c) ** Using the sampling formula provided in chapter 2 and that $\theta = 0.02$, compute the updated set of allele frequencies. *Hint*: Given zero alleles are IBD we have four different observations, whereas given one allele IBD we have three different observations)

d) ** Compute the LR using that $\theta = 0.02$. *Hint*: Use the results in b) and c))

e) ** Plot the LR versus the value of θ using the specifications in b).

4.5.2 X-chromosomal markers and FamLinkX

`FamLinkX` implements an 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, but comparisons to the other models will be made. For all calculations, unless stated otherwise, we consider X-chromosomal marker data and corresponding inheritance patterns.

Exercise 4.12 (A paternity case revisited. Video)**.**

	12	13
16	59	1
17	1	39

Table 4.15: Haplotype observations for Exercise 4.12 .

We will first revisit the paternity case (Duo) (for illustration see Figure 2.6) with hypotheses

H_1**:** The alleged father (AF) is the true father of the female child.

H_2**:** The alleged father and the child are unrelated.

a) Open `FamLinkX` and specify the frequency database. *Hint*: `File->Frequency database`. Create a new cluster and specify two diallelic systems, L1 and L2, with alleles 12, 13 and 16, 17 respectively. Let $p_{12} = 0.6$, $p_{13} = 0.4$ for L1 and $p_{16} = 0.6$, $p_{17} = 0.4$ for L2. Select the `Simple mutation model` with the mutation rate set to 0 for both systems. Set the genetic position to 10 cM for L1 and 10.1 cM for L2. Furthermore, specify haplotype observations according to Table 4.15. Why is it important that we explicitly specify the gender of all persons in calculations for X-chromosomal marker data?

b) What is the estimated recombination rate between the two loci? Use Haldane's mapping function.

c) Why do we specify the number of observations for each haplotype?

d) Use the equation below to calculate a measure of the association between the alleles

$$r^2 = \frac{(p_{12}p_{16} - p_{12,16})^2}{p_{12}p_{13}p_{16}p_{17}}. \tag{4.12}$$

e) Specify λ to 0.0001 in `Options`. We will discuss the importance of λ in Exercise 4.14 and will not dwell further on it now. In brief, setting a low λ gives large weight to the observed haplotypes in Table 4.15. Select the appropriate pedigrees using the Wizard. The alternative hypothesis depicts two unrelated girls, but the first hypothesis will override the genders. *Hint*: `File->New wizard` and specify the data for the father as 12 for L1 and 17 for L2 and the child as 12/12 for L1 and 17/17 for L2. Calculate the LR which should coincide (approximately) with the theoretical value of 100. Choose to save the file when asked. There may be a small deviation from the theoretical value, which we will return to in later exercises.

f) Change the genotypes for the child to 12/13 for L1 and 16/17 for L2. Calculate the LR.

g) Change the number of observations for each haplotype. What happens? Explain why!

h) * Discuss the high degree of LD and if this is a likely situation to occur in reality.

Exercise 4.13 (A case of sibship revisited)**.**

In the second exercise we revisit the example in Exercise 4.2 concerning disputed sibship. Two females, F1 and F2, are interested to find out whether they are siblings in some way. We specify hypotheses

H_1**:** F1 and F2 are full siblings

H_2**:** F1 and F2 are maternal half siblings

H_3**:** F1 and F2 are paternal half siblings

H_4**:** F1 and F2 are unrelated

a) Explain why we may distinguish H_2 from H_3 with X-chromosomal markers but not with autosomal markers.

b) Use the same frequency data, and haplotype observations, as in Exercise 4.12, alternatively open the file **Exercise4_13.sav**.

c) Specify $\lambda = 0.0001$ and select the relevant hypotheses. Note: if you also stored the case-related DNA data in the previous exercise, you may be asked if you wish to erase all DNA data, answer yes.

d) Enter data for both F1 and F2 as 12/12 for L1 and 17/17 for L2. Calculate the LR Scale against H_4. Comment on the importance of accounting for LD and linkage in the current case.

Exercise 4.14 (On the importance of λ)**.**

This exercise is intended to provide some insight into how the haplotype frequencies are estimated and the importance of the parameter λ. Our model

for haplotype frequency estimation is described by

$$F_i = \frac{c_i + p_i\lambda}{C + \lambda} \tag{4.13}$$

where F_i is the haplotype frequency for haplotype i, c_i is the number of observations for the haplotype, p_i is the expected haplotype frequency (assuming linkage equilibrium) calculated using the unconditional allele frequencies, C is the total number of observations for all haplotypes and λ 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 estimate the haplotype frequency solely based on the counts. The difficulty lies in the choice of a good λ. Our recommendation is to compute the LR for a number of different values and select the least extreme value, i.e., the value closest to 1.

To start, we specify a case with an aunt of a female child

H_1 : The female is the aunt of the child.

H_2 : The two females are unrelated.

a) Again use the same frequency data as in Exercise 4.12. Select the relevant hypotheses. *Hint*: Select the *Aunt/Uncle* as the main hypothesis.

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 d $\lambda = 0.0001$, 0.01, 1, 100 and 10000.

d) What happens for large and small values of λ? *Hint*: use the equation for haplotype frequency estimation above.

e) Change the data for the child to 13/13 for L1. Repeat c) and discuss the results.

Exercise 4.15 (Extended example. Combining `Familias`, `FamLinkX`)**.**

This exercise provides a challenge where the user needs 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 DNA samples. The hypotheses are

H_1 : The three females are all full siblings.

H_2 : Any other pedigree constellations.

Obviously H_2 cannot be used in the current setting, in a simple way, and we need to refine possibly alternative hypotheses. a), b) and c) involve the use of `Familias`, however you may also skip to d) for `FamLinkX`.

a) * Open the file **Exercise4_15.fam** in `Familias 3`. We may assume that all females are children with no children of their own. Specify that the three females are children. Also define two parents, a mother and a father and specify that they are both born 1970.

b) * Use `Familias` (Generate pedigrees) to find the pedigrees with a posterior probability above 0.001. Which are the most probable relationships according to the results? Interpret the results.

c) * Discuss the constraints specified in a) and their impact on the results in b)

d) Open the database **Exercise4_15.sav** in `FamLinkX`, which contains frequency and haplotype data for the Argus X12 kit from QIAGEN based on a Swedish population sample. Explore the haplotype frequency database.

e) Based on the results in b), we specify the hypotheses

H_1 : The three females are all full siblings

H_2 : Two females are full siblings and the third (named F3) is a paternal half sibling

H_3 : Two females are full siblings and the third (named F3) is a maternal half sibling

f) Import the DNA data, available in **Exerecise4_15.txt**. Make sure to import the data in the file to the correct corresponding persons, the person denoted F3 should be imported to 3.

g) Calculate the LR and interpret the results. Be patient, the computation may require some time >20 min. *Hint*: To speed the computations up, go into `File->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.

h) Discuss if the LR in g) may be combined with the results in b)? What is your final conclusion on the case?

Exercise 4.16 (Further discussion of λ)**.**

We will provide an example of how the value of λ may crucial to the conclusion in a case. We use anonymized data from a real case with two typed females (F1 and F2) and consider hypotheses

H_1 : F1 and F2 are paternal half siblings, with different mothers.

H_2 : The two females are unrelated.

a) Open the file **Exercise4_16.sav**, containing the frequency database.

b) Select appropriate pedigrees and import genotype data from the file **Exercise4_16.txt**.

c) Compute the LR for a number of different values on λ, e.g., 0.001, 1, 100 and 1000.

d) Discuss the results in c). What conclusion can be drawn?

e) * Explore the genotype data and see if you can find an answer to the results. Use your knowledge about haplotype phases under the different hypotheses. *Hint*: Use the frequency estimation tool in the `Edit cluster` dialog.

f) Discuss what value on λ should be chosen. What is most conservative?