This presentation will be subject to revision

Daniel Kling

2014-05-15

Familias 3 – Dropouts and other updates

Daniel Kling

Copenhagen May 20-23 2014

Contents

- Brief overview
- Repetition/Updates
  - DVI module
  - Blind search interface
  - Simulation interface
- Dropouts
- Familial searching (In brief)
- Create database feature (In brief)
- Other news

Familias 3 - Introduction

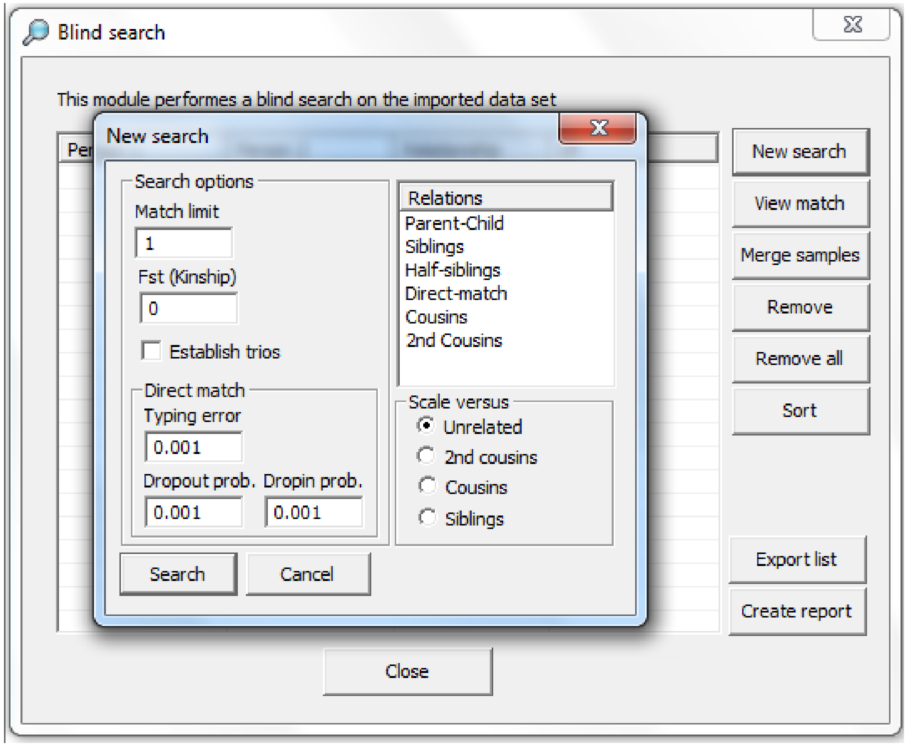
- Paper submitted
- Several bugs corrected
- Vast improvements
- New functions added
  - Log(LR) in simulations
  - Blind search options
  - View mutation matrix
  - Etc…

Familias 3 – Repetition

- DVI (Distaster Victim Identification)
  - What is a DVI operation?
  - Mass comparisons
  - New import option
    - CODIS xml-format
  - Faster!
  - New options on the priors
  - Export the result to text file (Open in Excel)

Familias 3 - Repetition

- Blind search
  - Search dataset for “unknown” relations
  - Detect relations among victims in DVI
  - Faster!



Blind search

This module performes a blind search on the imported data set

New search

Search options

Match limit: 1

Fst (Kinship): 0

☐ Establish trios

Direct match

Typing error: 0.001

Dropout prob.: 0.001

Dropin prob.: 0.001

Relations

Parent-Child
Siblings
Half-siblings
Direct-match
Cousins
2nd Cousins

Scale versus

● Unrelated
○ 2nd cousins
○ Cousins
○ Siblings

Familias 3 – Repetition

- Simulation
  - -What is a simulation?
    - -Based on a model
  - -Why do we want to simulate?
    - -Find distributions of LRs
    - -What can we expect
  - -The utility of simulations
    - -Assisting tool in decision prior to obtaining a case
    - -Assisting tool after a case is complete
  - -Further uses
    - -Use raw data to perform statistical calculations
    - -Use simulated data in other settings

Familias 3 – Repetition

- Simulation
  - Find good thresholds
    - False positive/negative rate
  - Investigate what we can expect
  - Include
    - Kinship
    - Mutations
    - Silent alleles
    - Multiple pedigrees
  - Investigate the number of persons we must genotype
  - Investigate the number of markers we must include

Familias 3 – News

Familias 3 – Direct matching

- Uses a “new” model presented by Kling et al.
- Two profiles, G1 and G2

H0: Profiles belong to one person

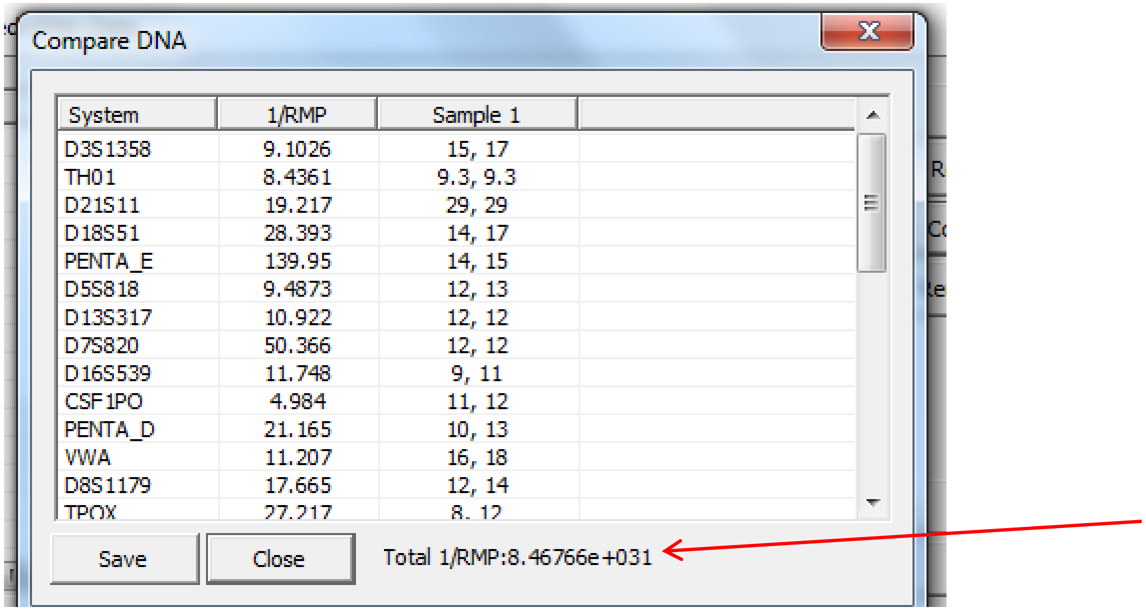
H1: Profiles belong to two persons

- Latent genotypes
- Conditional genotype probabilities

$$LR = \frac{P(G_1, G_2 \mid H_0)}{P(G_1, G_2 \mid H_1)} = \frac{\sum_{i=1}^{A} \sum_{j=i}^{A} P(G_{true,i,j}) P(G_1 \mid G_{true,i,j}) P(G_2 \mid G_{true,i,j})}{P(G_1) P(G_2)}$$

Familias 3 – Random match prob.

– Compare DNA data



System	1/RMP	Sample 1
D3S1358	9.1026	15, 17
TH01	8.4361	9.3, 9.3
D21S11	19.217	29, 29
D18S51	28.393	14, 17
PENTA_E	139.95	14, 15
D5S818	9.4873	12, 13
D13S317	10.922	12, 12
D7S820	50.366	12, 12
D16S539	11.748	9, 11
CSF1PO	4.984	11, 12
PENTA_D	21.165	10, 13
VWA	11.207	16, 18
D8S1179	17.665	12, 14
TPOX	27.217	8, 12

Total 1/RMP:8.46766e+031

Familias 3 – Import Genemapper file

- Export data from Genemapper
  - File -> Export table
  - Genotypes table

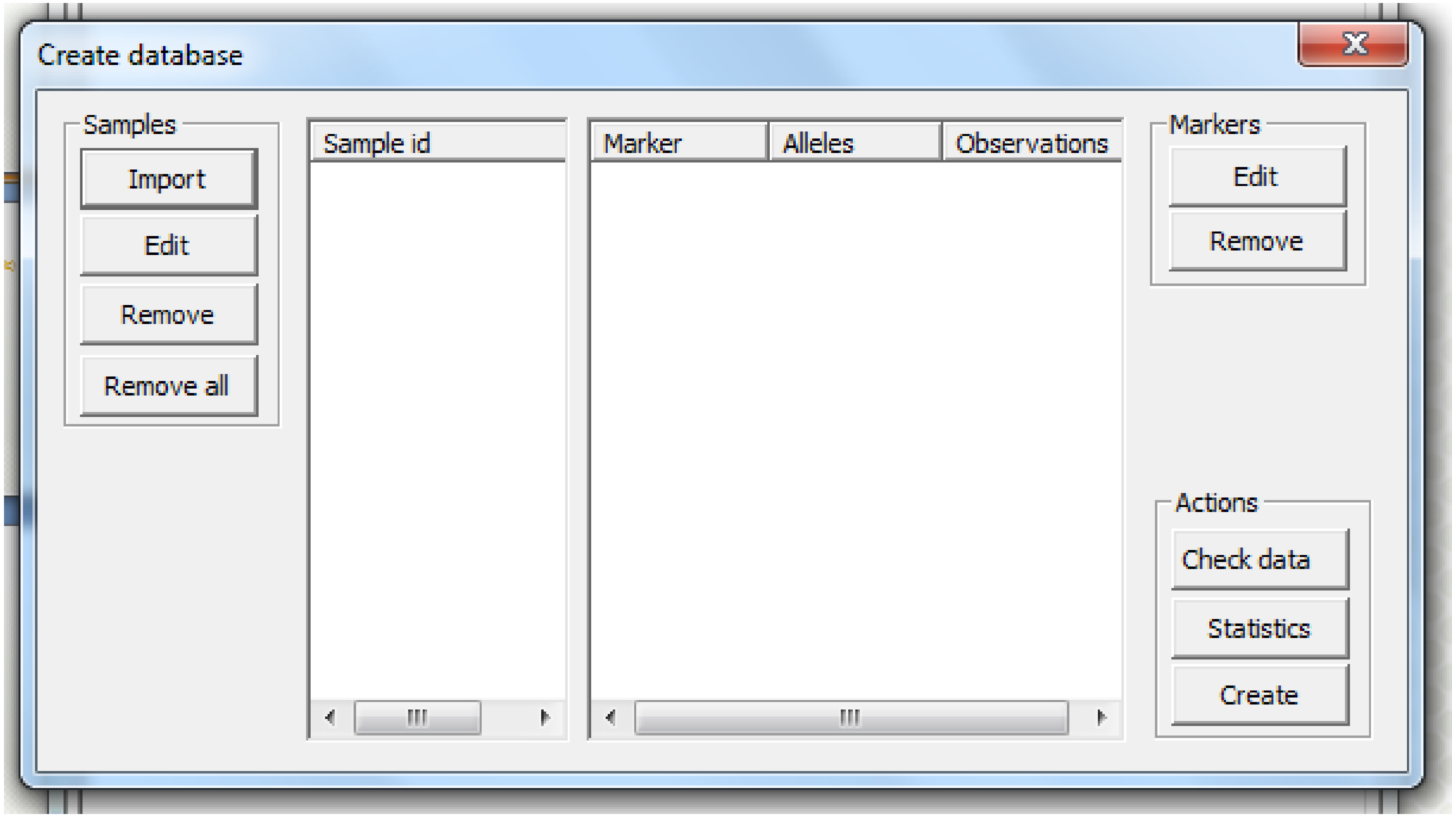
  Sample Name \t Marker \t Allele1 \t Allele2

Familias 3 – Familial searching (Beta)

- What is Familial searching?
    - Relative in database?
- Databases
    - Large searches
    - Spurious matches
- Mixtures
    - LR calculations
- Other uses
    - Find if the father of the unborn child of woman is in the criminal database

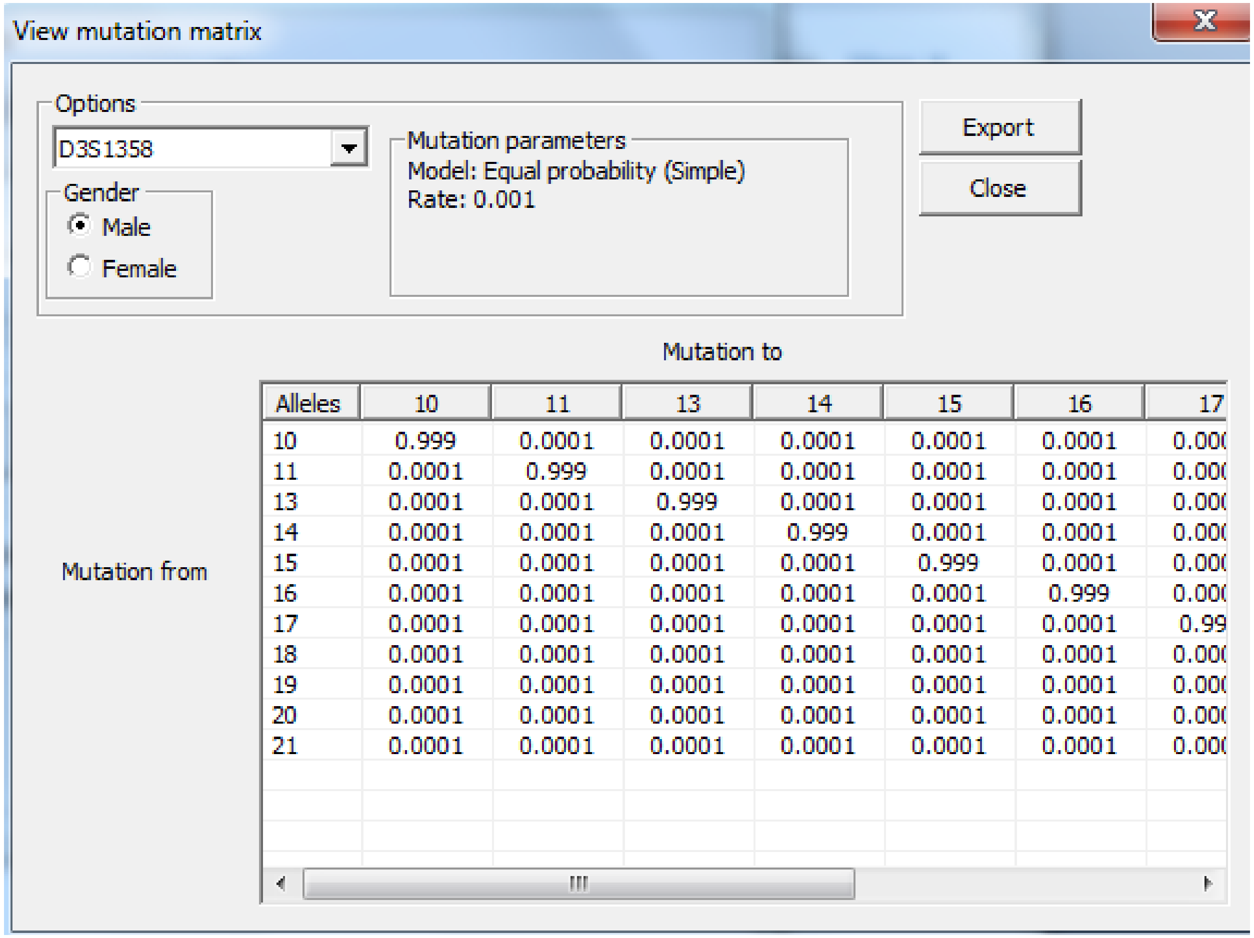
Familias 3 – Create database

- Use output directly from Genemapper



Familias 3 – View mutation matrix

- Advanced settings



Options: D3S1358

Gender:
● Male
○ Female

Mutation parameters
Model: Equal probability (Simple)
Rate: 0.001

Mutation to

Mutation from

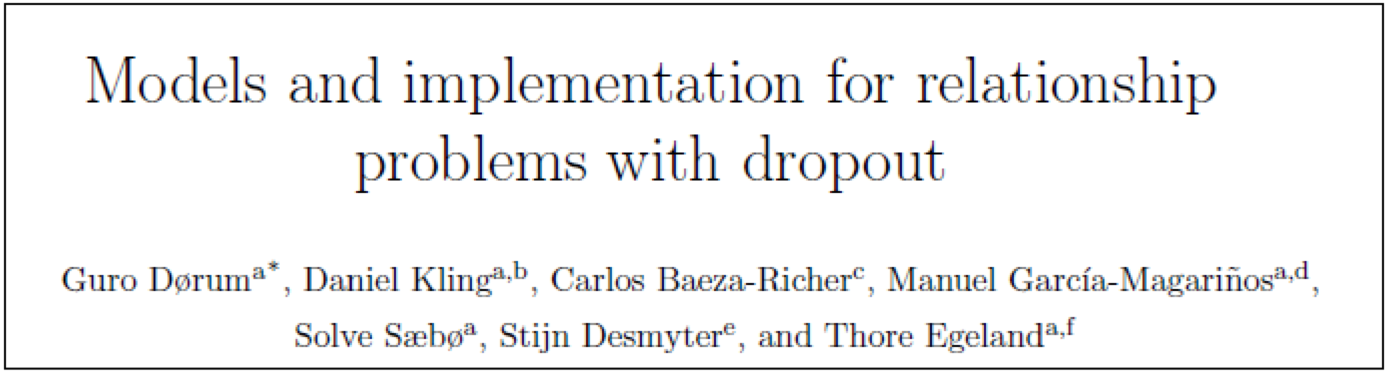
Alleles	10	11	13	14	15	16	17
10	0.999	0.0001	0.0001	0.0001	0.0001	0.0001	0.000
11	0.0001	0.999	0.0001	0.0001	0.0001	0.0001	0.000
13	0.0001	0.0001	0.999	0.0001	0.0001	0.0001	0.000
14	0.0001	0.0001	0.0001	0.999	0.0001	0.0001	0.000
15	0.0001	0.0001	0.0001	0.0001	0.999	0.0001	0.000
16	0.0001	0.0001	0.0001	0.0001	0.0001	0.999	0.000
17	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.99
18	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.000
19	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.000
20	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.000
21	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.000

Familias 3 - Dropouts

Familias 3 - Dropouts

- What is a dropout?
- The model
    - Conditional genotype probabilities
    - Dropout probability
- Low quality profiles
- Estimation of dropout probabilities
    - Logistic regression
- Implementation in Familias
    - LR for a number of dropout probabilities

Familias 3 - Dropouts



Models and implementation for relationship problems with dropout

Guro Dørum[a*], Daniel Kling[a,b], Carlos Baeza-Richer[c], Manuel García-Magariños[a,d],
Solve Sæbø[a], Stijn Desmyter[e], and Thore Egeland[a,f]

Familias 3 - Dropouts

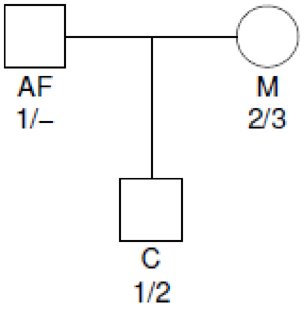


Figure 1: Pedigree in the motivational example.

The hypotheses tested are

H_1: AF is the father of C

H_2: Someone unrelated to AF is the father of C

LR without uncertainty: [P(1,1)P(2,3)0.5]/[P(1,1)P(2,3)p(1)0.5]=1/p(1)

Familias 3 - Dropouts

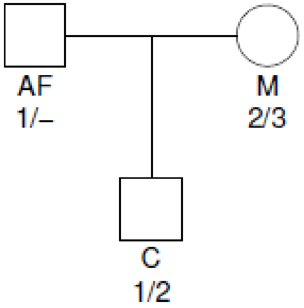


Figure 1: Pedigree in the motivational example.

The hypotheses tested are

H_1: AF is the father of C

H_2: Someone unrelated to AF is the father of C

LR with uncertainty:

$$LR = \frac{\sum_j P(G_C \mid G_M, G_{AF_j}) P(G^\star_{AF} \mid G_{AF_j}) P(G_{AF_j})}{P(G_C \mid G_M) \sum_j P(G^\star_{AF} \mid G_{AF_j}) P(G_{AF_j})}.$$

Familias 3 - Dropouts

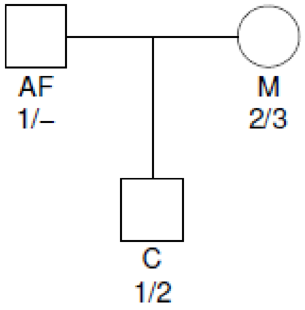


Figure 1: Pedigree in the motivational example.

The hypotheses tested are

H_1: AF is the father of C

H_2: Someone unrelated to AF is the father of C

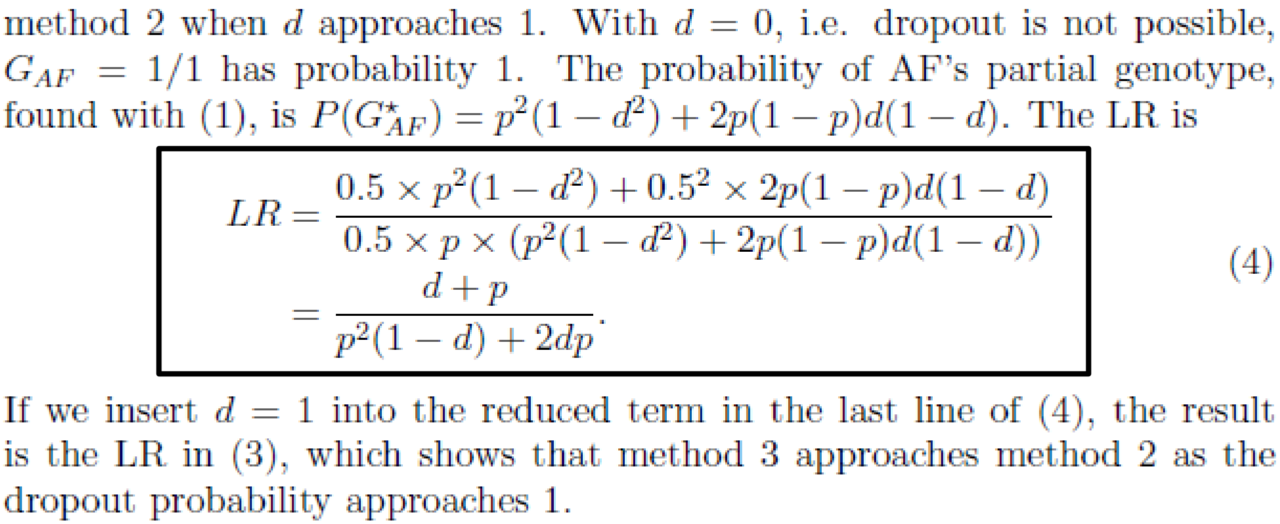
LR with uncertainty:

$$LR = \frac{\sum_j P(G_C \mid G_M, G_{AF_j}) \boxed{P(G^\star_{AF} \mid G_{AF_j})} P(G_{AF_j})}{P(G_C \mid G_M) \sum_j P(G^\star_{AF} \mid G_{AF_j}) P(G_{AF_j})}.$$

$$P(G^\star_{AF} = 1/- \mid G_{AF} = 1/1) = 1 - d^2$$
$$P(G^\star_{AF} = 1/- \mid G_{AF} = 1/Q) = d(1-d).$$

Familias 3 - Dropouts

LR with uncertainty:



method 2 when d approaches 1. With $d = 0$, i.e. dropout is not possible, $G_{AF} = 1/1$ has probability 1. The probability of AF's partial genotype, found with (1), is $P(G^\star_{AF}) = p^2(1-d^2) + 2p(1-p)d(1-d)$. The LR is

$$\boxed{\begin{aligned} LR &= \frac{0.5 \times p^2(1-d^2) + 0.5^2 \times 2p(1-p)d(1-d)}{0.5 \times p \times (p^2(1-d^2) + 2p(1-p)d(1-d))} \\ &= \frac{d+p}{p^2(1-d) + 2dp}. \end{aligned}} \tag{4}$$

If we insert $d = 1$ into the reduced term in the last line of (4), the result is the LR in (3), which shows that method 3 approaches method 2 as the dropout probability approaches 1.

Familias 3 - Dropouts

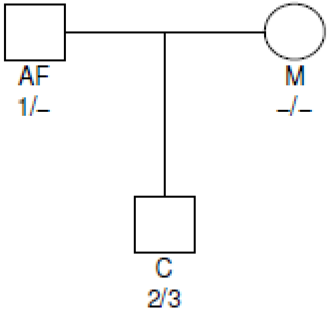
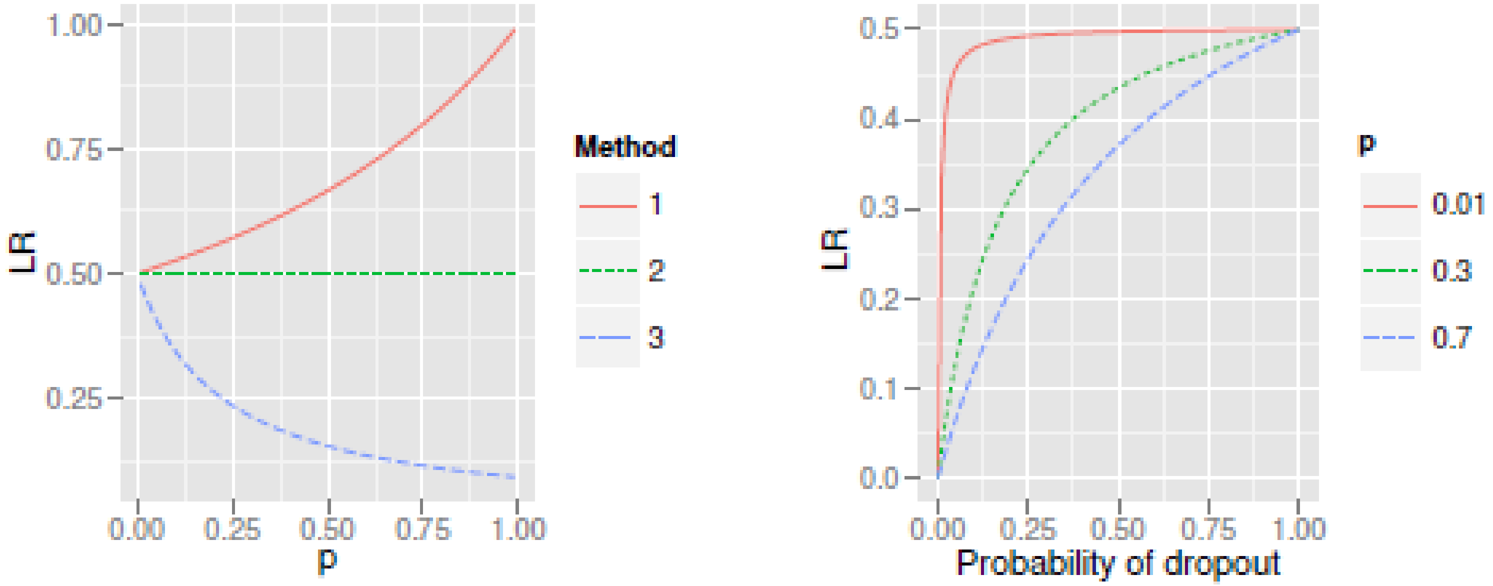


Figure 4: Pedigree in example 2.

Familias 3 - Dropouts

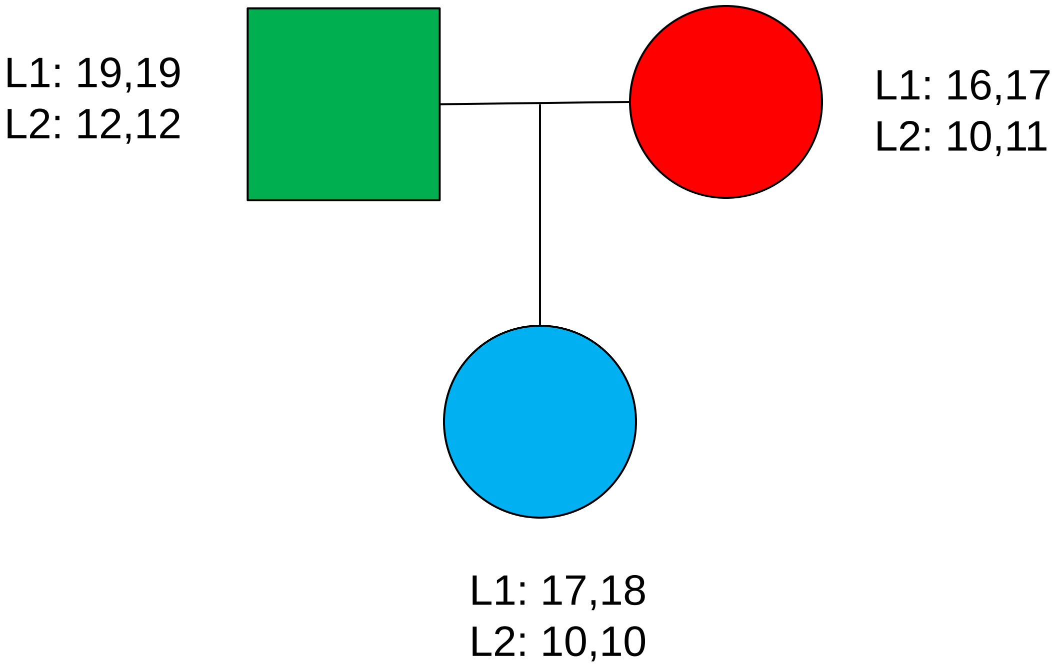


(a) Comparison of dropout methods (b) Method 3: LR vs dropout

Familias 3 - Dropouts

- Dropout versus silent allele

Example



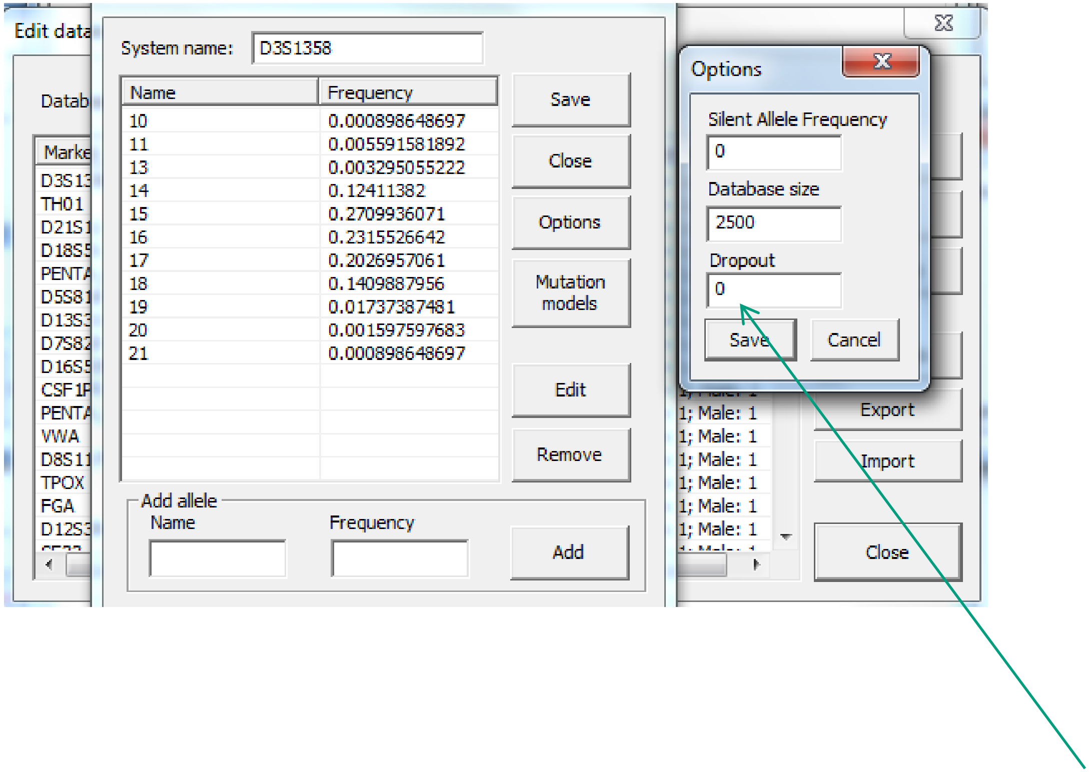
L1: 19,19
L2: 12,12

L1: 16,17
L2: 10,11

L1: 17,18
L2: 10,10

Familias 3 - Dropouts

– Define dropout probabilities for each system



System name: D3S1358

Name	Frequency
10	0.000898648697
11	0.005591581892
13	0.003295055222
14	0.12411382
15	0.2709936071
16	0.2315526642
17	0.2026957061
18	0.1409887956
19	0.01737387481
20	0.001597597683
21	0.000898648697

Options

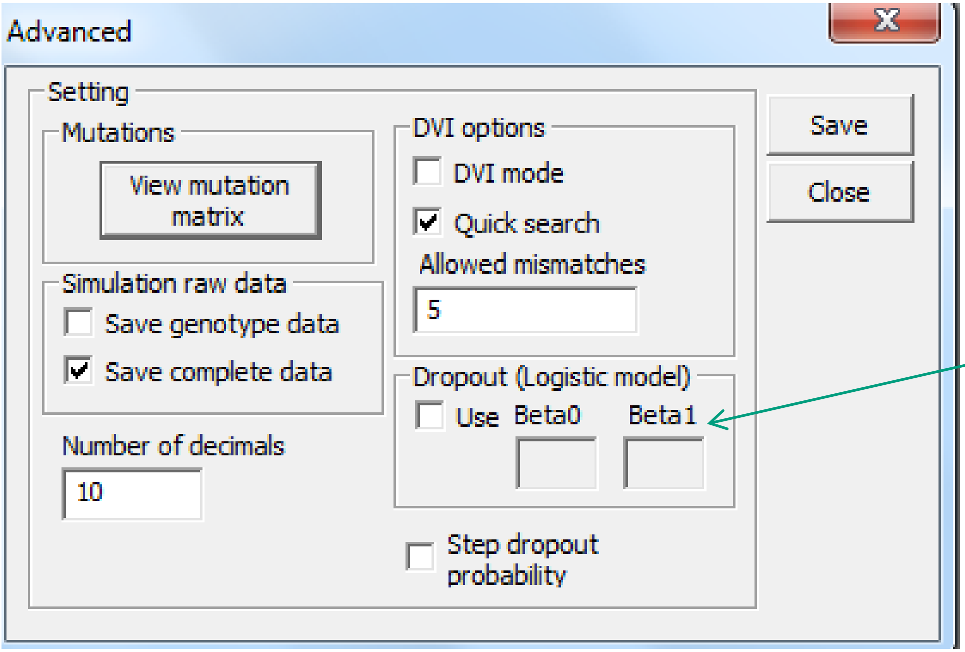
Silent Allele Frequency: 0

Database size: 2500

Dropout: 0

Familias 3 - Dropouts

- Define the use of logistic regression model



Setting

Mutations

Simulation raw data

☐ Save genotype data

☑ Save complete data

Number of decimals: 10

DVI options

☐ DVI mode

☑ Quick search

Allowed mismatches: 5

Dropout (Logistic model)

☐ Use Beta0 Beta1

☐ Step dropout probability

$$logit(d) = \log\left(\frac{d}{1-d}\right) = \beta_0 - \beta_1 \log(H)$$

Exercises Day 2

- Exercises in Familias 3.1.3
  - Do exercise 15-17
    - Dropouts
  - Complete the exercises in the Familias 3 – exercises
    - New: 7, 8 and 9