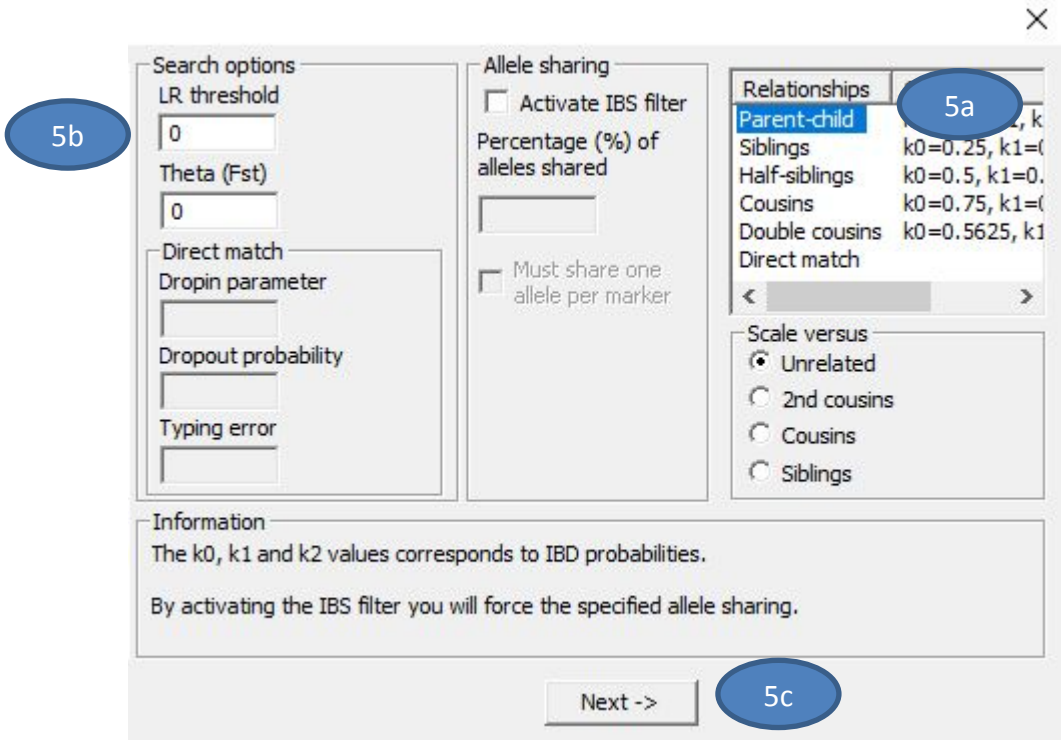
Define marker 1, enter Familial searching module 2a
Add persons and database profiles 2b Press Next 2c

Defining target(s)



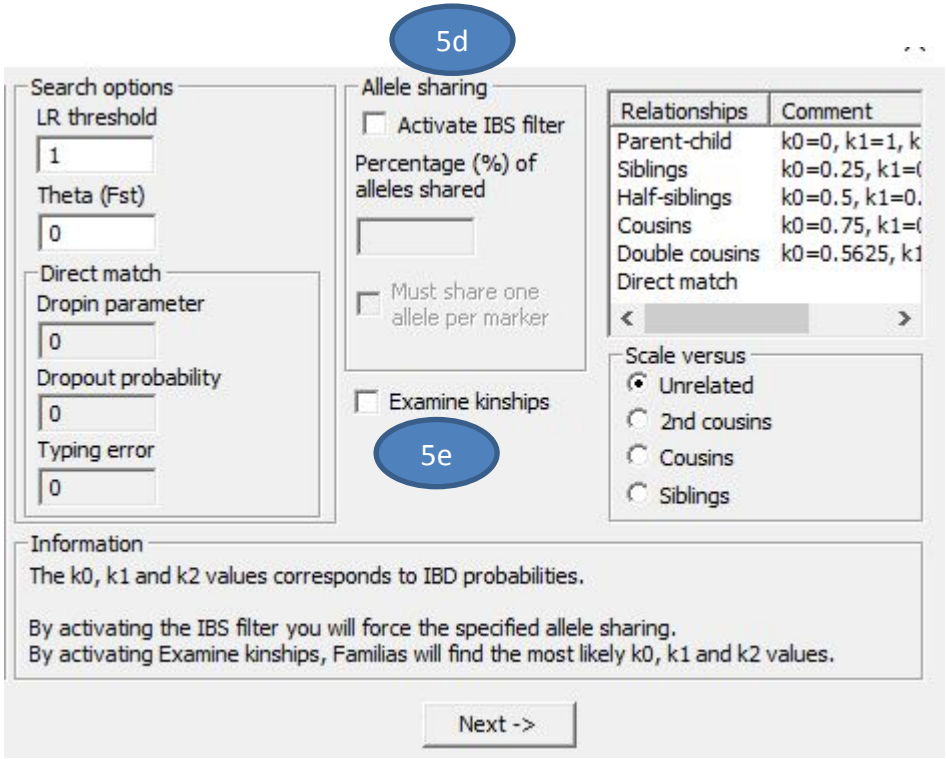
- Define the target, in this case a trace. Press 'Add' 4a
- Edit DNA data as indicated 4b
- We may set known contributors (e.g. Victims in mixtures) 4c

Defining search options



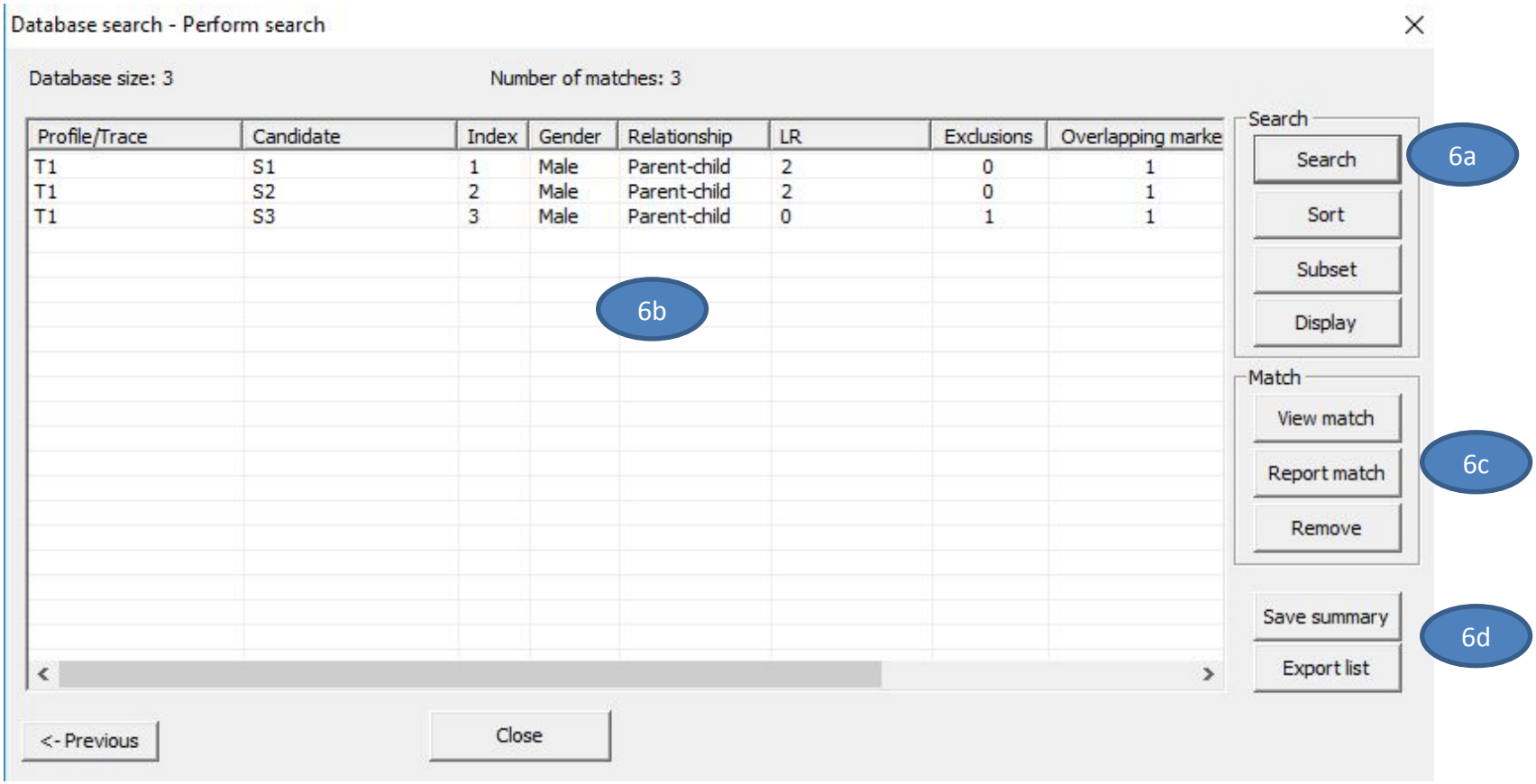
- We specify Parent-child (5a) as the relationship we want to include in the search.
- We set the LR threshold to 0 (zero) (5b) and leave the other parameters at their default values.
- Press 'Next' (5c)

Advanced options



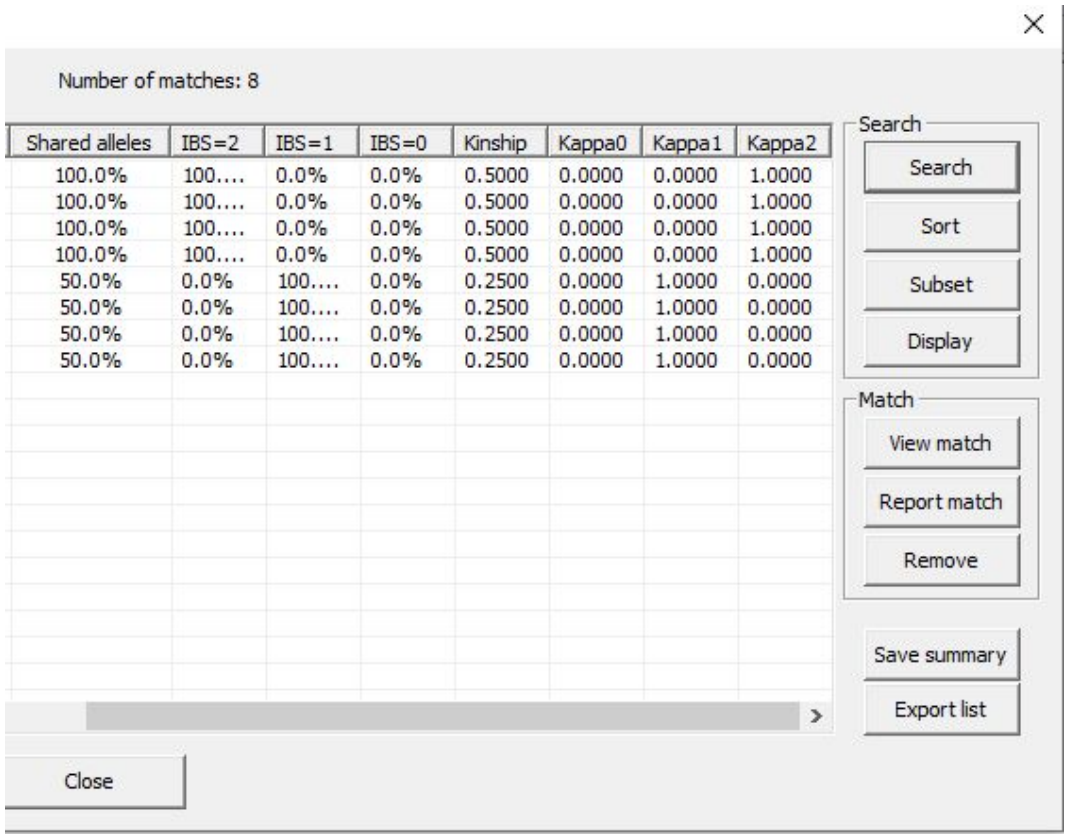
- We can use an IBS filter 5d to remove matches with low IBS sharing
- We can select Examine kinships 5e to have Familias perform a grid search over a small space of possible k0, k1 and k2 values. Computer intense and only performed for relationships exceeding the LR threshold.

The familial search - Basic results



- We press 'Search' 6a
- The results are displayed as matches between a Profile/Trace (target) and a Candidate (database element). Results are sorted according to the LR value. In addition some other metrics are displayed. 6b
- The results may be investigated closer 6c and some reports may be created 6d

The familial search - Advanced results



- By scrolling to the right you will see some other stats from the search.
- Shared alleles (total) as well as fraction of markers with 0, 1 or 2 alleles IBS
- If Examine kinships is activated in the search options, a grid search is performed where the most likely IBD values are reported as well as the kinship coefficient.

Familial search with import from files

- Start familias and import the frequency database:
  - **Tools > General DNA data** file: http://familias.name/tutorial/database-dvi-tutorial.txt
- Import the offender database
  - **Tools > Familial searching > Import** file: http://familias.name/tutorial/offenders-fs-tutorial.txt
- Import the target profile
  - **Tools > Familial searching > Next > Import**
    file: http://familias.name/tutorial/target-fs-tutorial.txt
- The search can be done next giving results as before. Further details on import files are in Section 8 below. Familias handle standard xml output from the CODIS software.

Recall: you can save the above files by using Ctrl+S keys of your keyboard in your internet browser once you have followed the link

8. Input files (formats)

8.1 Frequency database input

Standard (tab separated)

The standard format is recognized by all versions of Familias and is described below. It is saved as a text file.

TPOX	
8	0.2
9	0.3
9.3	0.5
D12S391	
20	0.1
21	0.3
22	0.4
23	0.2

Matrix format

The matrix format is commonly attached to population frequency publications. Important is the Allele or Alleles identifier in the top left element.

Alleles	TPOX	D12S391
8	0.2	
9	0.3	
9.3	0.5	
20		0.1
21		0.3
22		0.4
23		0.2

8.2 Case data input (samples)

Tab separated 1

Tab separated files are most easily created in Excel (or a similar tool) and saved as text file with tab separation. Below is an example of the format.

Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Person 1	8	9.3	20	21
Person 2	9.3	9.3	22	23

Tab separated 2

Familias also allows tab separated text files with single column markers.
The format is described below.

Sample	TPOX	D12S391
Person 1	8,9.3	20,21
Person 2	9.3,9.3	22,23

GeneMapper

Familias can import data from the genotyper software GeneMapper (or data on similar format). The format is described exactly below.

Sample	Marker name	Allele 1	Allele 2
Person 1	TPOX	8	9.3
Person 1	D12S391	20	21
Person 2	TPOX	9.3	
Person 2	D12S391	22	23

XML input

Familias recognizes xml input based loosely on the CODIS format. An example of the format is described to the right where the necessary tags are defined. Additional tags (normally added by CODIS) are disregarded.

An example is located at
http://familias.name/Files/example.xml

```
<?xml version="1.0" encoding="utf-8"?>
<CODISImportFile xmlns="urn:CODISImportFile-schema">
  <SPECIMEN SOURCEID="N/A">
    <SPECIMENID>Person 1</SPECIMENID>
    <SPECIMENCATEGORY>Father</SPECIMENCATEGORY>
    <LOCUS>
      <LOCUSNAME>TPOX</LOCUSNAME>
      <ALLELE>
        <ALLELEVALUE>8</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>9.3</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
      <LOCUSNAME>D12S391</LOCUSNAME>
      <ALLELE>
        <ALLELEVALUE>20</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>21</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
  </SPECIMEN>
</CODISImportFile>
```

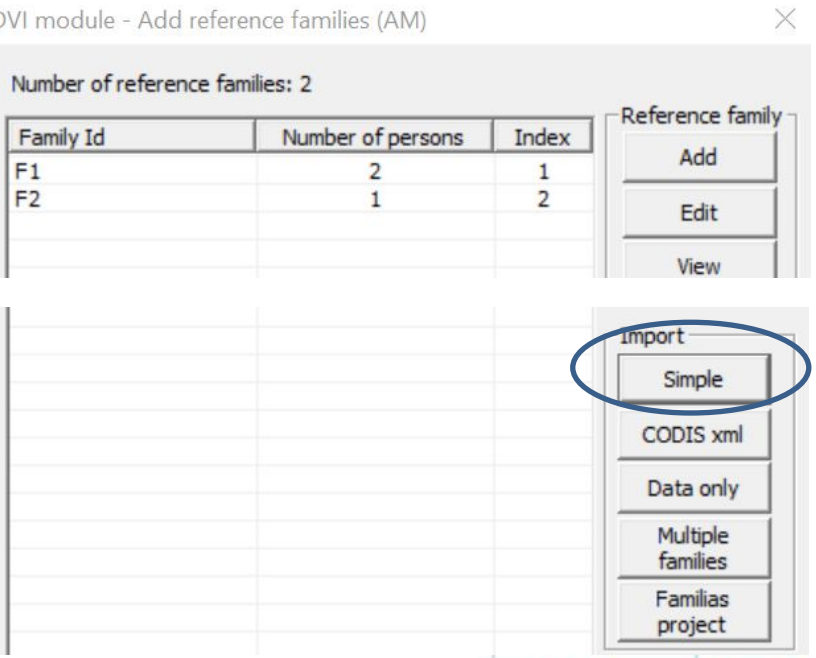
8.3 Reference family data (DVI)

Relationships

Familias recognizes some standard relationships for the subsequent import options. Below is an exhaustive list. Please check all imported data to assure Familias has made the correct interpretation. Here's an example file from the DVI section, more examples follow.

[Father], [Mother], [Parent], [Son], [Daughter], [Child], [Sister], [Brother], [Sibling], [Half Sister], [Half Brother], [Half Sibling], [Direct], [Identity], [Paternal Half Sister], [Maternal Half Sister], [Maternal Half Brother], [Paternal Half Brother], [Grandmother], [Grandfather], [Paternal Grandmother], [Maternal Grandmother], [Paternal Grandfather], [Maternal Grandfather], [Grandson], [Granddaughter], [Grandchild], [Uncle], [Aunt], [Niece], [Nephew], [Wife], [Husband]

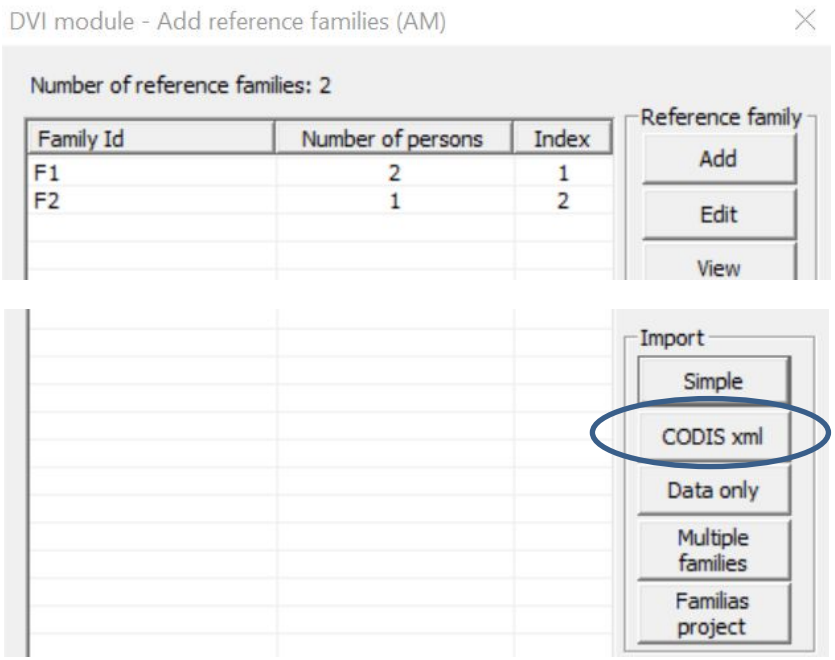
Simple



Similar to the tab separated files described previously for standard DNA data input. Several individuals may be included in the same file. Note, no family information can be included, each individual is placed into a single family. A line specifying a relation may be included, see below.

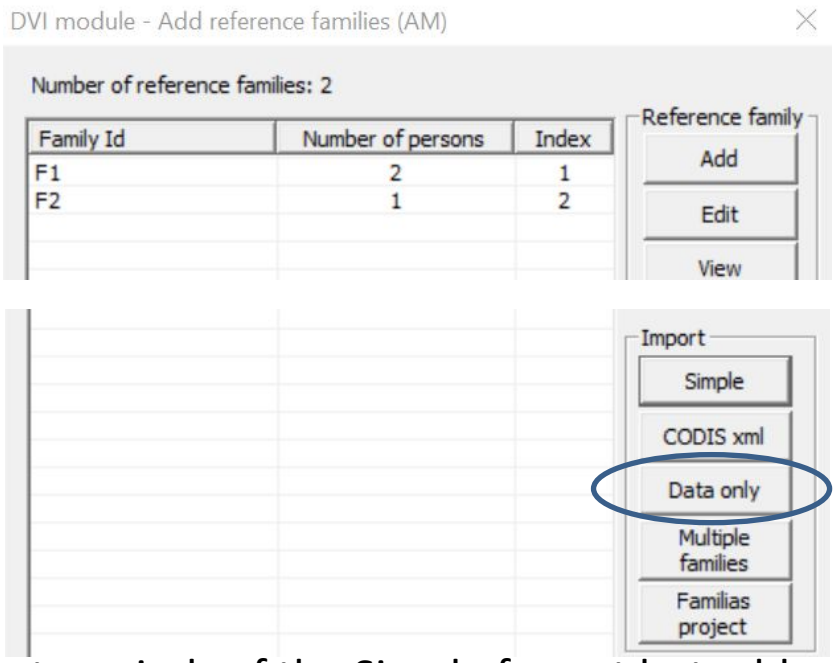
Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
[Brother]				
Person 1	8	9.3	20	21

CODIS xml



The format is described previously for standard DNA data input. This version will however recognize CODIS categories and create some family relationships based on those.

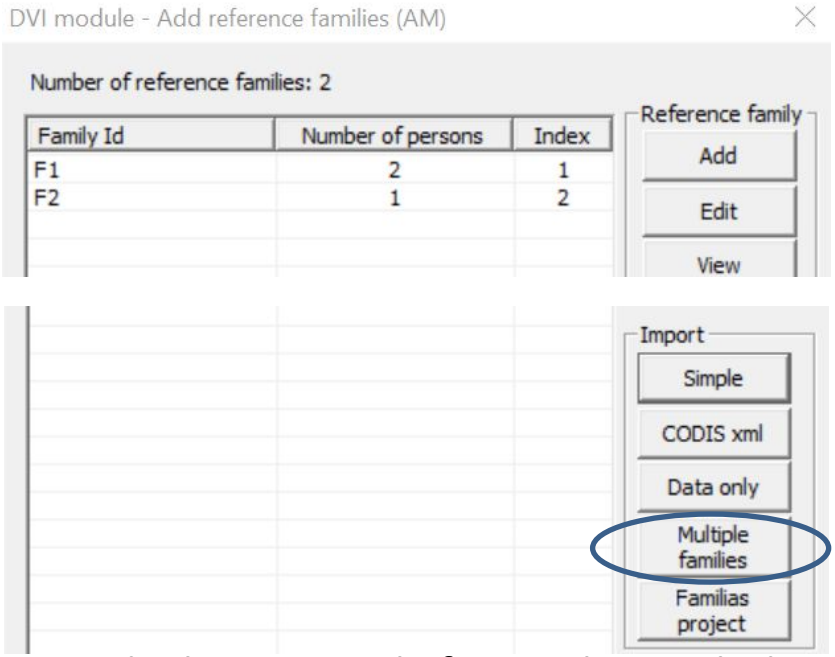
Data only



This format reminds of the Simple format but adds a column to define family designation, see below.

Family	Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Family 1	Person 1	8	9.3	20	21

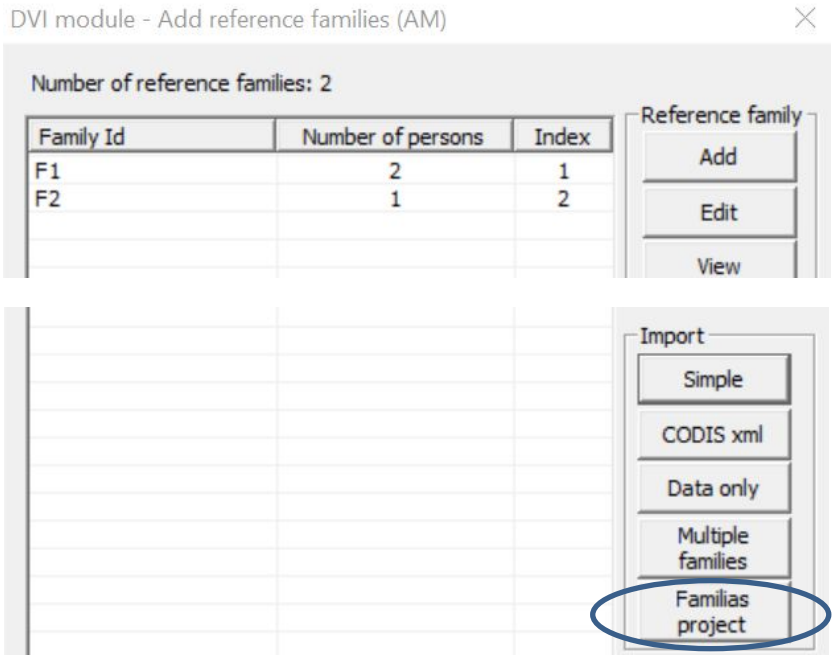
Multiple families



This format extends the Data only format but including an additional column specifying the relationship to the missing person within each family, see below.

Family	Relationship	Sample	TPOX 1	TPOX 2	D12S391 1	D12S391 2
Family 1	[Brother]	Person 1	8	9.3	20	21

Familias projects



This option allows the import of standard Familias project with some DNA data and pedigrees. Familias will automatically create families (one for each Familias file). The missing person should be named MISSING PERSON.

9. Output files and reports

9.1 Reports from Familias

rtf format

This is the standard output from Familias, i.e., the so called Familias report. It may be generated using different scopes, the simple containing only information about the LRs and pedigrees and the most extensive containing all information needed to reproduce the results.

The file is actually generated as a txt file with the rtf extension and is therefore not a true rich text formatted file.

csv format

This format will generate a comma (semicolons) separated file (csv). Output is illustrated below.

```
Output generated by Familias, version 3.2.2
Ped 1 vs Ped 2;
Marker;LR
SE33;8.22655
Total LR;8.226551889
Database;Norwegian
```

tab format

This format will generate a tab separated file. Output is illustrated below.

```
Output generated by Familias, version 3.2.2
Marker  LR
SE33    8.22655
Total LR    8.226551889
Database    Norwegian
```

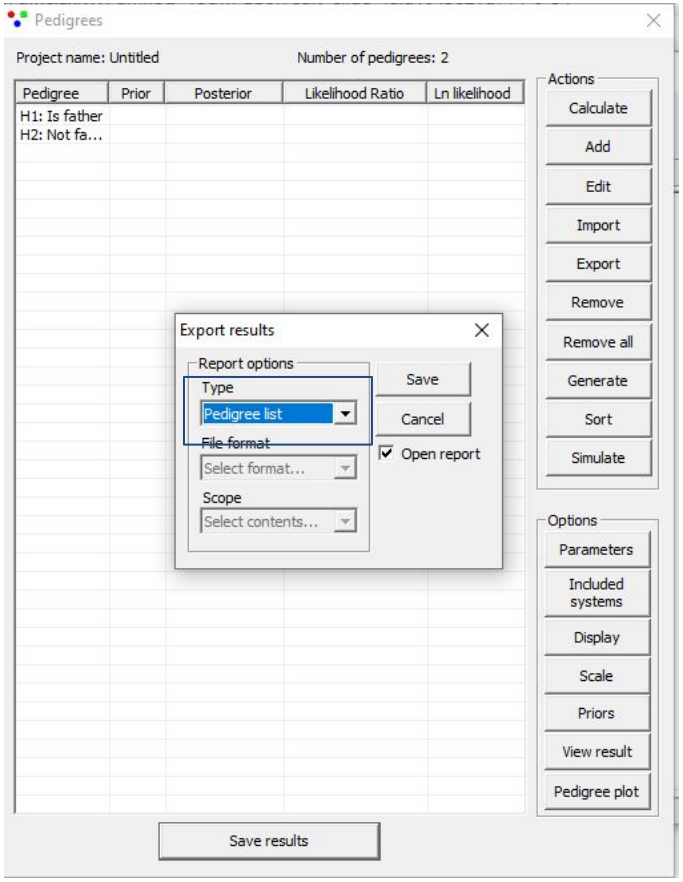
xml format

This format will generate an xml file with comprehensive content. This file may actually be opened as a project as well. An excerpt is illustrated below.

```
<?xml version="1.0" encoding="utf-8"?>
<FAMILIAS>
    <INFORMATION>
        <TITLE>Familias xml file</TITLE>
        <VERSION>3.2.2</VERSION>
        <TIMESTAMP>
        Mon Dec 11 09:53:31 2017
        </TIMESTAMP>
        <PROJECTNAME>testing_opt</PROJECTNAME>
        <DATABASEUSED>Norwegian</DATABASEUSED>
    </INFORMATION>
    <FREQUENCYDATABASES>
        <FREQUENCYDATABASE>
            <THETA>0</THETA>
            <DATABASENAME>Norwegian</DATABASENAME>
            <MARKERS>
                <MARKER>
                    <MARKERNAME>SE33</MARKERNAME>
                    <ISEXCLUDED>1</ISEXCLUDED>
                    <MARKERMODEL>
                        <FEMALEMUTATIONMODEL>Equal probability (Simple)</FEMALEMUTATIONMODEL>
                        <FEMALEMUTATIONRATE>0</FEMALEMUTATIONRATE>
                        <FEMALEMUTATIONRANGE>0</FEMALEMUTATIONRANGE>
                        <FEMALEMUTATIONRATE2>2.71134e-309</FEMALEMUTATIONRATE2>
                        <MALEMUTATIONMODEL>Equal probability (Simple)</MALEMUTATIONMODEL>
                        <MALEMUTATIONRATE>0</MALEMUTATIONRATE>
                        <MALEMUTATIONRANGE>0</MALEMUTATIONRANGE>
```

Pedigree list

This format will generate a tab-separated text file listing all the pedigrees as well as likelihoods, LRs and posteriors (if available)



9.2 Other

General output

Several dialogs/windows allow the export of the complete table of results, e.g. DVI search results. The output is simply a tab separated txt file that may be processed further in Excel or similar software.

Those exports are self-explained.

Summary reports

Several dialogs/windows allow the export of summaries, i.e., the blind search module allows the user to generate a summary of the results from a search, as does the DVI search function.

Export matrix

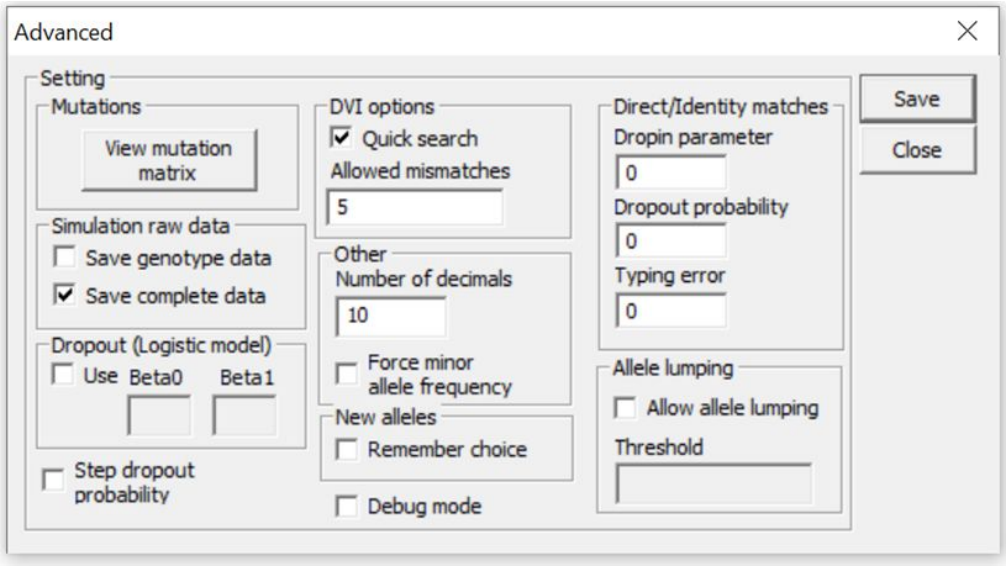
The blind search module allows the export of a distance matrix. The matrix will be of n x n dimensions where n indicates the number of individuals in the blind search. The matrix will compare all individuals and compute either,

1. A IBS distance between the two genotypes
2. The kinship coefficient, estimated using a maximum likelihood approach
3. The kappa2 and kappa0 estimates, using the same method as in 2.

10. Advanced

Description

There are some advanced options/settings in Familias reached via
File -> Advanced, see illustration on next page. Some of the settings are intended for expert users.



View mutation matrix – View and/or save the mutation matrix for the defined markers/systems.
Save genotype data – Decide whether or not to save the raw genotypes from a simulation
Save complete data – Decide whether or not to save all raw data (genotypes and likelihoods) from a simulation
Dropout – Enable a logistic dropout model whereby allele peak heights are needed (may be imported using the Genemapper format only). Recommended for advanced uses only.
Step dropout probability – If enabled together with dropout for a profile, Familias will compute the LR for a range of dropout probabilities (similar to a sensitivity analysis).
DVI options – By default a quick search feature is enabled whereby Familias first computes the LR disregarding mutations and then with mutations for markers where the likelihood is zero. Set the number of allow mismatches for Familias to even consider a computation. Disabling this feature may cause extensive computation time.
Number of decimals – Set the number of decimals to display and save when numbers are relevant.
Force minor allele frequency – If enabled Familias will force the use of the minor allele frequency (defined for each marker) whenever an allele frequency is below this value. Warning! May cause the sum of allele frequencies to sum to more than 1.
New alleles - Remember choice whenever a new allele is detected. For instance, the same frequency to all new alleles.
Debug mode - This will turn on some debug options, mainly through dumping files into the Familias install directory.
Direct/Identity matches – The values used in a direct match comparison. May also be changed in the blind search dialog.
Allele lumping – Not currently implemented. If enabled it will cause lumping of alleles which in turn will produce much shorter computation times. The lumping is an approximation whereby all transitions/mutations with a probability lower than the threshold is lumped.

11. Miscellaneous

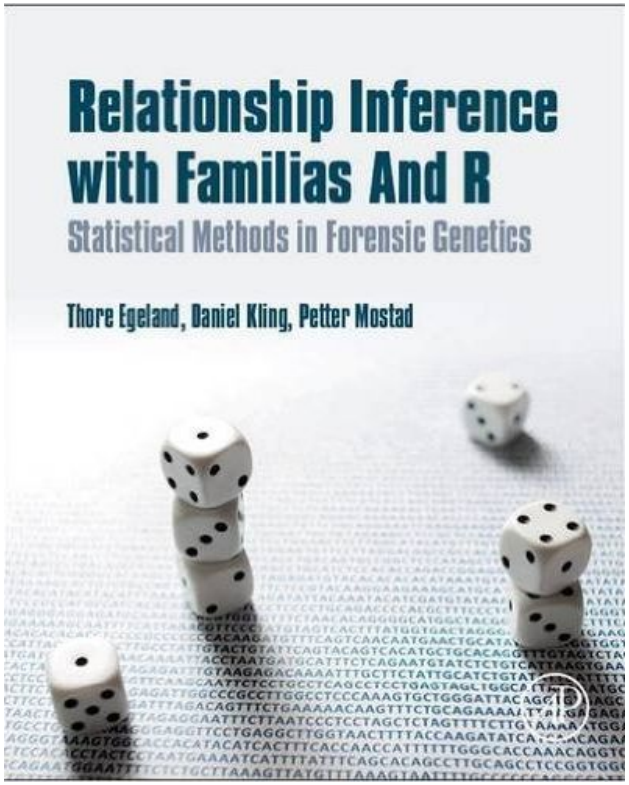
Create a frequency database

- **File > Create database** can be used to
  - create a database of allele frequencies
  - provide summary statistics, like heterozygosity
  - perform a blind search with the profiles of a population study
- A toy example (file) based on a sample of ten individuals for two markers

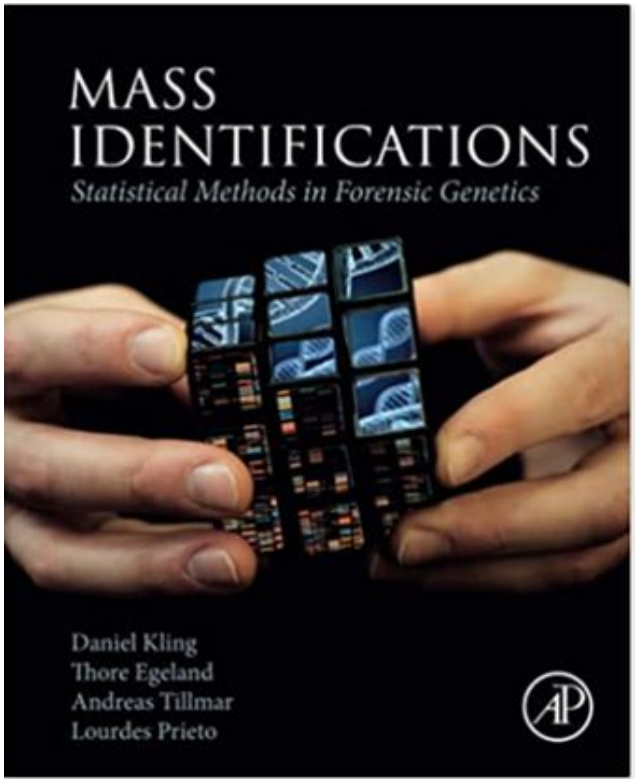
SampleID	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2
ID1	13	14	30	30
...	...	...	...	...
ID10	11	13	30	31.2

- Import: **File > Create database > Import**
- Create Familias file: **File > Create database > Create**
- Produce statistics: **File > Create database > Statistics**
  The file contains statistics at bottom and can also be imported to Familias
- Perform a blind search: **File > Create database > Check data**

More information: http://familias.no, http://familias.name and in the books:



Relationship inference



Mass Identifications

Useful links

Databases (ready to use):

- Norwegian
- Somali: Northeast Africa (Somalia, Ethiopia, Eritrea)
- Middle Asia (Pakistan/Afghanistan)
- Southeast Asia (Thailand, Vietnam)
- Spanish

Elias Hernandis has created a great tool to convert and produce data files for Familias (and other software). See https://leapdna.org/.

Papers referencing Familias. Validation

- It may be helpful to check papers that reference Familias
- An updated, but not exhaustive list, can be obtained by checking to
  - References to Egeland et al.(2000)
  - References to Kling et al. (2014)
- Kinship software, including Familias, was validated in
  - Drabek (2008)