Solutions to using FamLink/FamLinkX,

This document contains solutions to the exercises for the exercises. Questions and comments may be sent to help@famlink.se or Daniel.kling@fhi.no

FamLink

Exercise 1, Using FamLink

The first exercise considers

- H1: The alleged father (AF) is the true father of the child
- H2: Another man, not related to AF is the true father of the child

a) -
b) Several mapping functions exists, e.g. Morgan – where genetic distance is linear to the recombination rate (1 cM = 1% recombination rate), Haldane – exponential function, Kosambi – exponential function that account for double crossovers. The first one is implemented in FamLink.
c) 25
d) Nothing will happen for the current case as linkage does not affect the results.
e) LR=1/(2*p12)*1/(2*p12)=1/(4*0.1*0.1)=25

Exercise 2, A case of siblings

The second exercise involves two persons P1 and P2. We consider several hypotheses,

- H1: P1 and P2 are full siblings
- H2: P1 and P2 are half siblings
- H3: P1 and P3 are unrelated

a) -
b) See table below. We see that linkage has a high impact for the current case and accounting for recombination is essential.

	LR(linkage)	**LR(no linkage)**
Full sibling	2469	915
Half siblings	49.7	30

c) See table below. The LR(linkage) will change and converge with LR(no linkage).

	LR(linkage)	**LR(no linkage)**
Full sibling	915	915
Half siblings	30	30

d) No, both maternal and paternal halfsiblings have the same inheritance patterns and can thus not be distinguished with autosomal markers (linked or not linked)

Exercise 3, An immigration case

In the third case we will consider an immigration case with hypotheses,

- H1: P1 is the uncle of P2
- H2: P1 is the grandfather P2
- H3: P1 and P2 are unrelated

a) See table below. We observe a difference between H1 and H2 with LR(linkage) while for LR(no linkage) they do have the same values. The difference is in this case small but may be larger for other combinations of alleles.

	LR(linkage)	**LR(no linkage)**
Uncle	41.9	30
Grandfather	46.5	30

b) We will get the same LR:s for both H1 and H2, 30.25.

c) See table below. The difference between the LR(linkage) for H1 and H2 is larger than in a) but slightly closer to LR(no linkage) as may be expected.

	LR(linkage)	**LR(no linkage)**
Uncle	34	30
Grandfather	40.4	30

Exercise 4, **Create your own pedigree

FamLink includes a number of predefined pedigrees where the user only needs to select the required pictures indicating the family structure. We may also create our own pedigrees using the Merlin Input file notation (see http://www.sph.umich.edu/csg/abecasis/merlin/tour/input_files.html). Consider a case of three persons interested to know whether they are all full siblings or unrelated,

- H1: Three full siblings
- H2: Three unrelated persons

a) -

b) -

  a. The search and subtract method allows for the allele frequencies to sum to more than 1.0 prior to the calculations. The frequencies of the alleles in the current case are later used while the rest of the alleles are removed and a rest allele is added such that the sum of the frequencies is 1.0. Normalization will simply normalize all the frequencies such that they sum to 1.0.

c) The LR is large, as may be expected given the hypotheses and the data. LR(linkage) = 3979201 and LR(no linkage) = 996835. We see again the importance of accounting for linkage.

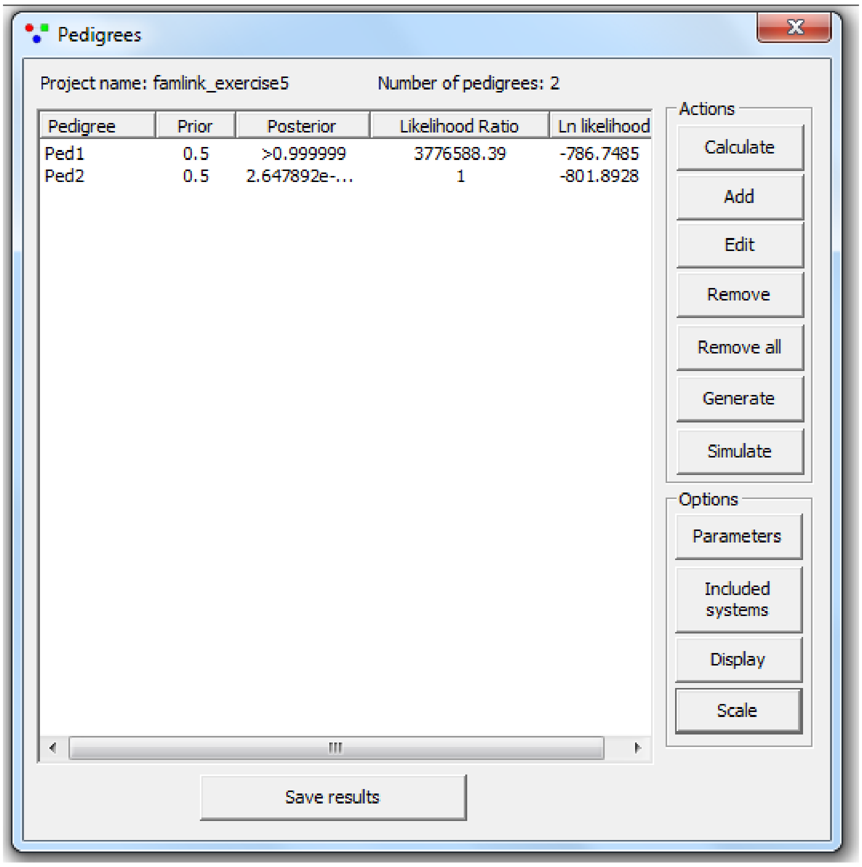
Exercise 5, Analyze Familias files in FamLink

In addition to creating our own pedigrees we may also analyze previous Familias projects (v 1.81 or above) to obtain an LR where linkage between a number of markers are considered. (All common markers as well as a number of uncommon markers are predefined with their genetic distances).0 We may now consider more markers and more complicated pedigree structures. Consider,

- H1: See Figure 1, where the alleged father is the true father of the child
- H2: See Figure 1, where the alternative father is instead the father of the child

a) -

b) See figure below.



c) Open the Quick analysis interface in FamLink. Mark generate report and browse to the file famlink_exercise5.txt and open.

d) -

e) The computed LR in FamLink is 1.02079e+006, which is about 37765588.39/(1.02079*10^6)=37 times lower than the LR computed in Familias.

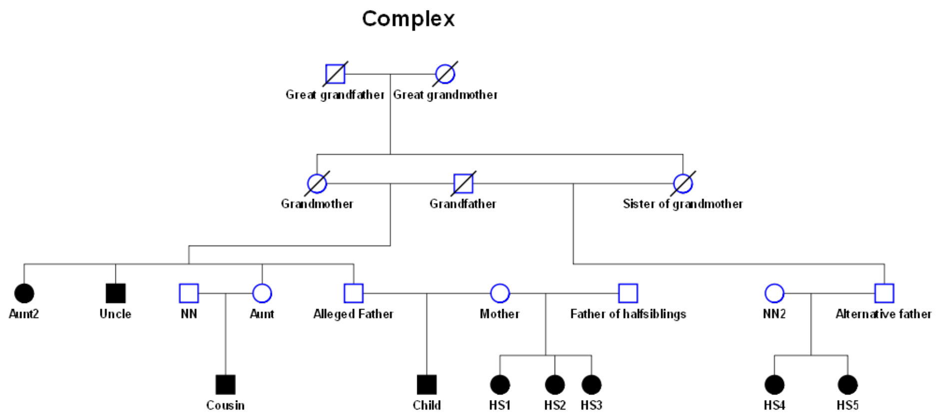


Figure 1, extended pedigree where the alleged father and an alternative father are related

FamLinkX

FamLinkX implements a new algorithm for linked markers on the X-chromosome. In addition to linkage the software accounts for linkage disequilibrium (allelic association) and mutations. The software is intended to be user-friendly but may provide obstacles for the inexperienced user. FamLinkX provides the likelihood ratios using three different methods, M1: Exact model, considering linkage, linkage disequilibrium and mutations; M2: Cluster approach, see manual for Merlin, linkage and linkage disequilibrium is considered but not recombinations within clusters and not mutations; M3: Only linkage is considered between markers. In the following exercises we are interested in M1 as this is the preferred model, specially for STR markers.

Exercise 6, Using FamLinkX

Comment [ET1]: klare ikke å lese famlinkx_exercise6.sav. Fint om vi kan se på FamlinkX sammen, jeg hadde problemer

The first exercise is meant to give a general overview of the functions of FamLinkX using an example. Consider the sample example as in exercise 1. Consider a paternity case (Duo) with hypotheses,

- H1: The alleged father (AF) is the true father of the female child
- H2: Another man, not related to AF is the true father of the female child

a) For X-chromosomal data the inheritance pattern is different for females and males. Specifying genders is crucial.

Create a new cluster and specify two allele systems, L1 and L2 with alleles 12, 13 and 16, 17 respectively. Let p(12)=0.4, p(13)=0.6 for L1 and p(16)=0.4, p(17)=0.6 for L2. Select the Simple mutation model with the mutation rate set to 0 for both systems. Set the genetic distance to 10 cM for L1 and 10.1 cM for L2.

Comment [ET2]: hvordan?

b) That depends on what mapping function you use to convert the genetic distance to recombination rate, but for small distances we get pretty much the same results, 0.001 for markers separated by 0.1 cM.

Furthermore, specify haplotype observations according to the table below..

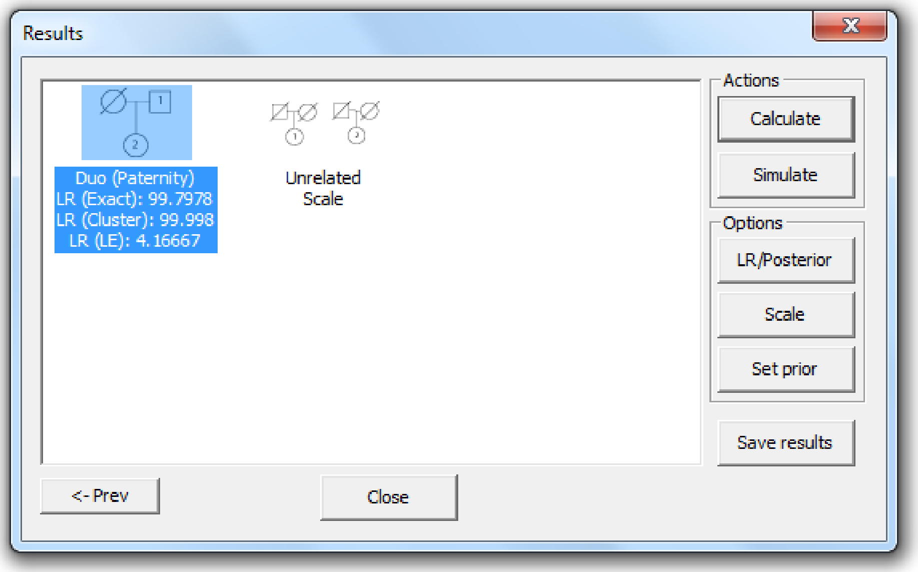
	12	13
16	39	1
17	1	59

c) We need this to estimate haplotype frequencies as for the tightly linked markers within each clusters we also have to consider linkage disequilibrium.

d) -

e) -

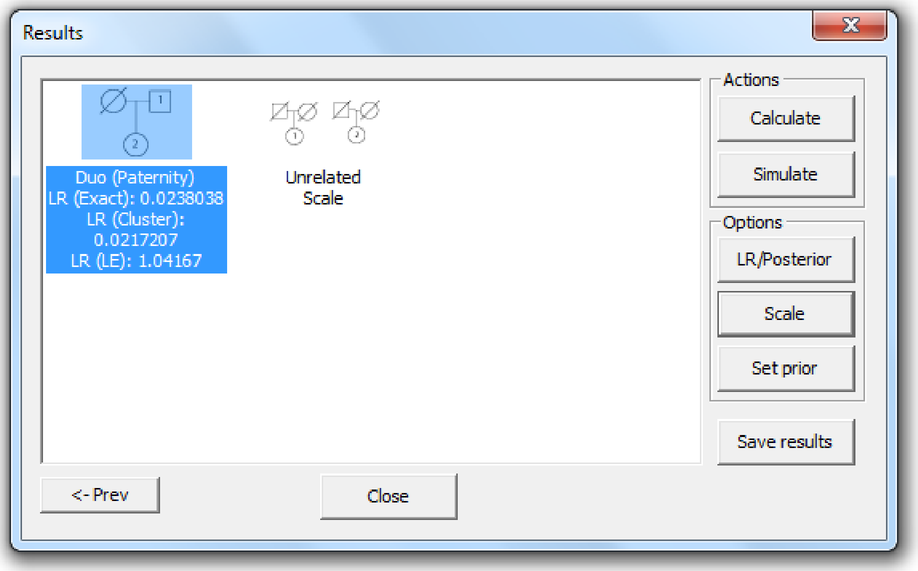
f) See figure below.



g) N/A for now, ask the teacher for comments

h) See figure below

Comment [TE3]: Jeg får litt annet: 0.0217208 og (Cluster) 0.0217207; suste samme



i) The constructed data is currently in extreme LD, see table in b). Adding observations such that the data is "more" in LE will yield a smaller difference from the LE model.

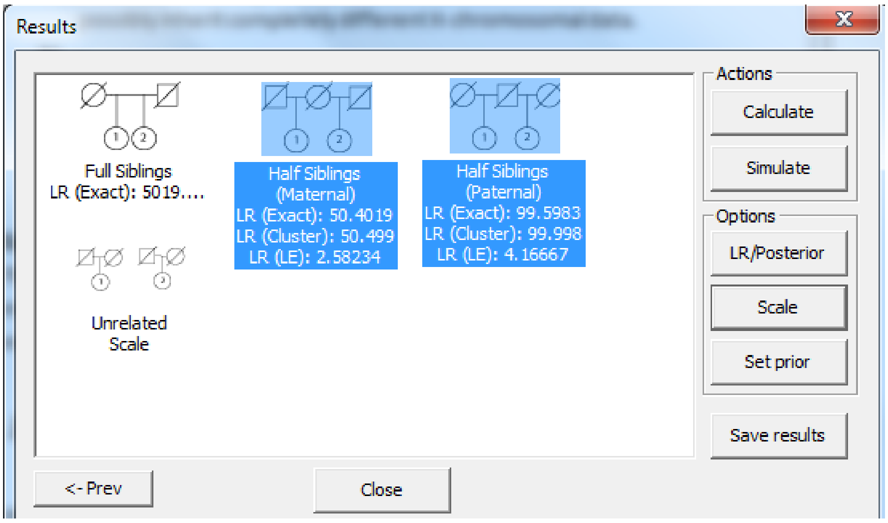
Exercise 7, A case of siblings revisited

In the second exercise we revisit the example in exercise 2 with siblings. Two females, P1 and P2, are interested to find out whether they are siblings in some way. We specify hypotheses,

- H1: P1 and P2 are full siblings

- H2: P1 and P2 are maternal half siblings
- H3: P1 and P2 are paternal half siblings
- H4: P1 and P3 are unrelated

a) Since paternal and maternal half siblings has different inheritance patterns. Whereas paternal half siblings must share at least one allele at each locus, maternal half siblings can possibly inherit completely different X-chromosomal data.
b) -
c) -
d) -
e) See figures below.



Exercise 8, Importance of Lambda

The third exercise is intended to provide some knowledge about how the haplotype frequencies are estimated and the importance of the parameter lambda. Our model for haplotype frequency estimation is described in the equation below.

$$H_i = \frac{c_i + \lambda p_i}{C + \lambda}$$

Where H_i is the haplotype frequency, c_i is the number of observations for the haplotype, p_i is the expected haplotype frequency (linkage equilibrium) calculated using the allele frequencies, C is the total number of observations for all haplotypes and lambda is a parameter giving weight to the expected haplotype frequencies. This model allows for unobserved haplotypes to be accounted for, in contrast to models which estimates the haplotype frequency solely based on the counts. The difficulty lies in the implementation of a “good” Lambda. Our recommendation is to compute the LR for a number of different values and select the most conservative one.

We specify a case with an aunt of a female child

- H1: The female is the aunt of the child
- H2: The two females are unrelated

a) -
b) -
c) See table below.

Lambda	LR
0.0001	13
1.0	10.7
10	4.5
100	1.5
1000000	1.02

d) The LR converges with the LE approach, i.e. by setting a high value for lambda we give high prior weight to the expected haplotype frequencies, which is basically equal to considering the markers to be in linkage equilibrium.
e) The opposite as in d), we now only use the observations, meaning that unobserved haplotypes will have very low frequencies and thus cases where they do appear will tend to have very high LR:s, given that the persons share that haplotype.

*Exercise 9, A challenge

This last exercise provides a challenge where the user need to combine the results from Familias and FamLinkX to obtain a final result. The data is extracted from a real case where three females provided data and the following hypotheses,

- H1: The three females are all full siblings
- H2: Any other pedigree constellation

Obviously H2 cannot be used in the current setting and we need to refine possibly alternative hypotheses.

a) -.
b) The most probable pedigrees are that they are all full siblings or that the person denoted 2 are paternal or maternal halfsibling of the other two. Obviously for autosomal markers the latter two cannot be distinguished and we must resort to X chromosomal markers. See solution file for details.
c) -

d) We specify three hypotheses, all three persons as full siblings and alternative hypotheses that 2 are paternal halfsibling or maternal half siblings to the other two (them still being full siblings)
e) -
f) See FamLinkX report in the exercise folder. The LR is largely in favor of the full siblings hypotheses.

g) -