Relationship Inference with Familias and R.

Statistical Methods in Forensic Genetics

Supplementary: Familias 3 Tutorial

Thore Egeland

Daniel Kling

Petter Mostad

October 20, 2015

Chapter 1

Familias tutorial

This online tutorial[1] supplements and refers to the Elsevier-book "Relationship Inference with Familias and R. Statistical Methods in Forensic Genetics", by Egeland, Kling and Mostad, and is based on the windows program `Familias` 3.1.9, referred to as `Familias`. There are some references to the mentioned book in this tutorial. However, these references are not necessary for the understanding of the tutorial. The purpose to explain briefly the *basic functionality* of the program, a complete description is provided in the manual available from `http://familias.no`.

The tutorial starts by discussing a standard paternity case. Then, we address the most important complicating factors: mutation, theta-correction and silent alleles. Finally, we present a more complicated example where more than two alternatives are considered. Some relevant papers include [72][2],

[1] Available from `http://familias.name/book.html`

[2] `Familias 3` reference: Kling et al. "Familias 3–Extensions and new functionality". FSI: Genetics, 13:12-127, 2014

and [36][3].

Four basic steps

There are four basic steps involved in a typical application of the program as illustrated in Figure 1.1. These steps suffice to perform the calculations for standard paternity cases. Below these steps are detailed for the paternity case summarised in Figure 1.2.

1. `General DNA data` window, Figure 1.3. Click `Add` to enter a marker. In the new window, enter `Marker1` and the two alleles `A` and `B`, both with frequencies 0.05. Enter the `C` allele with frequency 0.9. Press `Save`.

2. `Persons` window, Figure 1.4. Enter the persons: `AF` (alleged father), `Mother`, and `CH` (child) and their gender. Close window (this should generally be done before continuing).

3. `Case DNA data` window, Figure 1.5. Double-click each person to enter his or her DNA data as given Figure 1.2. In the new window, enter the appropriate allele system (use the pull-down menu) and the observed alleles for this person, then press `Add` and `OK`.

4. `Pedigrees` window, Figure 1.6. Click `Add` to enter the pedigree corresponding to hypothesis H1 (paternity). Enter `H1: AF father` as `Pedigree name`. Enter the `Mother` as the parent of `CH` in the pull-down menu and click `Add`. Similarly, enter `AF` as the parent of `CH`.

[3]Drábek. "Validation of software for calculating the likelihood ratio for parentage and kinship". FSI: Genetics, 3:112-118, 2009

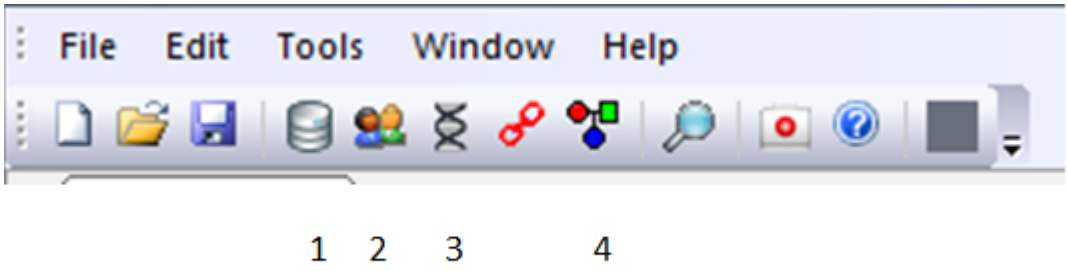


Figure 1.1: The basic windows of `Familias`.

Click `OK` to finish the definition of the pedigree corresponding to hypothesis H1. Click `Add` in the pedigree window once more to add the pedigree corresponding to hypothesis H2. Enter `H2: Unrelated` as `Pedigree name`. Press `Calculate`. Normally one would answer `Yes` when asked to save.

The output is shown in Figure 1.7, page 11. The $LR = 20$ as it should according to the equation $LR = 1/p_A = 1/0.05 = 20$. Furthermore, a prior probability of 0.5 for each alternative, gives the posterior as $LR/(LR+1) = 20/21 = 0.952381$, also indicated in the output window.

Advanced software is not required for the simple paternity case considered so far. However, mutation, theta corrections and silent alleles complicate matters as described below.

Specific mutation models

The default value for mutation rates is zero. However, if it is known or reasons to suspect that there is a non-zero mutation rate, it should be specified.

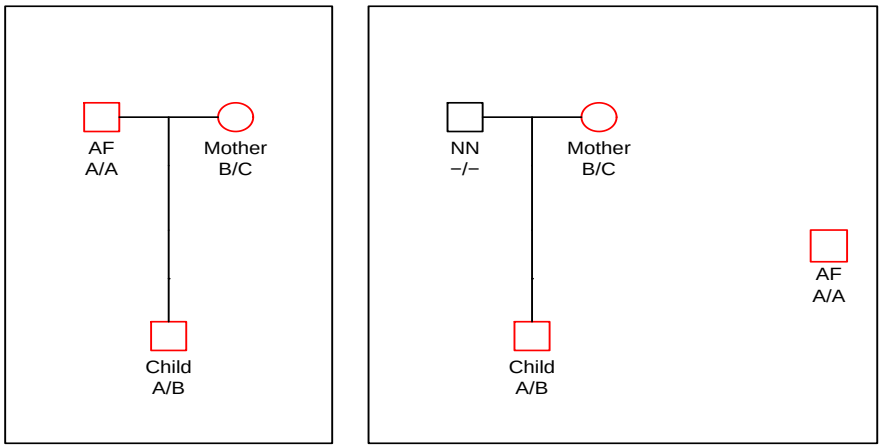


Figure 1.2: A standard paternity case. Left panel corresponds to hypothesis H_1, AF being the father while AF is unrelated, H_2, to the right.

A reasonable mutation rate could be around 0.005. The program offers the possibility to distinguish between male and female mutation rates. The reason for this is that paternal alleles tend to mutate more often than maternal alleles. There are 5 different mutation models to choose from as shown in Figure 1.8:

1. `Equal probability (Simple)`
2. `Proportional to freq.`
3. `Stepwise (Unstationary)`
4. `Stepwise (Stationary)`
5. `Extended stepwise`

If a model is stationary this implies that adding irrelevant persons will not affect the result. Conversely, for unstable models adding irrelevant persons may

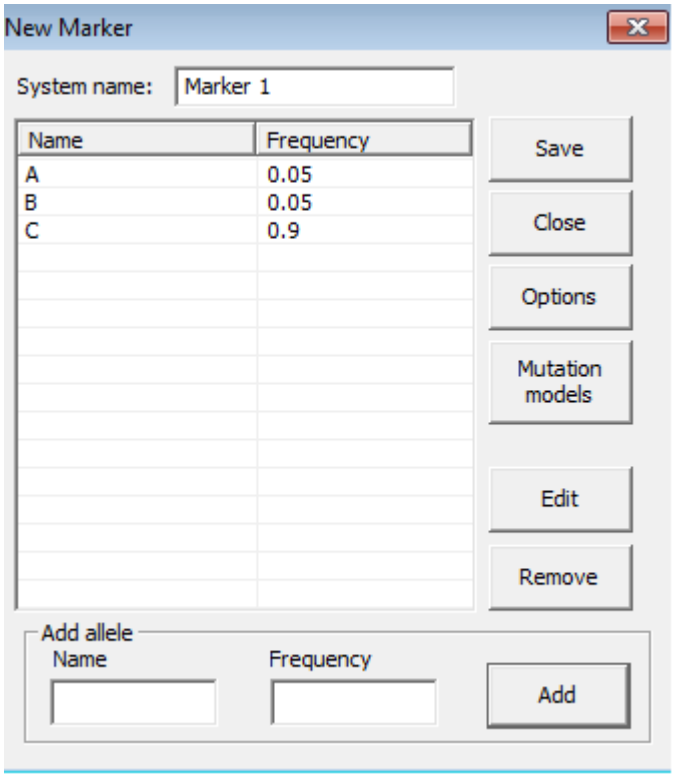


Figure 1.3: Defining a marker.

lead to slightly different results. Stationarity is not a natural biological condition, as allele frequencies do change over time. However, non-stationarity has the somewhat unpleasant consequence that the exact LR will change by including extra irrelevant persons in the calculations. Furthermore, a person's allele frequencies will be different if they are derived directly from the data base compared to if they are derived, with mutations, from parents having the database allele frequencies. Models 2 and 4 above are stationary.

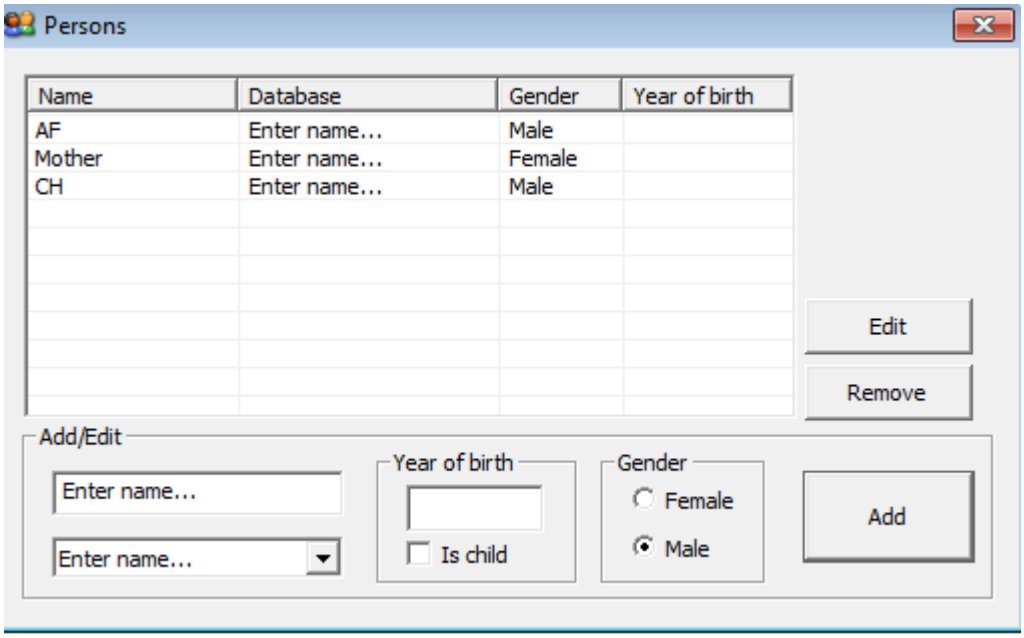


Figure 1.4: Defining persons.

Exercise 2.7, available from `http://familias.name/book.html`, exemplifies the above mutation models. The alleged father is 14/15 and the child 16/17. Without a model for mutations, the likelihood ratio would be 0. Using Model 1, `Equal probability (Simple)`, all mutations are equally likely. With reasonable parameter choices for Models 3, 4, and 5, the shortest mutation, the one from 15 to 16, is the more likely.

Theta correction

Deviation from Hardy Weinberg Equilibrium is the simplest case where the so called theta correction is needed. The input is illustrated in Figure 1.9.

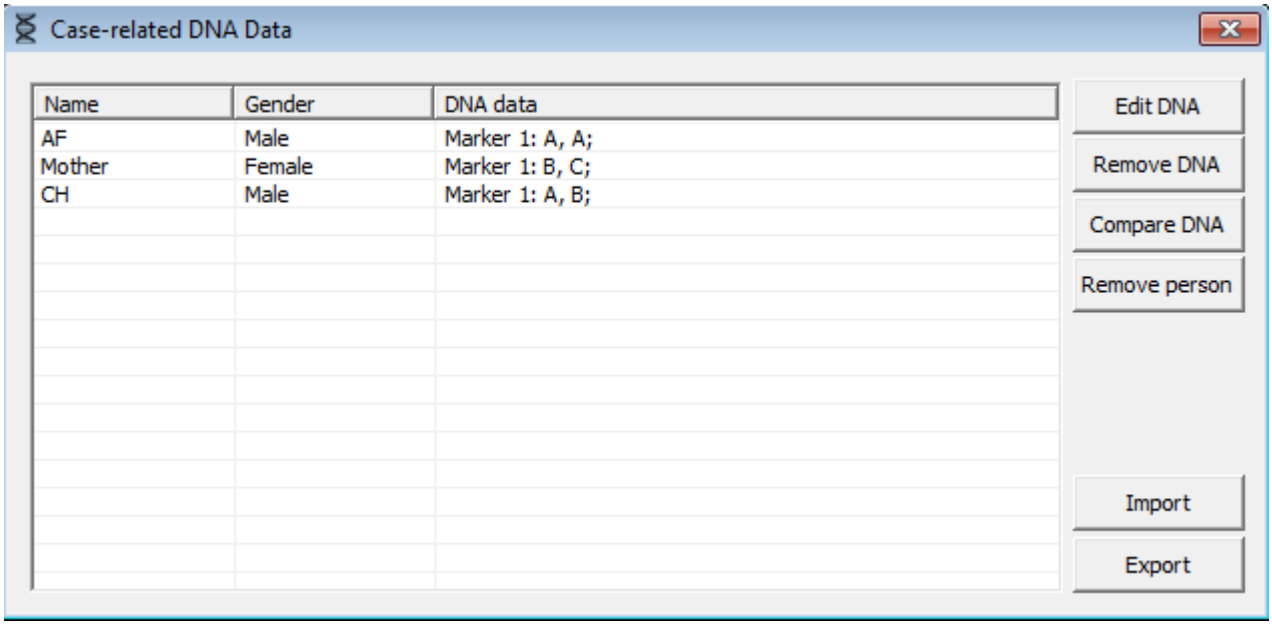


Figure 1.5: Genotype data.

Silent alleles

Silent alleles may be present when some homozygotes are observed. Figure 1.10 demonstrates the input. Note that both alleles need to be provided in the `Case DNA data` window.

Example. Brothers?

A woman M has 3 sons S1, S2, and S3, and the question is if a putative father PF is the father of all, some, or none of these sons. DNA data is available

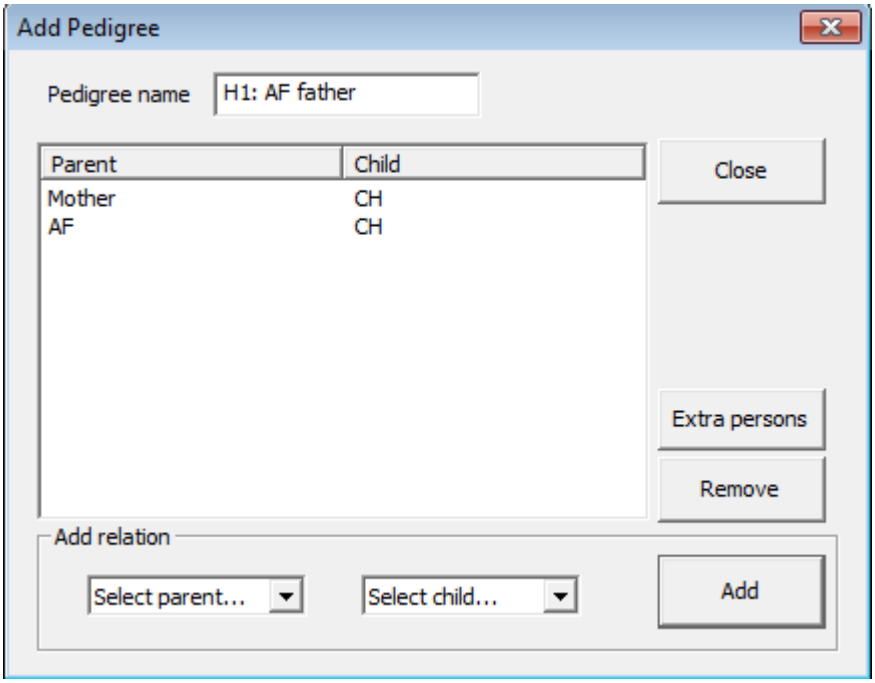


Figure 1.6: Pedigree definition of trio.

for S1, S2 and S3.[4] Data from 8 loci is given. In all loci, all alleles have frequency 0.05. The alleles are numbered 1, 2, 3, 4. With this notation, S1, S2, and S3 have the observations given in Table 1.1.

Note that `Familias` contains functions for automatic generation of sets of pedigrees. This may be useful in situations when a large number of pedigrees should be considered possible. In the example above, clicking the button Generate (and keeping the default settings) will generate a total of 8 pedigrees provided S1, S2 and S3 are defined as children, M and PF are given the same birth data and `M` is fixed to be the mother of S1, S2 and S3. The results are

[4]The file `http://familias.name/TutorialBrothers.fam` is available for those who would like to skip manual input.

Locus	sys1	sys2	sys3	sys4	sys5	sys6	sys7	sys8
S1	1/2	1/2	1/2	1/2	1/2	1/1	1/2	1/2
S2	3/4	3/3	3/4	3/4	3/4	1/2	1/2	2/3
S3	3/4	1/2	1/2	3/4	3/3	3/3	3/4	3/4

Table 1.1: Genotype data for brother example.

given in Figure 1.12. We are using by default a flat prior of $1/8 = 0.125$. Observe that pedigree8, the full brother alternative gives a likelihood (and hence a posterior) of 0. To understand this consider sys5 in Table 1.1. S3 is homozygous 3/3. This implies that the two other can display at most two alleles different from 3. However, they have three alleles, 1,2 and 4. The alternative specifying S2 and S3 as full brothers and half brother of S1, is the most likely. This pedigree appears as `Ped4` in Figure 1.11, The posterior probability 0.6245 from `Familias` can be confirmed by

$$\frac{\exp(-120.2663)}{4\exp(-122.6019)+\exp(-120.2663)+\exp(-122.058)+\exp(-123.3108))}$$

where the numerator is the likelihood for `Ped 4` and the denominator the sum of the likelihoods. Rather, than reporting the posterior probability, we can obviously report more conventional LR-s. Then a choice of reference, a denominator, need to be decided on, and several values have to be reported and this may be inconvenient.

Brother example continued

A stepwise stationary mutation model with mutation rate 0.005 and range 0.1 is used for all markers. As can be seen from Figure 1.13, the results

are now completely changed. The full brother alternative now comes out as the by far most likely alternative. This makes sense intuitively as there are several marker where pairs of individuals share both alleles.

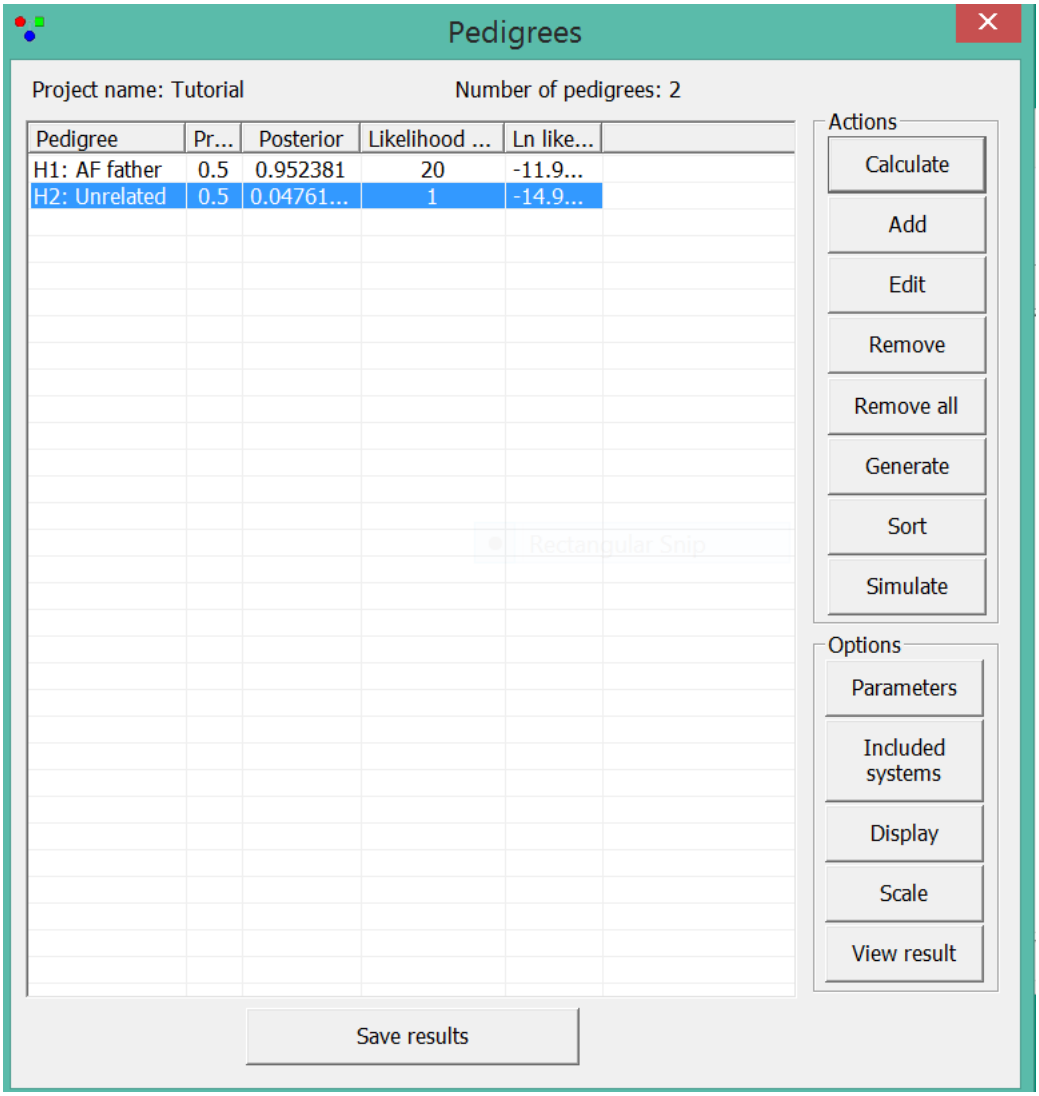


Figure 1.7: `Familias` output.

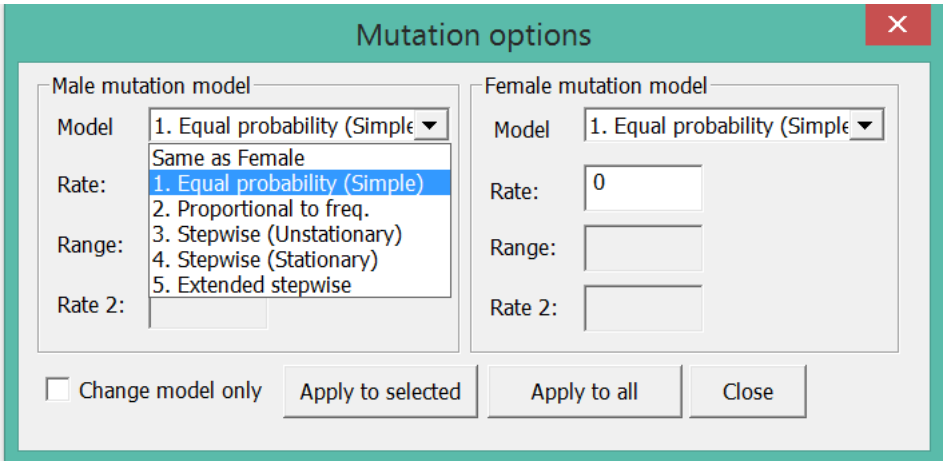


Figure 1.8: Mutation models

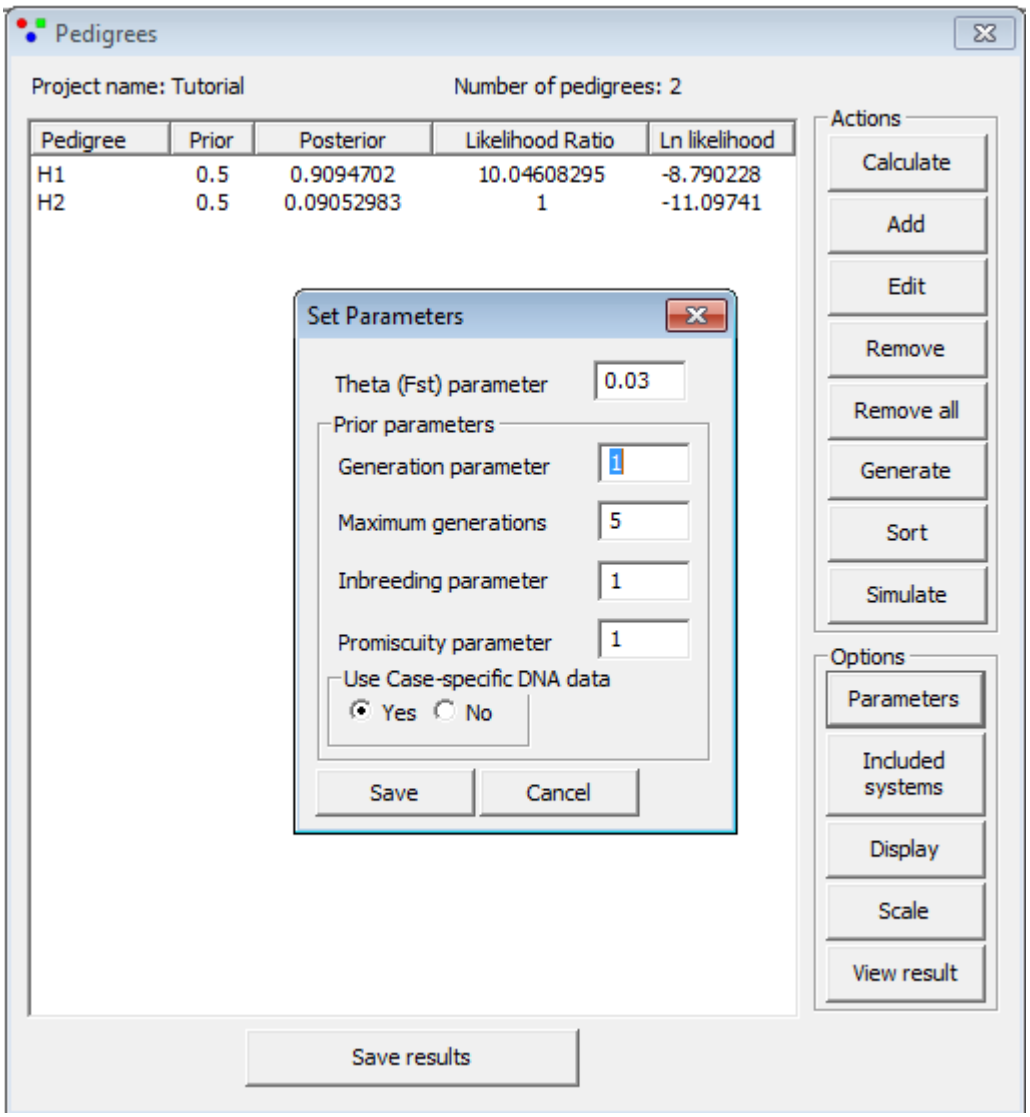


Figure 1.9: Press `Parameter` in the `Pedigrees` window to provide a θ value, 0.03 in the example. The LR is reduced from 20 to 10.05. The remaining parameters of this window are normally left unchanged. The default values correspond to a flat prior and changing them only affects the posterior probability, not LR.

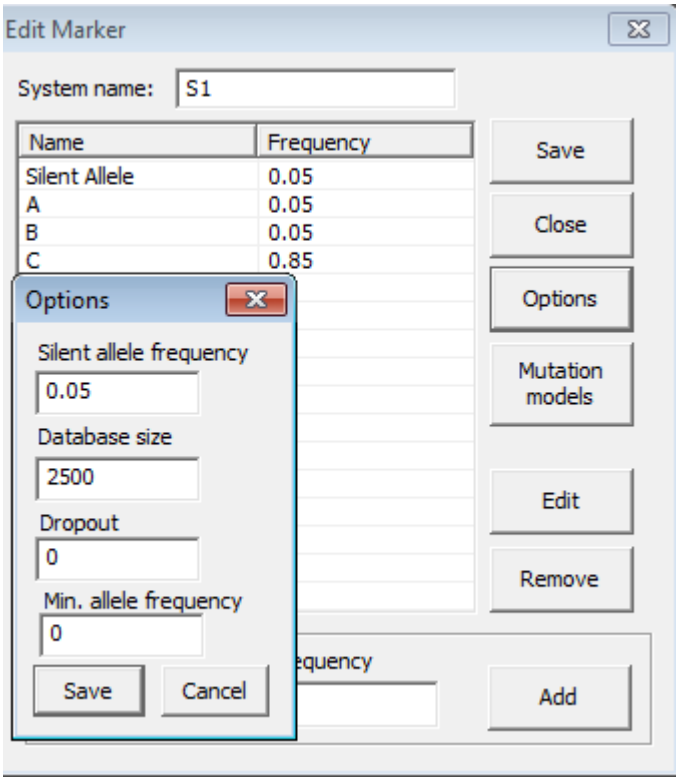


Figure 1.10: Press `Options` in the `Edit Marker` window to provide a silent allele frequency, 0.05, in the example. Note that allele frequencies including the slient allele must sum to 1. This can be achieved by reducing the frequency of the `C` allele to 0.85. The LR is reduced from 20 to 13.33.

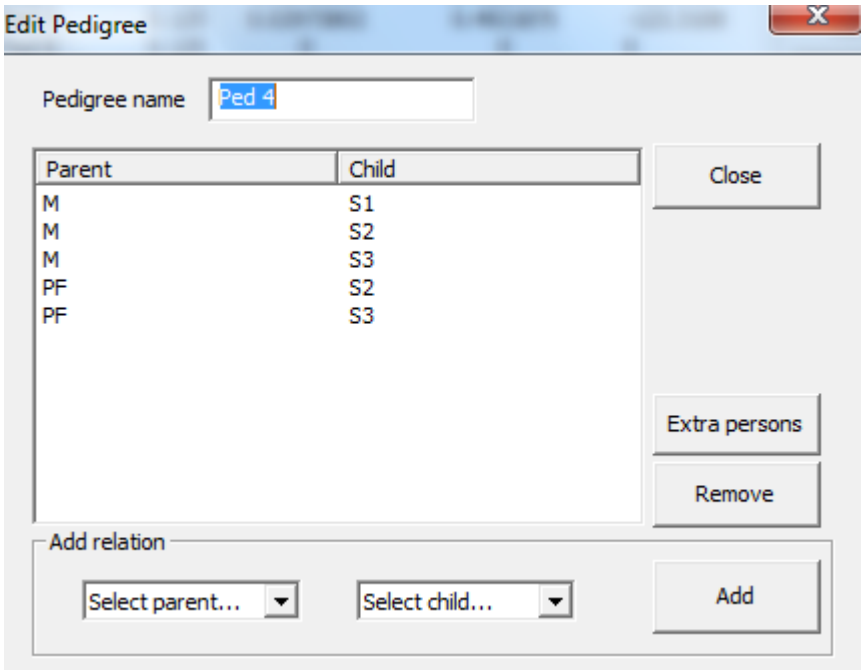
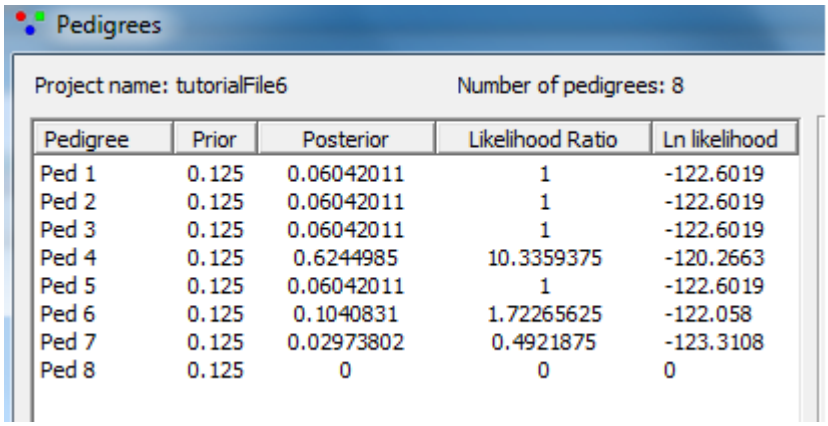
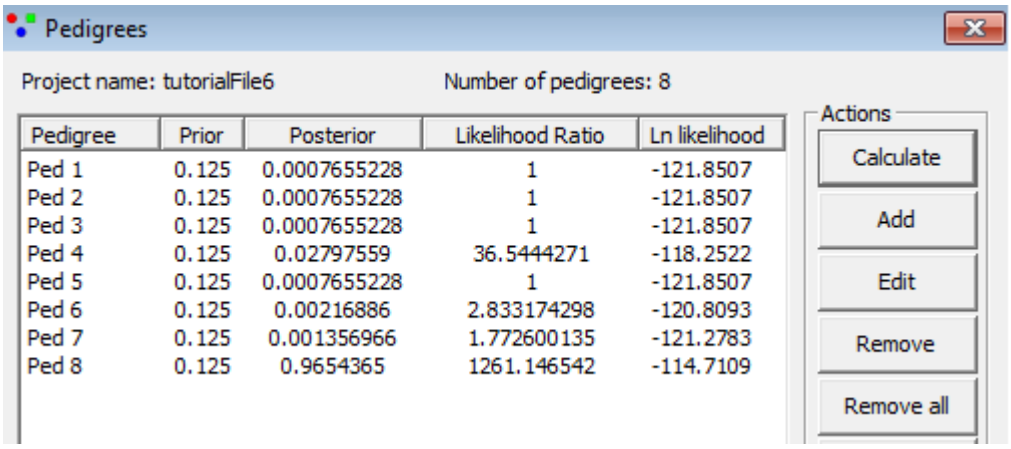


Figure 1.11: 8-pedigree example. The most likely pedigree.



Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
Ped 1	0.125	0.06042011	1	-122.6019
Ped 2	0.125	0.06042011	1	-122.6019
Ped 3	0.125	0.06042011	1	-122.6019
Ped 4	0.125	0.6244985	10.3359375	-120.2663
Ped 5	0.125	0.06042011	1	-122.6019
Ped 6	0.125	0.1040831	1.72265625	-122.058
Ped 7	0.125	0.02973802	0.4921875	-123.3108
Ped 8	0.125	0	0	0

Figure 1.12: `Familias` output for 8-pedigree example.



Pedigree	Prior	Posterior	Likelihood Ratio	Ln likelihood
Ped 1	0.125	0.0007655228	1	-121.8507
Ped 2	0.125	0.0007655228	1	-121.8507
Ped 3	0.125	0.0007655228	1	-121.8507
Ped 4	0.125	0.02797559	36.5444271	-118.2522
Ped 5	0.125	0.0007655228	1	-121.8507
Ped 6	0.125	0.00216886	2.833174298	-120.8093
Ped 7	0.125	0.001356966	1.772600135	-121.2783
Ped 8	0.125	0.9654365	1261.146542	-114.7109

Figure 1.13: 8-pedigree example with mutation model.