PRELIMINARY VERSION, THE PRESENTATION WILL BE SUBJECT TO CHANGES

Linked markers

Daniel Kling
Norwegian Institute of Public Health *and* Norwegian University of Life Sciences
daniel.kling@fhi.no

Copenhagen, Oct 9, 2013

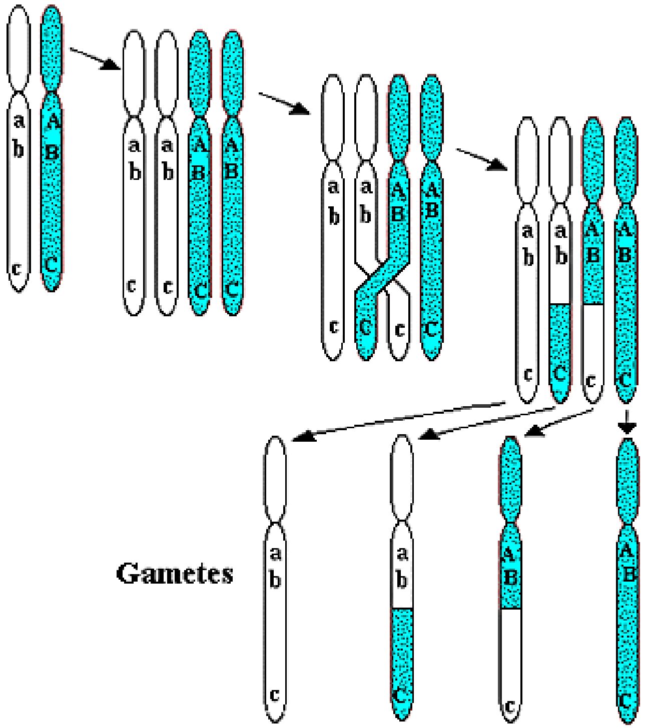
Outline

- Linkage explained
- Linkage disequilibrium explained
- FamLink
- FamLinkX

Linkage

- What is linkage?

ation



Crossing-over and recombination during meiosis

http://www.accessexcellence.org/RC/VL/GG/comeiosis.php

Linkage

- Worked example

Linkage disequilibrium

- What is linkage disequilibrium
  - Allelic association
  - Two alleles (at two different markers) which is observed more often/less often than can be expected.
  - Effects the allele probabilities not the transmission probabilities.

Example

Marker1 (vWa): Alleles 13 and 14, frequencies 0.2 and 0.8

Marker2 (D12S391): Alleles 16 and 17, frequencies 0.4 and 0.6

*Expected frequency of [13, 16] is 0.2*0.4 = 0.08*

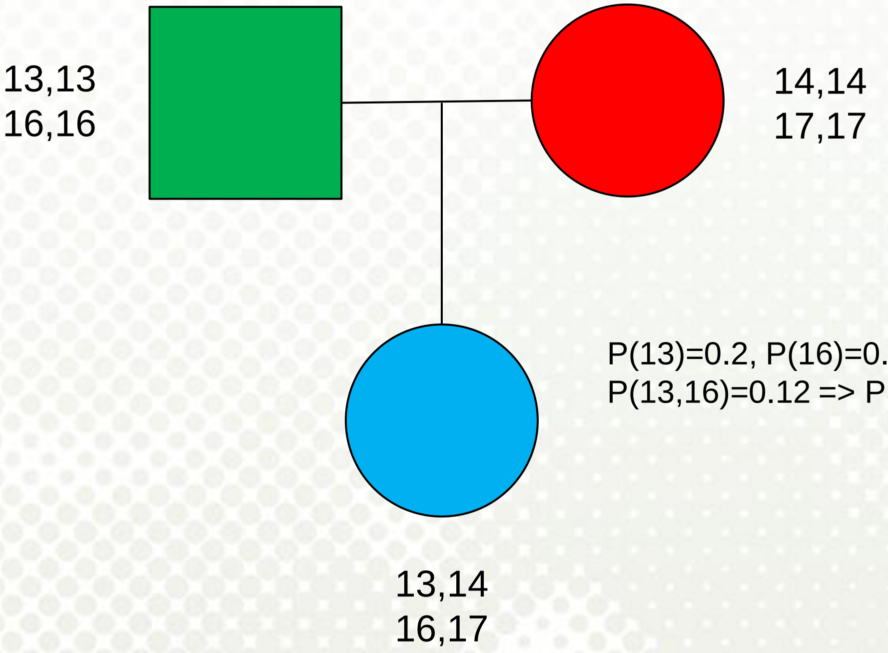
Observed frequency of [13, 16] is 0.12

$$r_{ij}^2 = \frac{\left[p_{ij} - p_i \times p_j\right]^2}{p_i(1-p_i)p_j(1-p_j)} \qquad \left\{\begin{matrix} i = 13,14 \\ j = 16,17 \end{matrix}\right\}$$

$$p_{13,16} = 0.12,\ p_{13} = 0.2,\ p_{16} = 0.4 \Rightarrow r_{13,16}^2 = \frac{(0.12-0.08)^2}{0.2(1-0.2)0.4(1-0.4)} \simeq 0.042$$

Linkage disequilibrium

- Worked example, paternity with two markers



13,13
16,16

14,14
17,17

13,14
16,17

P(13)=0.2, P(16)=0.4 (Linkage equilibrium)
P(13,16)=0.12 => P(16|13)=0.6 (Linkage disequilibrium)

LR1 = 1/P(13)*1/P(16)=12.5 (Linkage equilibrium)

LR2 = 1/P(13)*1/P(16|13)=8.33 (Linkage disequilibrium)

Summary

Linkage	Linkage disequilibrium
Dependency between neighbouring markers	Dependency between alleles at different loci
Observed within a pedigree	Observed in a population
Extends long distances >10 cM	Usually extends short distances <1 cM
Do not affect random match probability (unless related)	Affect random match probabilities
Take into account for extended pedigrees	Always take into account for all pedigrees
Always take into account if also LD is present, for all pedigrees	Measured by the deviation from expectations, decays with recombinations
Measured by the recombination rate, constant	Used to find alleles associated with a disease, in the population
Used to find markers linked to a disease, in families	

Recent example (Linkage)

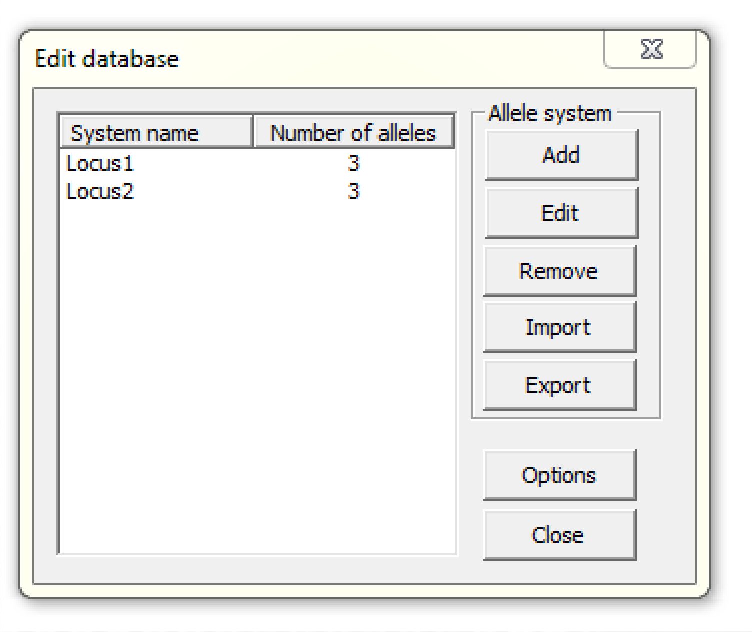
- Several publications suggest linkage should be accounted for between vWA and D12S391
- vWA and D12S391 is located only 6.3 Mb apart on chromosome 12
- Other STR markers may as well be linked, e.g. SE33 and D6S1043
- Following the discussion in the ISFG meeting in Vienna 2011, a new software to handle linked markers, was developed.

FamLink

- What can FamLink do?
  - User-friendly software
  - Similar to Familias in many ways
  - Perform likelihood calculations on a number of predefined pedigrees
  - Takes linkage between markers into account
  - Does not implement a good mutation model
  - Simulation interface
  - Implements the Lander-Green algorithm
    - Different than Familias
    - Merlin

FamLink – At a glance

- Define allele frequency database

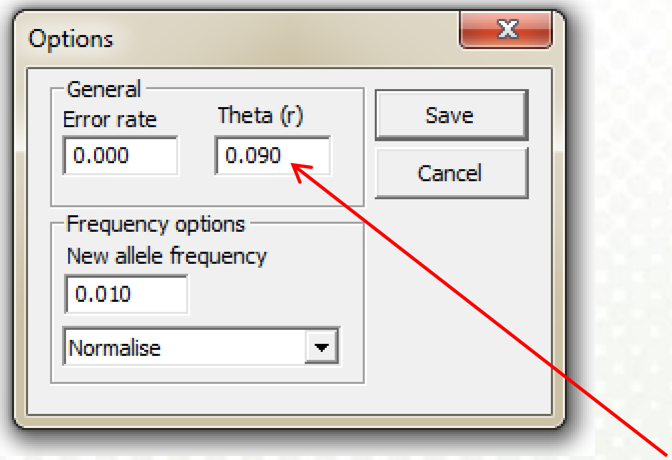


System name	Number of alleles
Locus1	3
Locus2	3

- Currently restricted to two markers

FamLink – At a glance

- Define recombination rate



Error rate: 0.000

Theta (r): 0.090

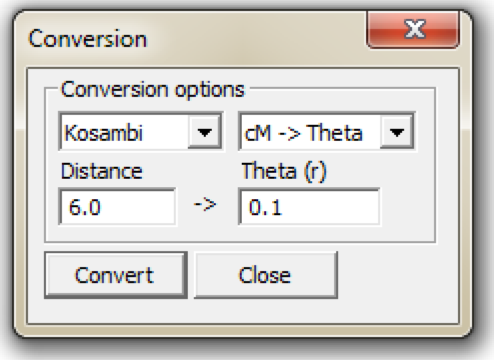
New allele frequency: 0.010

Normalise

- Between 0 and 0.5

FamLink – At a glance

- Convert genetic distance (cM) to recombination rate



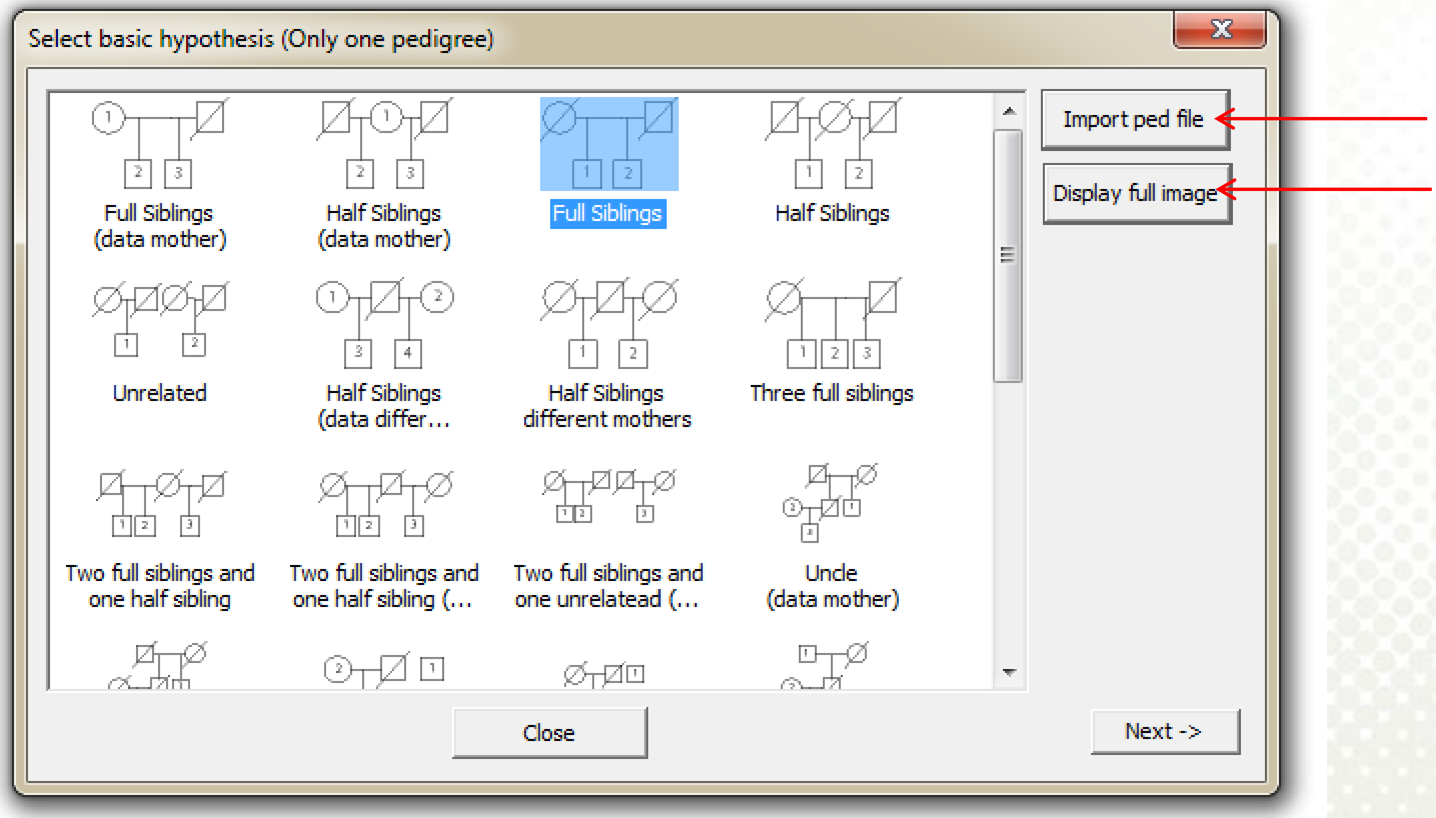
Kosambi

cM -> Theta

Distance: 6.0 -> Theta (r): 0.1

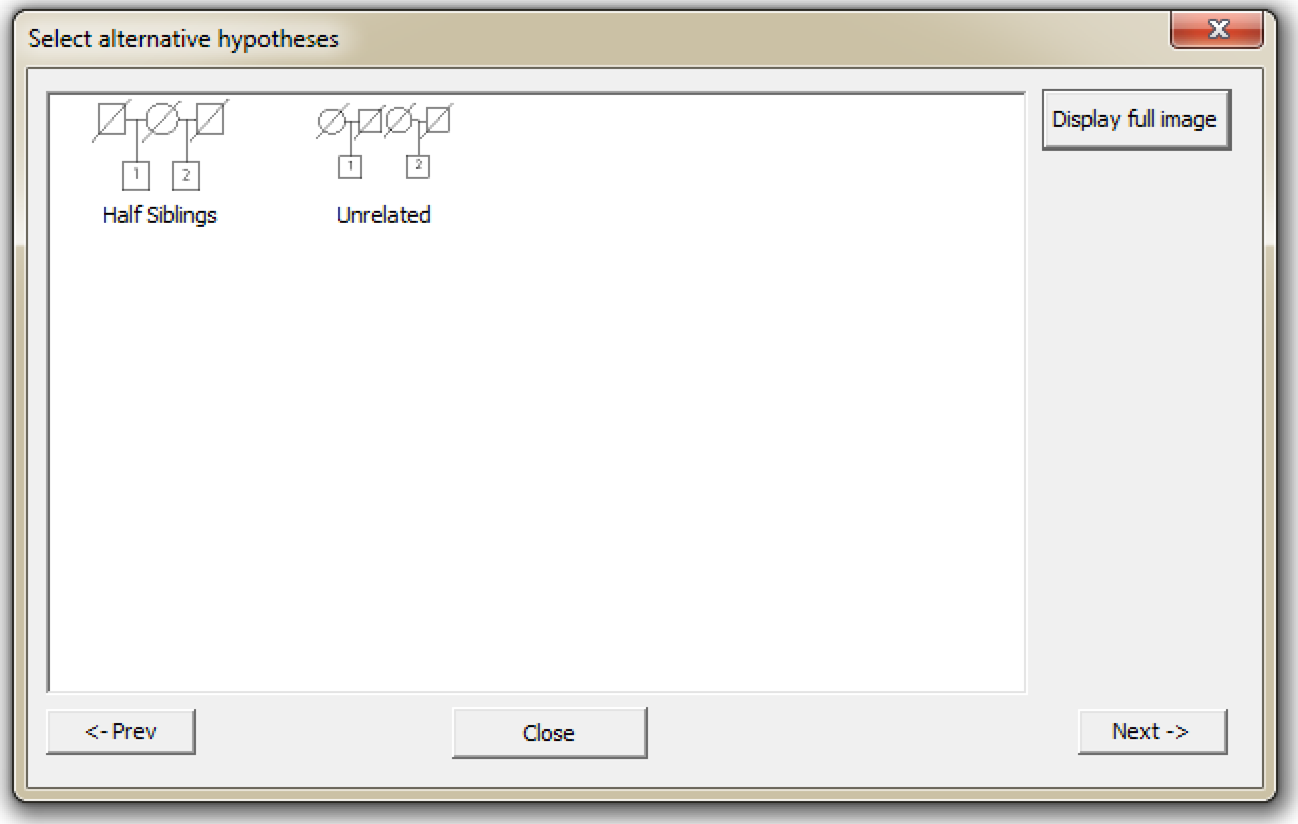
FamLink – At a glance

- Select main hypothesis



FamLink – At a glance

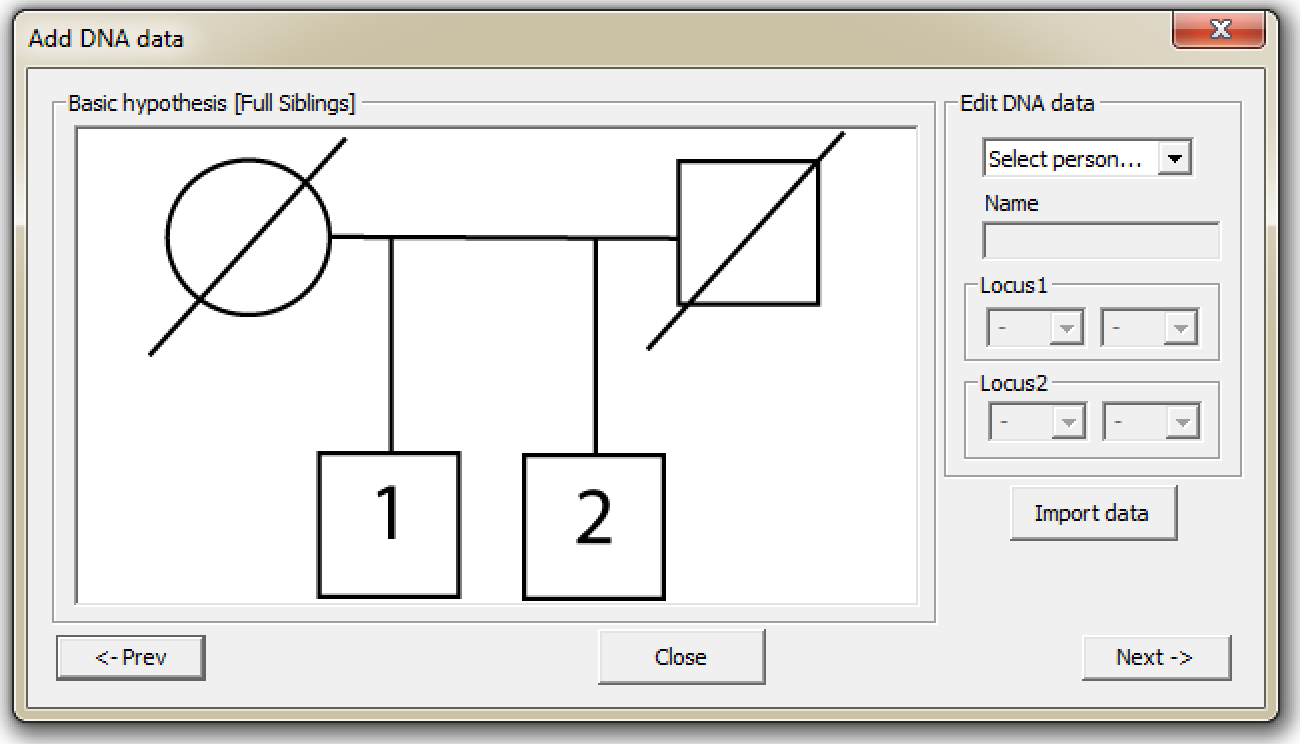
- Select alternative hypotheses



Select alternative hypotheses

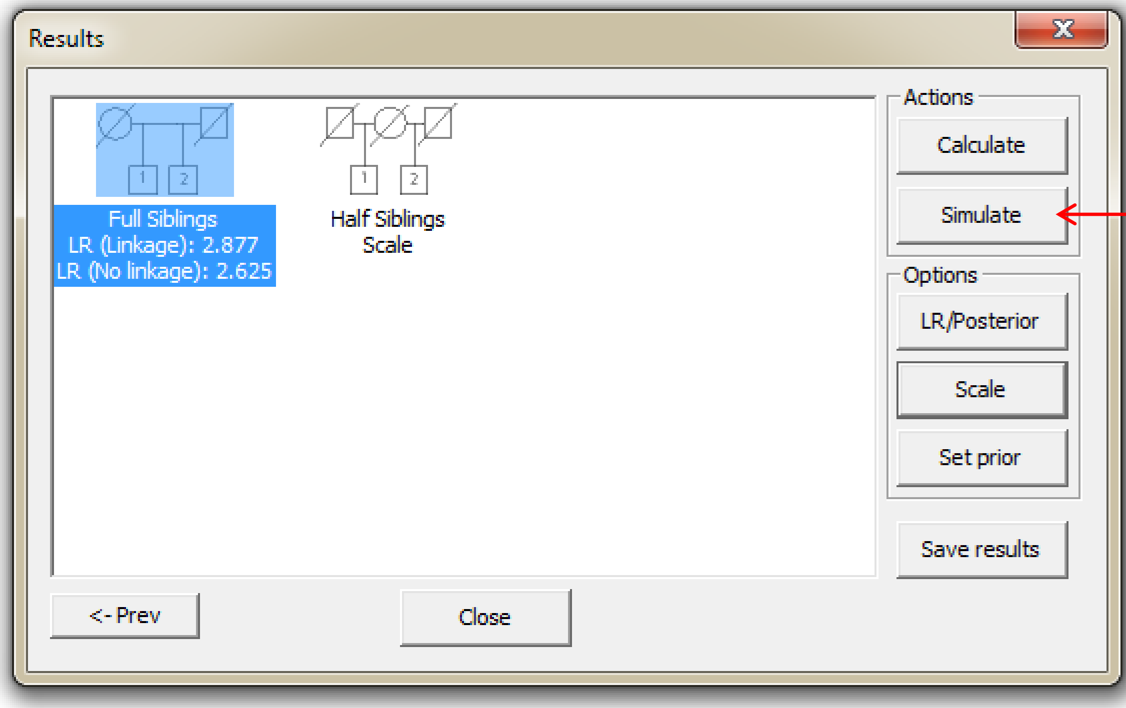
FamLink – At a glance

- Define DNA data



FamLink – At a glance

- Calculate likelihoods



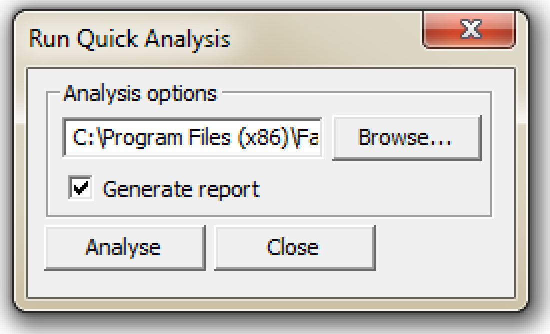
Full Siblings
LR (Linkage): 2.877
LR (No linkage): 2.625

Half Siblings
Scale

- We display two methods

FamLink – Familias connection

- Calculate likelihoods for a Familias project



Run Quick Analysis

Analysis options

C:\Program Files (x86)\Fa

☑ Generate report

- Get likelihoods for all your markers
  - Linkage considered for all markers

FamLink – Questions?

BREAK

X-chromosomal markers

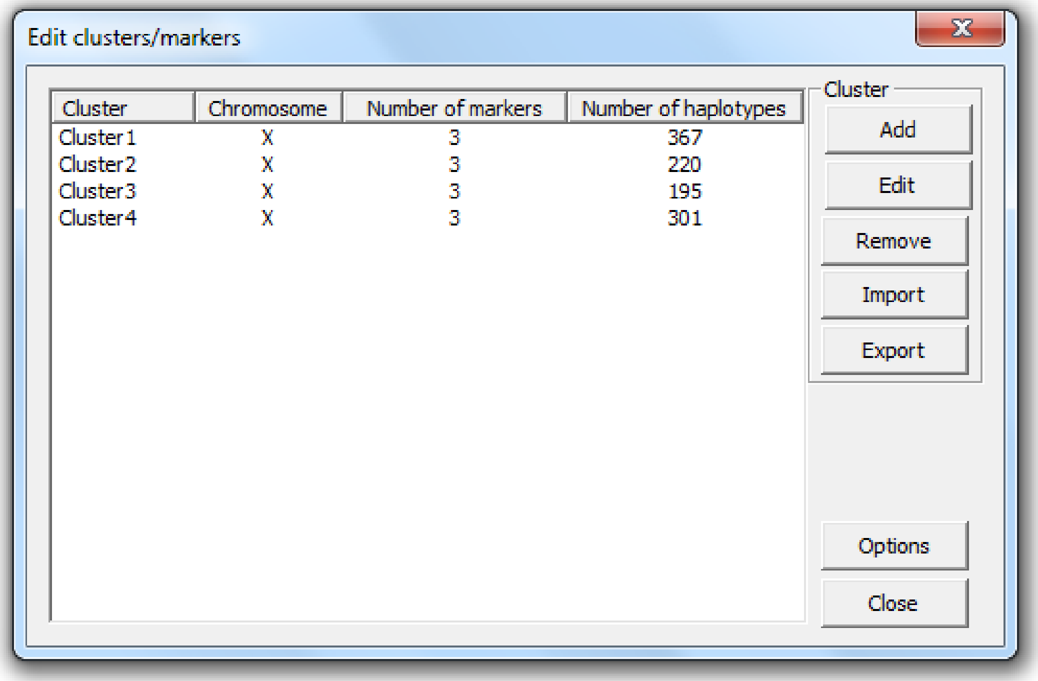
- Used where autosomal markers fail
- Argus X12
  - (4 clusters with three tightly linked markers)
- Linkage
- Linkage disequilibrium
- Mutations
- FamLinkX!
  - New joint probability model
  - Released autumn 2013

FamLinkX

- Markov chain to handle linkage
  - Similar to Lander-Green
- Multistep Markov chain to handle LD
- Uses a Dirichlet distribution to estimate haplotype frequencies
- Formula

FamLinkX – At a glance

- Define clusters of markers

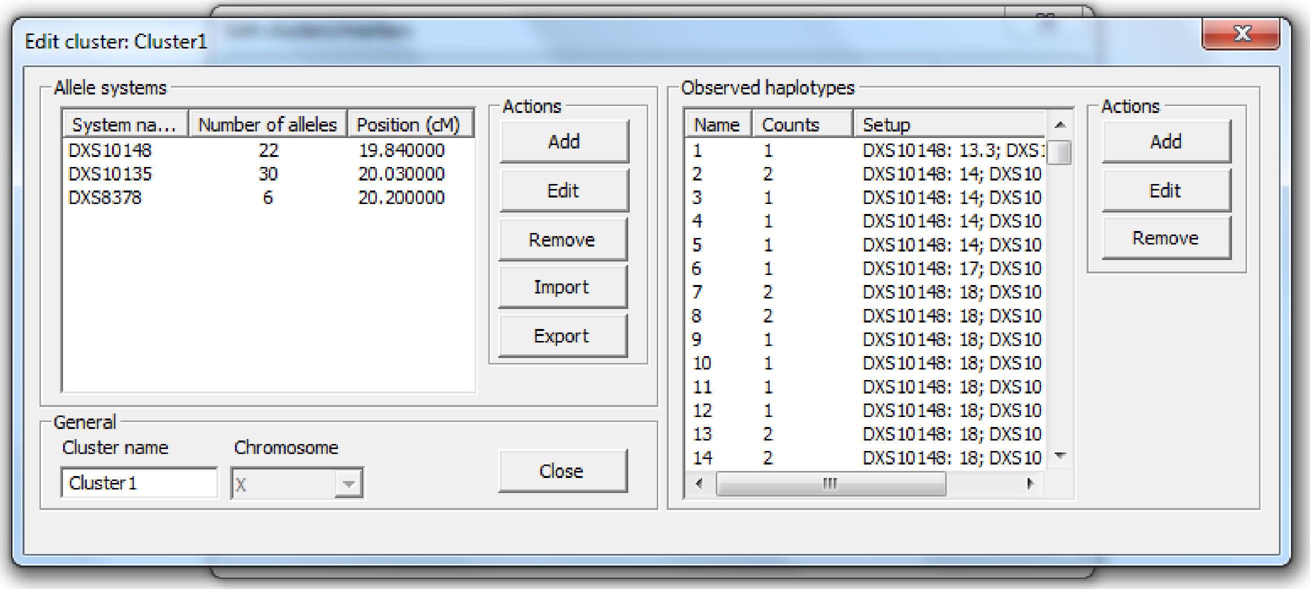


Cluster	Chromosome	Number of markers	Number of haplotypes
Cluster1	X	3	367
Cluster2	X	3	220
Cluster3	X	3	195
Cluster4	X	3	301

- Account for linkage between clusters
- Account for linkage and LD within each cluster

FamLinkX – At a glance

- Add markers
  - Define genetic position
  - Mutation parameters



Edit cluster: Cluster1

Allele systems

System na...	Number of alleles	Position (cM)
DXS10148	22	19.840000
DXS10135	30	20.030000
DXS8378	6	20.200000

General

Cluster name: Cluster1

Chromosome: X

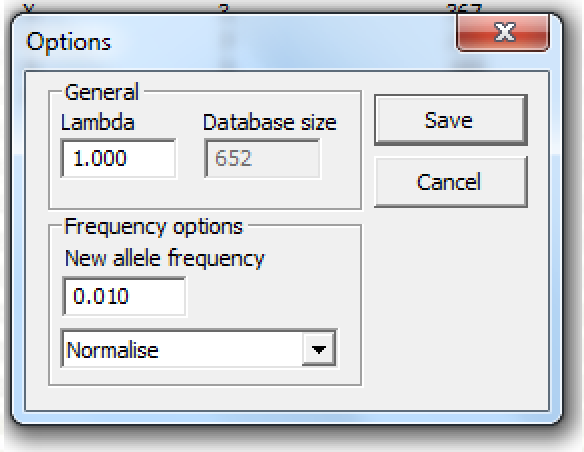
Observed haplotypes

Name	Counts	Setup
1	1	DXS10148: 13.3; DXS
2	2	DXS10148: 14; DXS10
3	1	DXS10148: 14; DXS10
4	1	DXS10148: 14; DXS10
5	1	DXS10148: 14; DXS10
6	1	DXS10148: 17; DXS10
7	2	DXS10148: 18; DXS10
8	2	DXS10148: 18; DXS10
9	1	DXS10148: 18; DXS10
10	1	DXS10148: 18; DXS10
11	1	DXS10148: 18; DXS10
12	1	DXS10148: 18; DXS10
13	2	DXS10148: 18; DXS10
14	2	DXS10148: 18; DXS10

- Add haplotypes(!)
  - Define setup
  - Counts

FamLink – At a glance

- Selecting value for Lambda



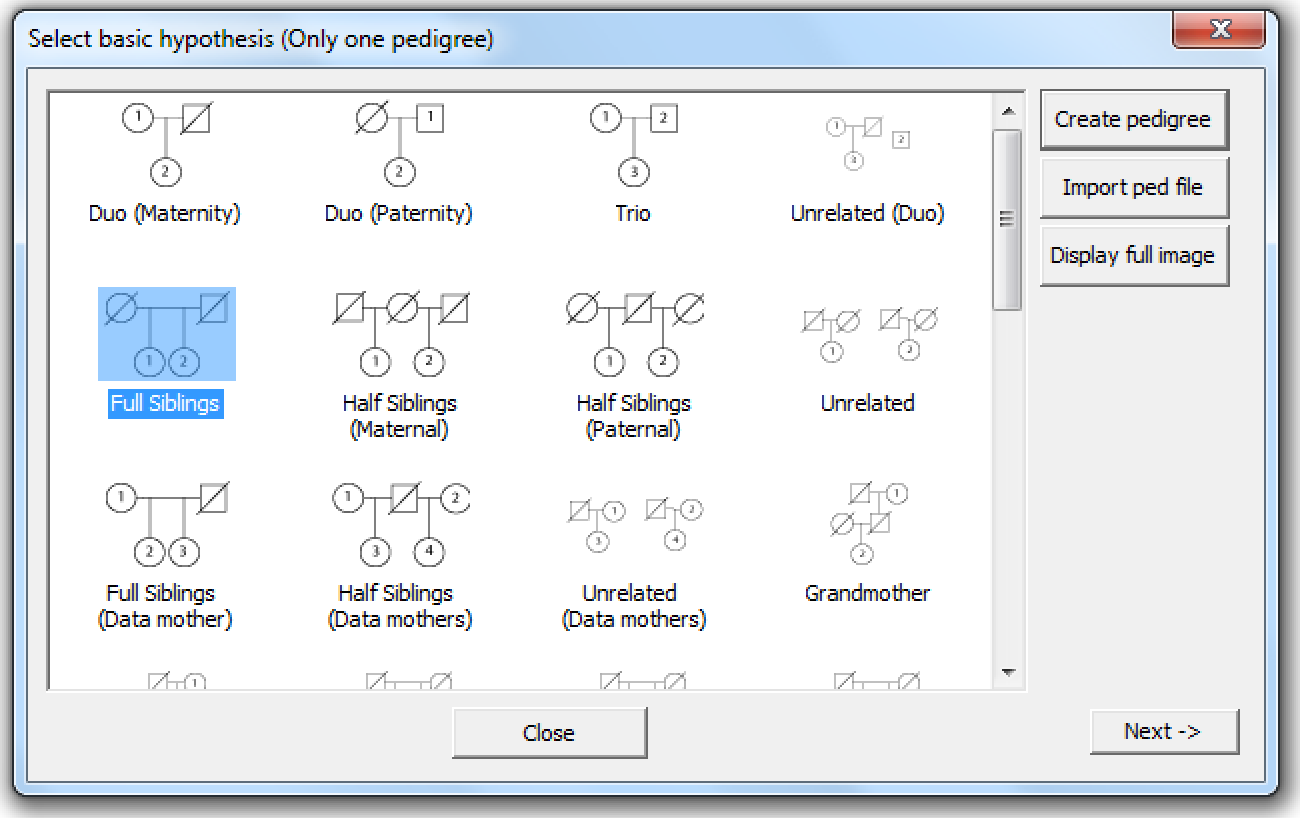
General
Lambda: 1.000
Database size: 652

Frequency options
New allele frequency: 0.010
Normalise

- We display two methods

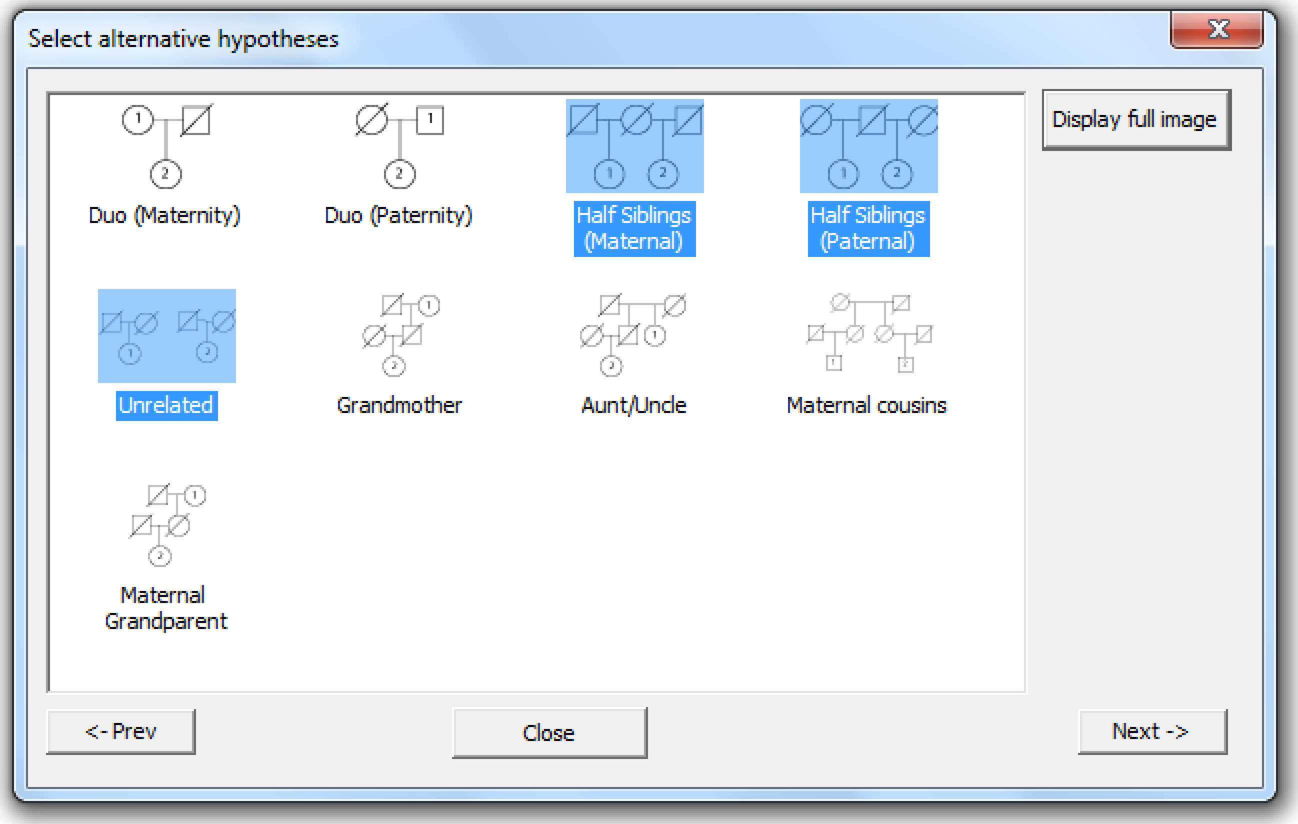
FamLinkX – At a glance

- Select main hypothesis



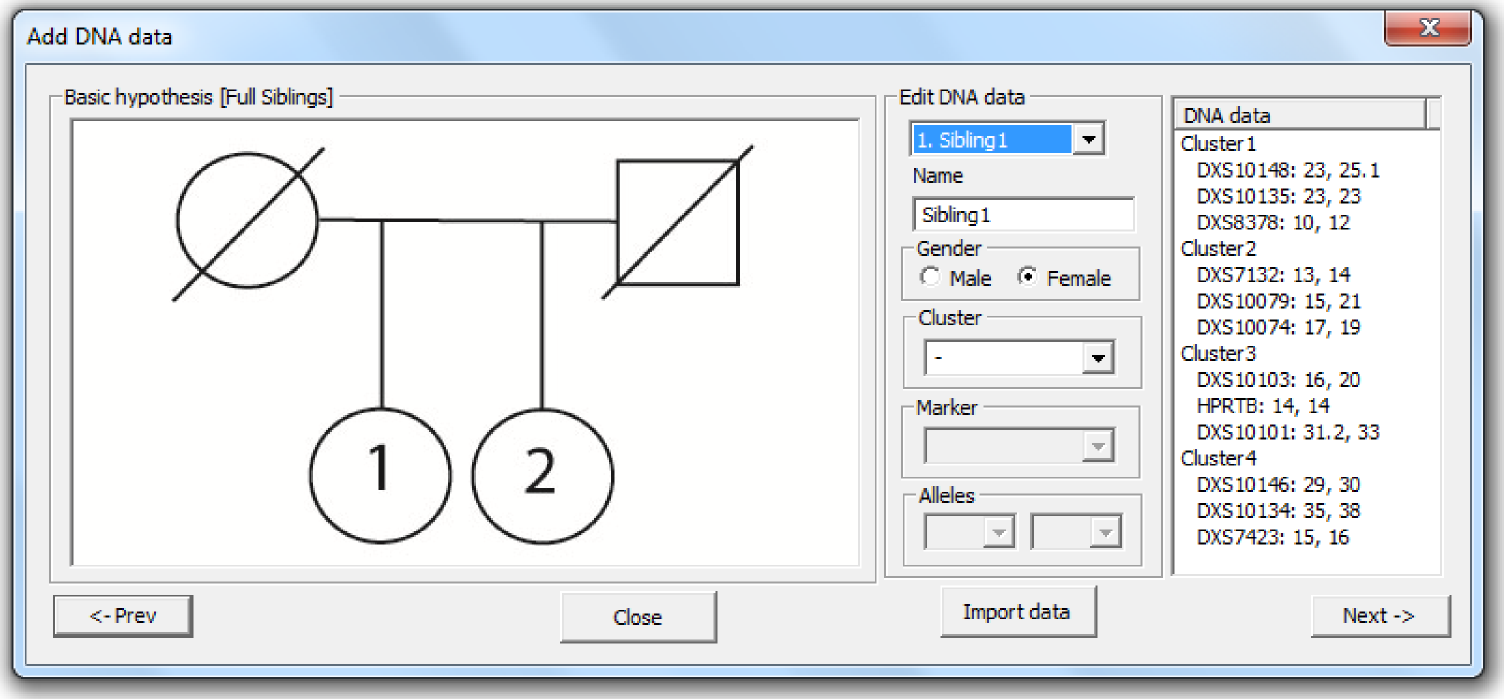
FamLinkX – At a glance

- Select alternative hypotheses



FamLink – At a glance

- Define DNA data



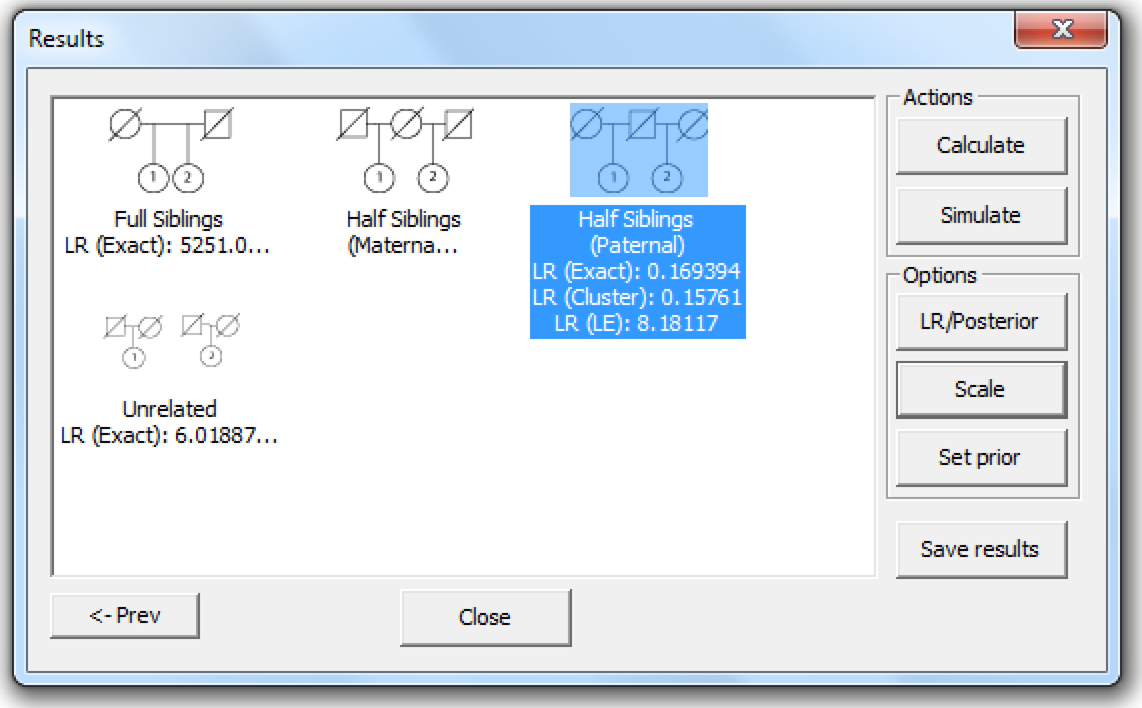
Basic hypothesis [Full Siblings]

Edit DNA data
1. Sibling1
Name: Sibling1
Gender: ○ Male ● Female
Cluster: -
Marker
Alleles

DNA data
Cluster1
  DXS10148: 23, 25.1
  DXS10135: 23, 23
  DXS8378: 10, 12
Cluster2
  DXS7132: 13, 14
  DXS10079: 15, 21
  DXS10074: 17, 19
Cluster3
  DXS10103: 16, 20
  HPRTB: 14, 14
  DXS10101: 31.2, 33
Cluster4
  DXS10146: 29, 30
  DXS10134: 35, 38
  DXS7423: 15, 16

FamLink – At a glance

– Calculate likelihoods



Results

Full Siblings
LR (Exact): 5251.0...

Half Siblings
(Materna...

Half Siblings
(Paternal)
LR (Exact): 0.169394
LR (Cluster): 0.15761
LR (LE): 8.18117

Unrelated
LR (Exact): 6.01887...

– We display three computation methods

FamLinkX – Questions?

EXERCISES