Exercises using FamLink/FamLinkX,

The exercises are intended to provide an introduction to FamLink and FamLinkX, freely available at http://www.famlink.se. For introductory examples/exercises on the general use of FamLink, see the webpage. Exercises with one * is considered to be intended for intermediate users, while questions with ** are for advanced users.

FamLink

Exercise 1, Using FamLink

We will first consider a simple exercise where the purpose is to get familiar with the user interface and create your frequency database. Consider a paternity case (Duo) with hypotheses,

- 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) Specify two allele systems, L1 and L2 with alleles 12, 13 and 14 for both systems. Let p(12)=0.1, p(13)=0.2 and p(14)=0.7 for both systems. The loci are closely linked, specify the recombination rate to 0.01.
b) In several situations we only have the genetic position/distance of the markers. How can we convert between genetic distance and recombination rate?
c) Select the appropriate pedigrees and specify the data for the father as homozygous 12, 12 for both loci and the child heterozygous 12, 13 for both loci. Calculate the LR which should coincide with the theoretical value of 25. (Hint: press the LR/Posterior (W) button to change the displayed results)
d) Try changing the recombination rate to 0.5 and calculate the LR again. What happens? Explain why?
e) Use pen and pencil to show that LR=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) Specify two allele systems as in the previous exercise: L1 and L2 with alleles 12, 13 and 14 for both systems. Let p(12)=0.1, p(13)=0.2 and p(14)=0.7 for both systems. The loci are closely linked, specify the recombination rate to 0.01. Select the appropriate pedigrees and specify the data for P1 as homozygous 12, 12 for both loci and P2 as homozygous 12, 12 for both loci.

b) Calculate the LR. Scale versus the Unrelated pedigree and discuss the LR:s for full siblings and half siblings
c) Try changing the recombination rate to 0.5 and calculate the LR again. What happens? Why?
d) *Can linked autosomal markers be used to distinguish maternal from paternal half siblings?

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 of P2
- H3: P1 and P2 are unrelated

Use the same allele systems as in Exercise S2 but specify the recombination rate to 0.1. Specify the alleles for P1 to 12, 12 for both loci and for P2 to 12, 12 for both loci.

a) Calculate the LR and scale versus Unrelated. Discuss the results
b) Without changing the recombination rate to 0.5, discuss what will happen if you were to do this.
c) Calculate the LR also for r=0.25 and compare the results from a), b) and c)

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

We will use the same frequency database as in Exercise S2, but specify the recombination rate to 0.1

a) Specify the needed relationships in a text file and rename it to exerciseS4.ped. Specify the data for all persons to 12, 12 at the first locus. For the second locus we will add a previously unseen allele, denoted 11. Specify the data for all persons to 11, 12.
b) Import the pedigree into FamLink. (This is done byFile>New wizard>Import pedigree). When the new allele is imported we require a frequency. Specify 0.05 for allele 11 and select “Search and substract” as method.
    a. Discuss the “Search and substract” approach compared to “Normalization”
c) Calculate the LR and discuss the results.

Exercise 5, Analyze Familias files in FamLink

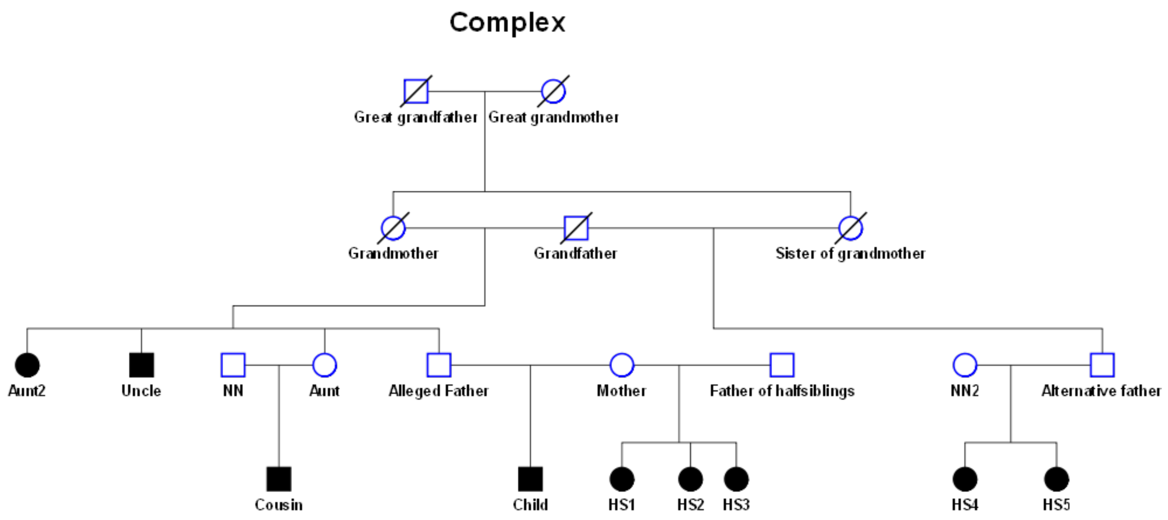
In addition to creating our own pedigrees we may also analyze previous Familias projects (v 1.81 or above) to obtain an LR where linkage between a number of markers are considered. (All common markers as well as a number of uncommon markers are predefined with their genetic distances.) We may now consider 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) Open the file famlink_exercise5.txt in Familias 3 and explore the project. (Note that the person marked as cousin in Figure 1 is not included since Familias version <3.0 could not handle the complexity of the project). The project contains genetic marker data for 35 autosomal STR markers, see Familias.

b) Calculate the LR in Familias. (Be patient, the computations may take >20 minutes depending on the performance of your computer. Possibly Familias may hang due to running out of memory)

c) Open the Quick analysis interface in FamLink. Mark generate report and browse to the file famlink.exercise5.txt and open.



d) Perform an analysis of the Familias project by pressing Analyze.

e) Open the generated report (found in the same directory as the exercise file) and browse the contents. Find the LR and compare it to the one obtained in Familias.

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

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) Why is it important that we specify the gender of all persons?

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.

b) What is the corresponding recombination rate between L1 and L2?

Furthermore, specify haplotype observations according to table below.

	12	**13**
16	39	1
17	1	59

c) Why do we specify the number of observations for each haplotype?
d) Specify Lambda to 0.0001 in Options. (We will discuss the importance of Lambda in exercise 8 and will not dwell further on it now). In brief, setting a low Lambda means we give a high weight to the observed haplotypes.
e) Select the appropriate pedigrees and specify the data for the father as 12 for L1 and 17 for L2 and the child homozygous 12, 12 for L1 and 17, 17 for L2.
f) Calculate the LR which should coincide (approximately) with the theoretical value of 100. There may be a small deviation from the theoretical value, see exercise 8.
g) ** Try deriving the theoretical formula.
h) Change the genotypes for the child to 12, 13 for L1 and 16, 17 for L2. Calculate the LR.
i) Change the number of observations for each haplotype. What happens? Explain why?

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) Explain why we may distinguish H2 from H3 with X-chromosomal markers but not with autosomal markers.
b) Use the same frequency data as in exercise 6. (Specify Lambda=0.0001) Select the relevant hypotheses.
c) Enter data for both P1 and P2 as 12, 12 for L1 and 17, 17 for L2.
d) Calculate the LR.
e) Scale against H4: Unrelated. Comment on the importance of accounting for LD and linkage in the current case.

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) Again use the same frequency data as in exercise 6. Select the relevant hypotheses.
b) Enter data for the child as 12, 12 for L1 and 16, 17 for L2 and for the aunt as 12, 13 for L1 and 17, 17 for L2.
c) Calculate the LR for different values of lambda, see table below.

Lambda	**LR**
0.0001	
1.0	
10	
100	
1000000	

d) What happens for large values of lambda?
e) What happens for small values of lambda? (Hint: use the equation for haplotype frequency estimation above)

*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 (anonymized) 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, in a simple way, and we need to refine possibly alternative hypotheses.

a) We may assume that all females are children with no children of their own. Also define two parents, a mother and a father and specify that they are both born 1970.
b) Use Familias (Generate pedigree) to find the pedigrees with a posterior probability above 0.001. (Use the database database_5.fam from the Familias exercises and the data provided in exercise_9.txt). Which are the most probable relationships according to the results? Interpret the results.
c) Open the database database_9.sav in FamLinkX, which contains frequency and haplotype data for the Argus X12 kit from Qiagen based on a Swedish population sample.
d) Use the results in b) to specify hypotheses in FamLinkX.
e) Specify the DNA data, available in exercise_9_X.txt. (Make sure to import the data in the file to the correct corresponding persons, hint: the person denoted 2000 should be imported to 3.)
f) Calculate an LR and interpret the results. (Be patient, the computation may require some time >20 min). To speed the computations up, go into Advanced, select and edit the hypothesis we have selected to investigate. For each pedigree set the Threshold value to 0.001 and the Steps value to 0.
g) Discuss if the LR in b) may be combined with the results in a)? What is your final conclusion on the case?