

Solutions to exercises using Familias 3, Copenhagen 2014

Exercise 1, The new mutation model and its implications

- a) See table below.

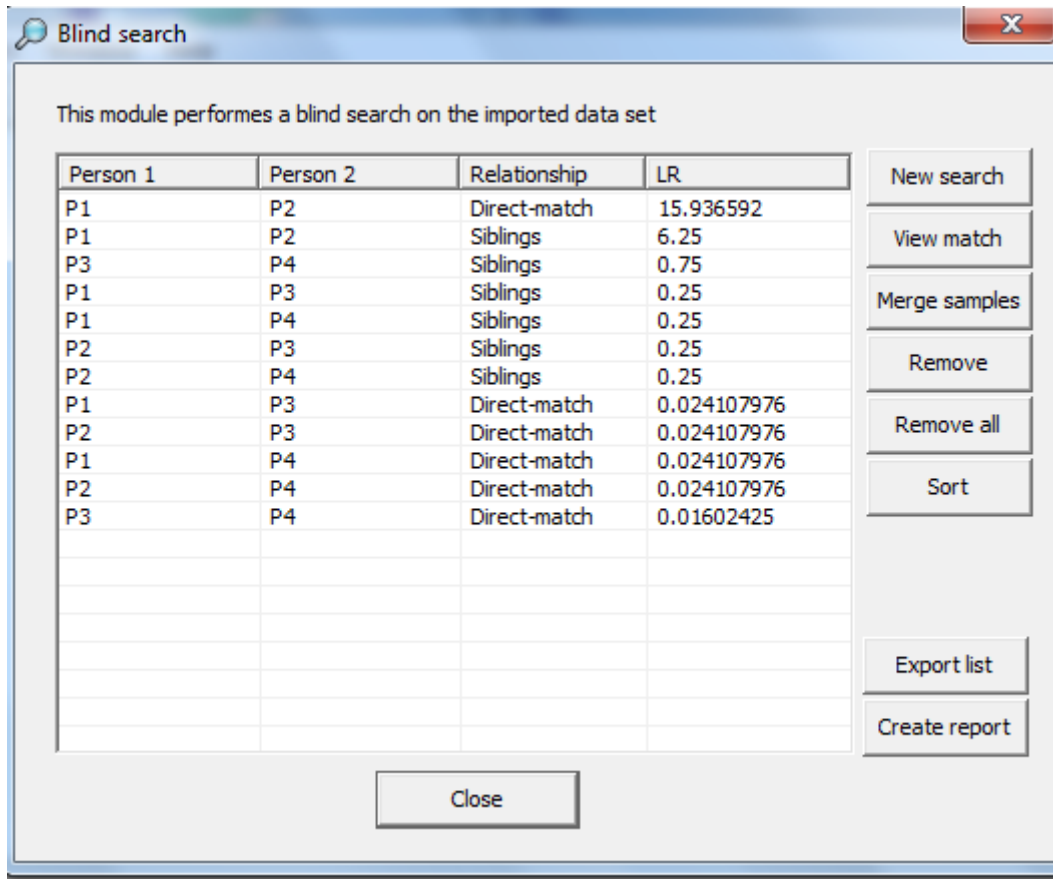
	9	9.3	10	12	15
9	0.99899	1.00E-07	0.001	1.00E-05	1.00E-08
9.3	2.50E-06	0.99999	2.50E-06	2.50E-06	2.50E-06
10	0.000918	9.18E-08	0.99899	9.18E-05	9.18E-08
12	8.41E-05	8.41E-07	0.000841	0.99899	8.41E-05
15	9.89E-07	9.89E-06	9.89E-06	0.000989	0.99899

- b) LR=0.00236966...
- c) LR=0.004168747...
- d) The result is different due to the fact that “old” implementation of the stepwise model in Familias do not look at actual step differences, but rather their location in the allele list. Also, old Familias do not consider microvariants as intermediate alleles.
- e) LR=4.545E-7. (It may for obvious reasons be interesting to consider silent alleles)
- f) -
- g) *Discuss stationary versus non-stationary mutation models.
- One large drawback with the stationary models is that we may obtain a different mutation rate than what we specify due to the matrix operations during stabilization.
 - We obtain different likelihoods if we add unrelated persons. Also the mutation matrix is not stable in multigenerational pedigrees.

Exercise 2, Using the Blind search interface

The blind search interface is, as the name suggests, a tool to blindly search for predefined pairwise relationships in a data set. The interface currently allows the user to search for parent-child, sibling, half-sibling relations as well as direct matches. It may be used in connection with the DVI module, see Exercise 4, to search a set of unidentified remains for direct matches or relationship within the data set. Another application may be to investigate a data set for unspecified relations before conducting a medical study or prior to creating a frequency database. The computations are relatively swift and may be used to search large data set for relations. We will test the module on a smaller data set.

- a) -
- b) -
- c) Results should look like the picture below (the matches <1.0 may not be present if the match limit is set to 1.0)



- d) P1 and P2 seems to be related, possibly a Direct-match, i.e. they are the same person, or they may also be siblings. Of course, this is a very simplified exercise using only one system. With several STR markers, two siblings will most probably not appear as a Direct-match due to the fact that they will most likely share no alleles (or only one) in some systems.
- e) -

Exercise 3, The simulation interface and its applications

The simulation interface has many uses and we will explore one of them. The interface uses the frequency database in Familias to simulate the defined pedigrees and calculates likelihood ratio distribution for the given case. Founder alleles are sampled using allele frequencies. Non-founder

alleles are subsequently sampled using the specified mutation model. We may define which persons are genotyped and those who are not.

Your lab has now received a request to whether you can reach a conclusion in a given case. We have,

- H1: The father of woman is also the alleged father of her child.
- H2: Another man, unrelated to the alleged father is the true father of the child.

The alleged father is dead and we may only receive samples from the mother and the child. In both cases the alleged father is the undisputed father of 'mother'.

- a) -
- b) We need only define the alleged father, the mother and the child
- c) -
- d) -
- e) Select only the mother and child as genotyped (Will be genotyped)
- f) -
- g) See picture below

Pedigree	Versus	Median	95%	5%
H1	H2 (TRUE)	0.002847	0.5371	1.849e-005
H1 (TRUE)	H2	842.6	6.248e+005	1.347

H2 (TRUE)	H1	351.2	5.355e+004	1.83
H2	H1 (TRUE)	0.001187	0.7414	1.582e-006

- h) By setting a LR limit of 100 we see that the false positive rate is 0% while the false negative rate is 30%. Going below the threshold of 100 is usually not advisable and increasing the threshold will likely result in a result which is inconclusive.
- i) We would probably need at least the markers in the frequency database to reach a conclusive result in the given case.

Exercise 4, A small accident – using the DVI module to identify the remains

In the last exercise we consider a small scale accident. Consider the crash of a small aeroplane with 10 passengers. We receive reference data from 5 different families.

- a) -
- b) -
 - a. -
 - b. It may be realistic that we do not have data for all missing persons as an accident may disperse the remains. Also we may start the Familias matching before all data is

collected and some materials may take longer time to extract, e.g. bones and teeth, than other, e.g. muscle tissue and blood. Also, we may have more unidentified remains than the number of actual missing persons. Possibly due to that several of the unidentified remains are one and the same person, or simply because we do not have a complete list of missing persons.

c) -

a. -

b. This is definitely a point of discussion. The priors may, to some extent, be changed using the Size option. If we use a large size > number of unidentified persons, a low prior weight will be given to each match, will using the Use list option is equal to using flat priors.

d) -

a. -

b. -

i. -

ii. -

iii. -

iv. -

c. -

i. -

ii. -

d. -

e) -

f) See table below. Not all unidentified remains were identified. We see that the match for PM1 has a relatively low LR (930) and may need further investigation.

DVI module - Results

Project name is: familias_exercise4

Family id	Unidentified person	Prior	Posterior	LR
Family 1	Reference pedigree	0.74194	0.00107386	1
Family 1	PM1 (Is father)	0.032258	0.998926	930.21659
Family 2	Reference pedigree	0.74194	9.51171e-007	1
Family 2	PM6 (Is Brother)	0.032258	>0.999999	1051334.1
Family 3	Reference pedigree	0.74194	1.48625e-005	1
Family 3	PM2 (Is mother)	0.032258	0.999985	67282.515
Family 4	Reference pedigree	0.74194	0.000787292	1
Family 4	PM3 (Is son)	0.032258	0.999213	1269.1764
Family 5	Reference pedigree	0.74194	3.23018e-008	1
Family 5	PM4 (Is brother)	0.032258	>0.999999	30958002

Search

Search

Sort

Limit list

Display

Match

View match

Confirm match

Remove

Save results

<- Previous

Close

DVI module - Results

Project name is: familias_exercise4_advanced

Family id	Unidentified person	Prior	Posterior	LR
Family 1	Reference pedigree	0.74194	9.2264e-005	1
Family 1	PM1 (Is father)	0.032258	0.0858255	930.21659
Family 1	PM1 (Is brother)	0.032258	0.000722744	7.8334347
Family 1	PM4 (Is brother)	0.032258	0.000221265	2.3981721
Family 1	PM5 (Is brother)	0.032258	0.913138	9897.019
Family 2	Reference pedigree	0.74194	9.51171e-007	1
Family 2	PM6 (Is Brother)	0.032258	>0.999999	1051334.1
Family 3	Reference pedigree	0.74194	1.48625e-005	1
Family 3	PM2 (Is mother)	0.032258	0.999985	67282.515
Family 4	Reference pedigree	0.74194	0.000787292	1
Family 4	PM3 (Is son)	0.032258	0.999213	1269.1764
Family 5	Reference pedigree	0.74194	3.23018e-008	1
Family 5	PM4 (Is brother)	0.032258	>0.999999	30958002

Search

Search

Sort

Limit list

Display

Match

View match

Confirm match

Remove

Save results

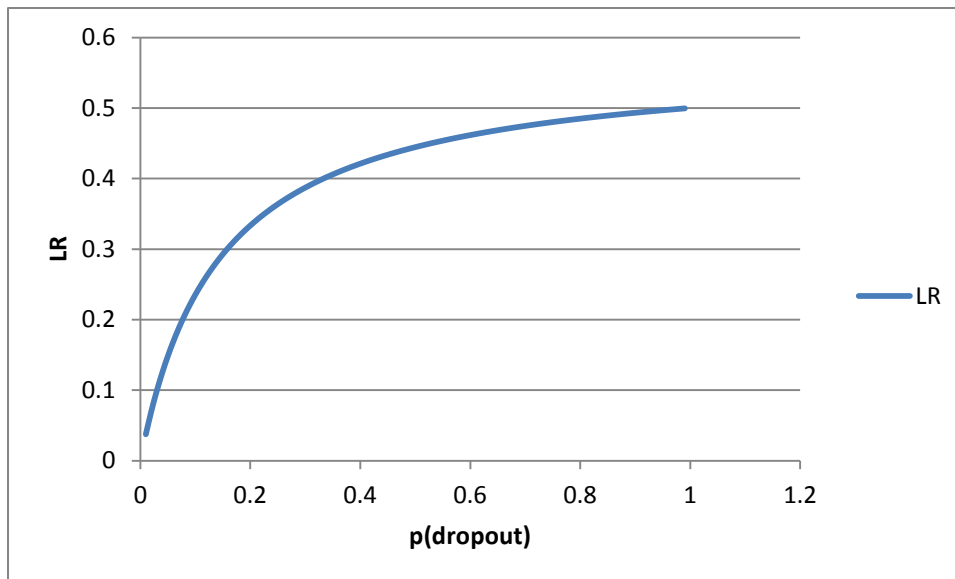
The two next exercises (5 and 6) do not have any suggested solution. Discuss with Daniel if you have any comments/questions.

Exercise 7, A case with low quality profiles (dropouts)

We consider a paternity case with a mother and her daughter and an alleged father,

- H1: The alleged father is the true father of child
- H2: Some other man, unrelated to the alleged father is the true father of the child

-
-
-
-
- The LR with $p(\text{dropout})=0.1$ should be 0.23529.
- See graph below for the LR as a function of the dropout probability



- g) **Now let us consider something more advanced. Go to Advanced setting and select Logistic dropouts (Use) and leave the regression parameters at their default values.
- a. Setup a series of dilutions where we are certain that dropout will occur. The concentration is less important as we are interested in the peak height for the remaining allele, i.e. the allele that has not dropped out. See also recommendations from the ISFG. Regression parameters are obtained through logistic regression, with the dropout probability as response and $\log(\text{peak height})$ as predictor. **Ask the teachers for further details**
 - b. -
 - c. $LR=0.1398$ with the peak height=100. If we decrease the peak height to something lower we will obtain an LR above 0.1398 as the probability of dropout will increase, see graph on e).
- h) Silent alleles are inherited throughout a pedigree whereas dropouts occur independently for different profiles. The latter meaning that the profile from the alleged father may be of poor quality and we may observe dropouts, while this has nothing to do with the quality of the profile for the child. A silent allele could not explain the data as the child is heterozygous and must have inherited the silent allele from the alleged father.

Exercise 8, Searching a database for relatives (Familial searching)

This exercise is meant to provide a brief overview of the new upcoming database searching feature in Familias 3.1. (The feature is still largely untested) We will consider a fictive database with >50.000 individuals, see Database8.txt. All individuals are genotyped for one STR marker S1. Again, use the same allele data as in Exercise 2.

- a) -
- b) -
- c) -
- d) -

could be siblings and we may wish to look further into these matches. For safety, one may wish to remove one of each from these two.

h) -

i) If you removed all the samples in g), i.e. samples 21-25, you should obtain output similar to the picture below.

j) -