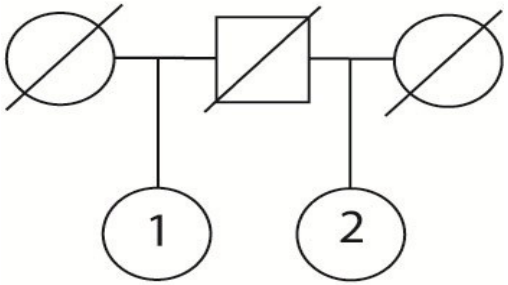
X-chromosomal markers and FamLinkX

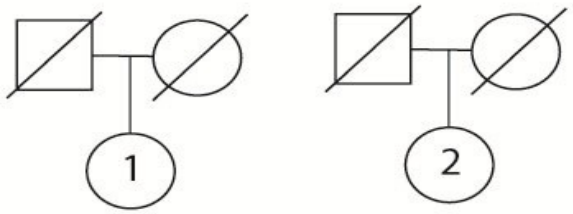
Athens, May 29, 2014

Use of markers on the X-chromosome to solve relationship issues

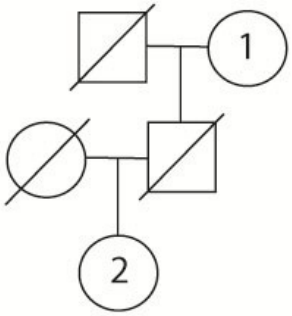
- X-chromosomal typing in males reveals their haplotype.
- Males transmit their whole chromosome X to their daughters.
- All sisters share their paternal ChrX haplotype.
- Furthermore, it is very likely that haplotypes of linkage groups remain stable throughout many generations. Consequently, they are a powerful means to demonstrate kinship

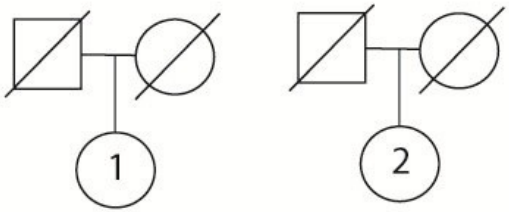
Case scenarios where X-STR typing is helpful





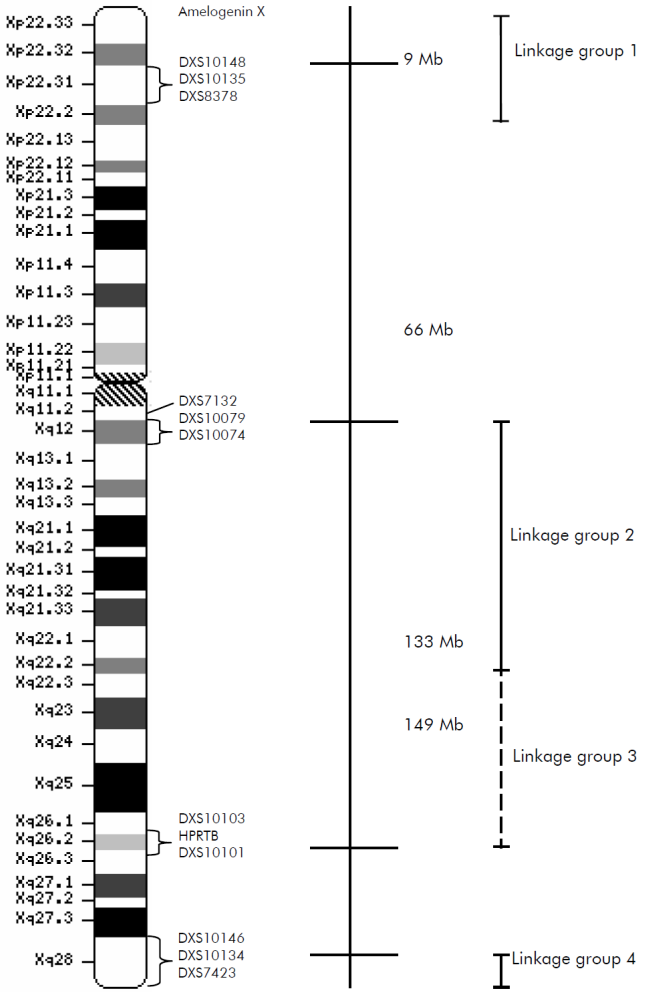
If the same father, 1 and 2 should share 1 allele (IBD) for each typed marker.





If related, 1 and 2 should share 1 allele (IBD) for each typed marker

Argus X12

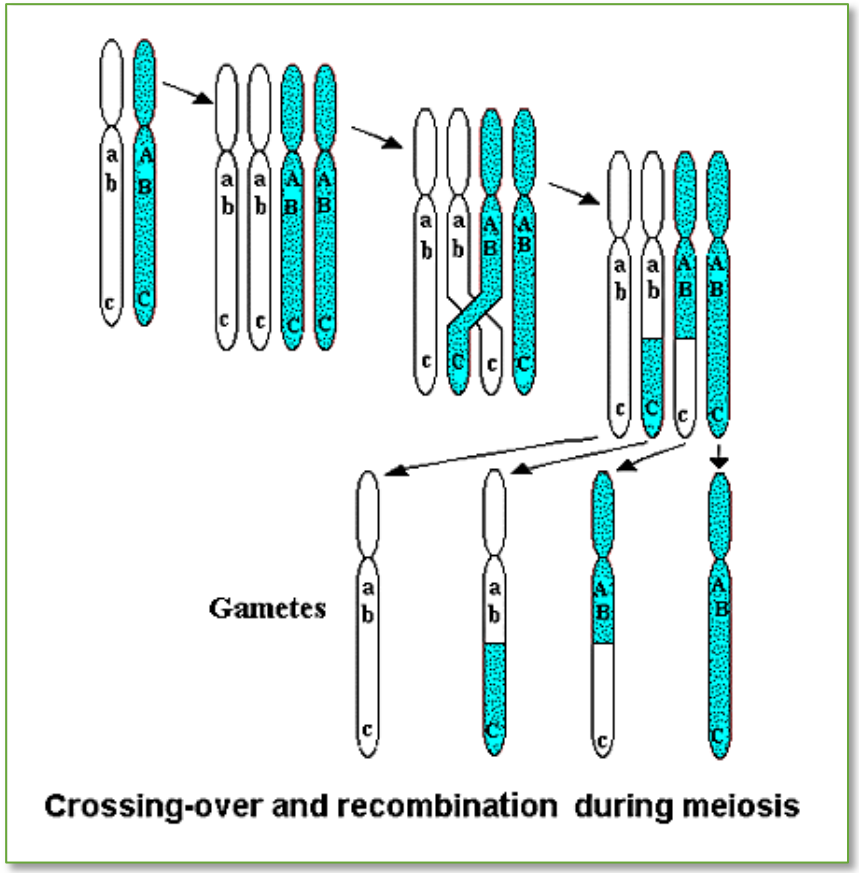


12 STRs in 4 clusters (3 STRs in each cluster)

LD and linkage should be accounted for

Linkage

- What is linkage?



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

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*

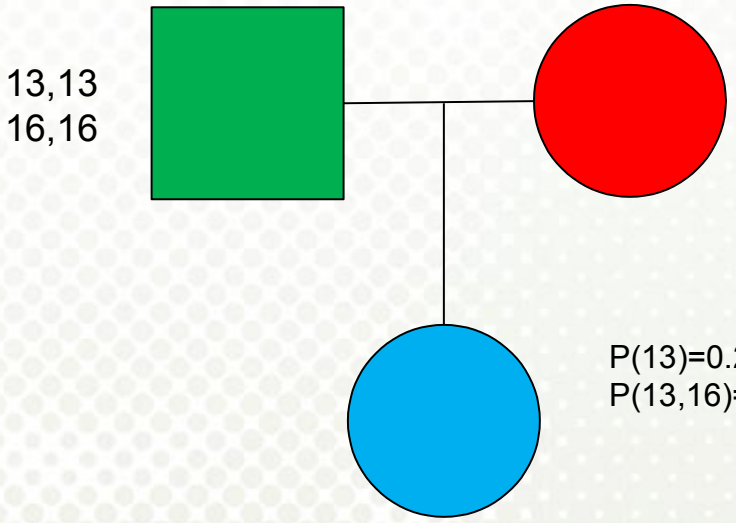
Observed frequency of [13, 16] is 0.12

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

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

Linkage disequilibrium

- Worked example, paternity with two markers



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,
Measured by the recombination rate constant decays with recombinations	Used to find alleles associated with a disease, in the population
Used to find markers linked to a disease, in families	

X-chromosomal markers

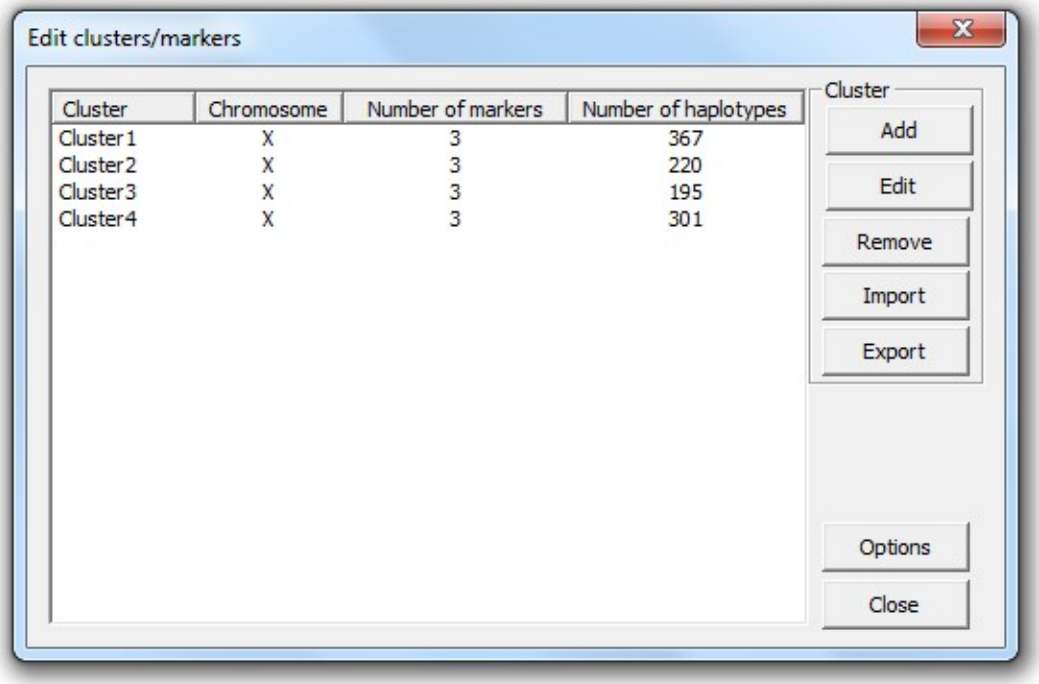
- Used where autosomal markers fail
- Argus X12
  - (4 clusters with three tightly linked markers)
- Linkage
- Linkage disequilibrium
- Mutations
- FamLinkX!
  - New joint probability model
  - Released autumn 2013

FamLinkX

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

FamLinkX – At a glance

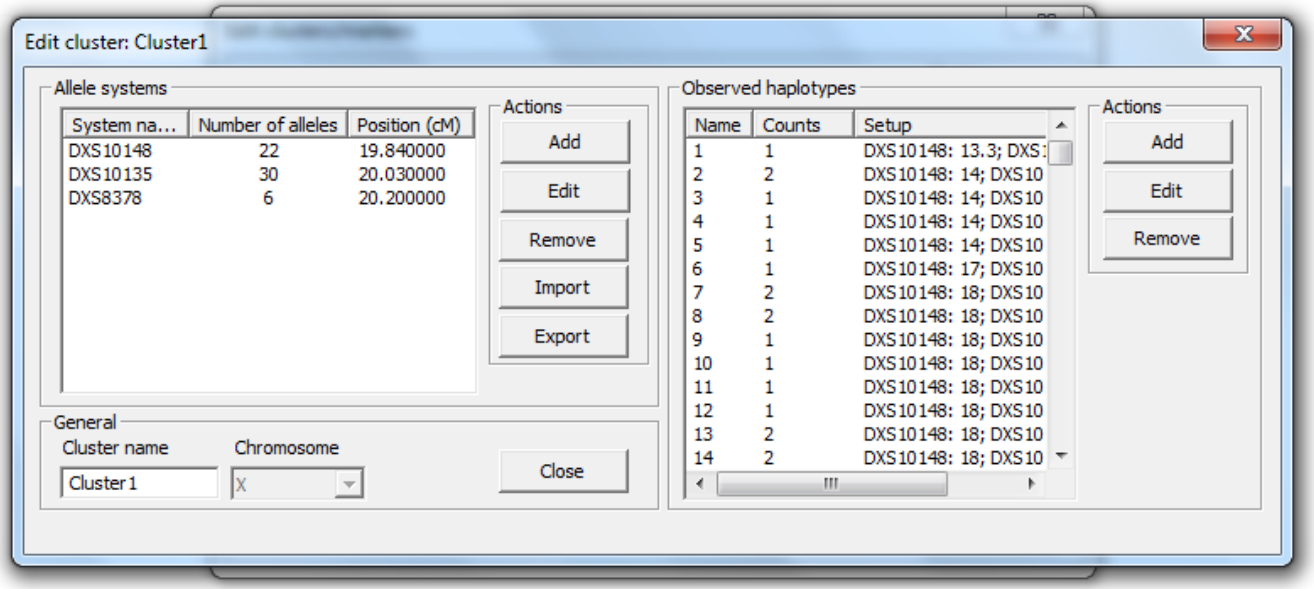
- Define clusters of markers



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

FamLinkX – At a glance

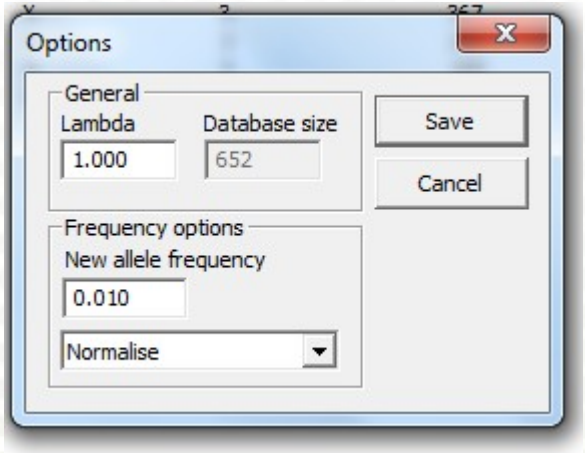
- Add markers
  - Define genetic position
  - Mutation parameters



- Add haplotypes(!)
  - Define setup
  - Counts

FamLink – At a glance

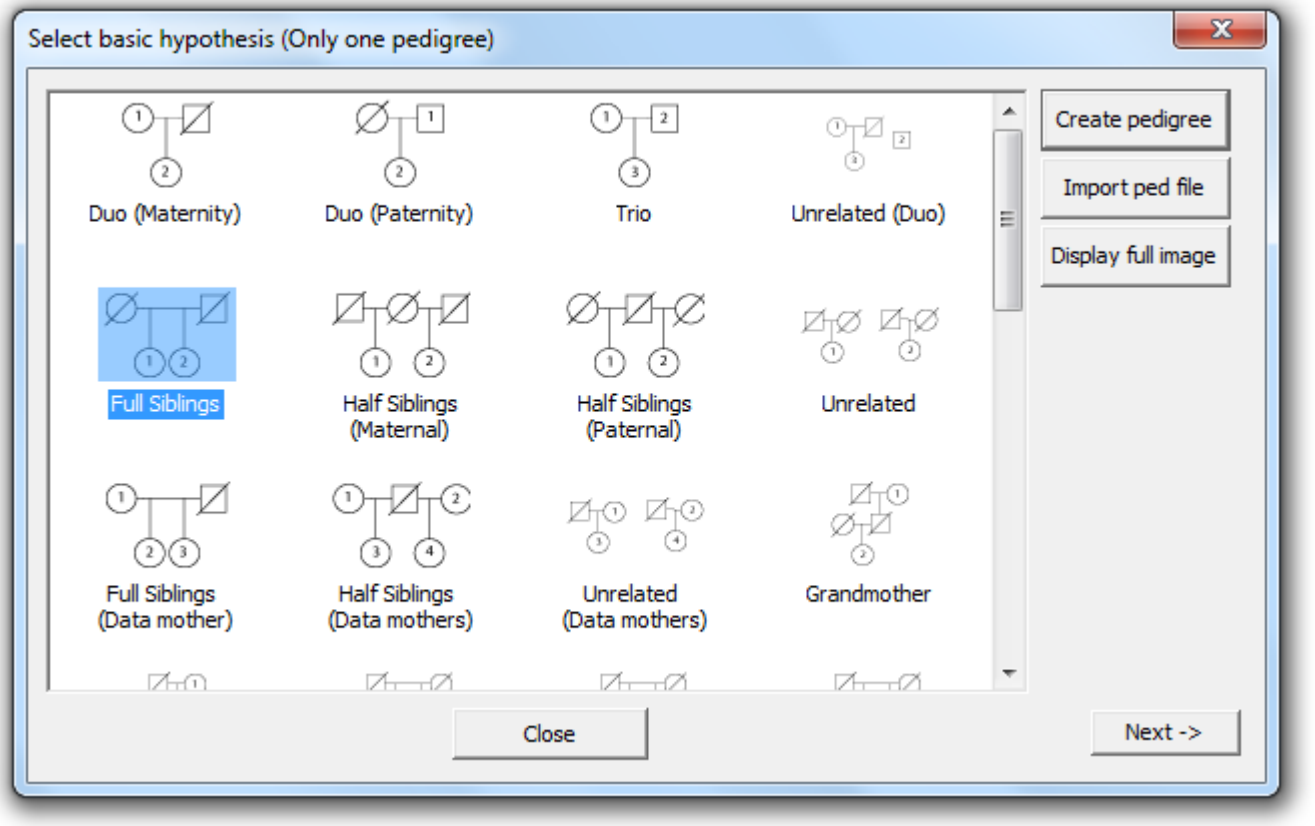
- Selecting value for Lambda



- We display two methods

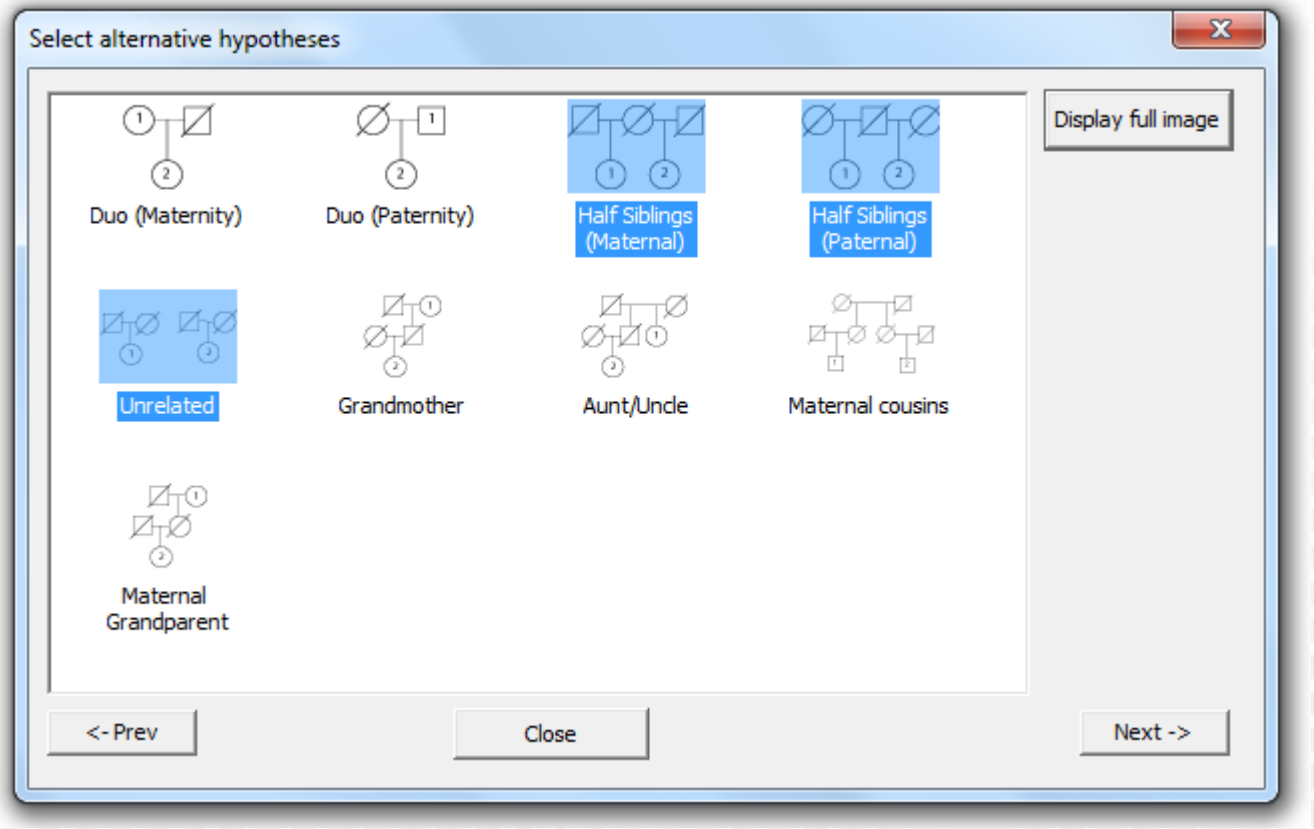
FamLinkX – At a glance

- Select main hypothesis



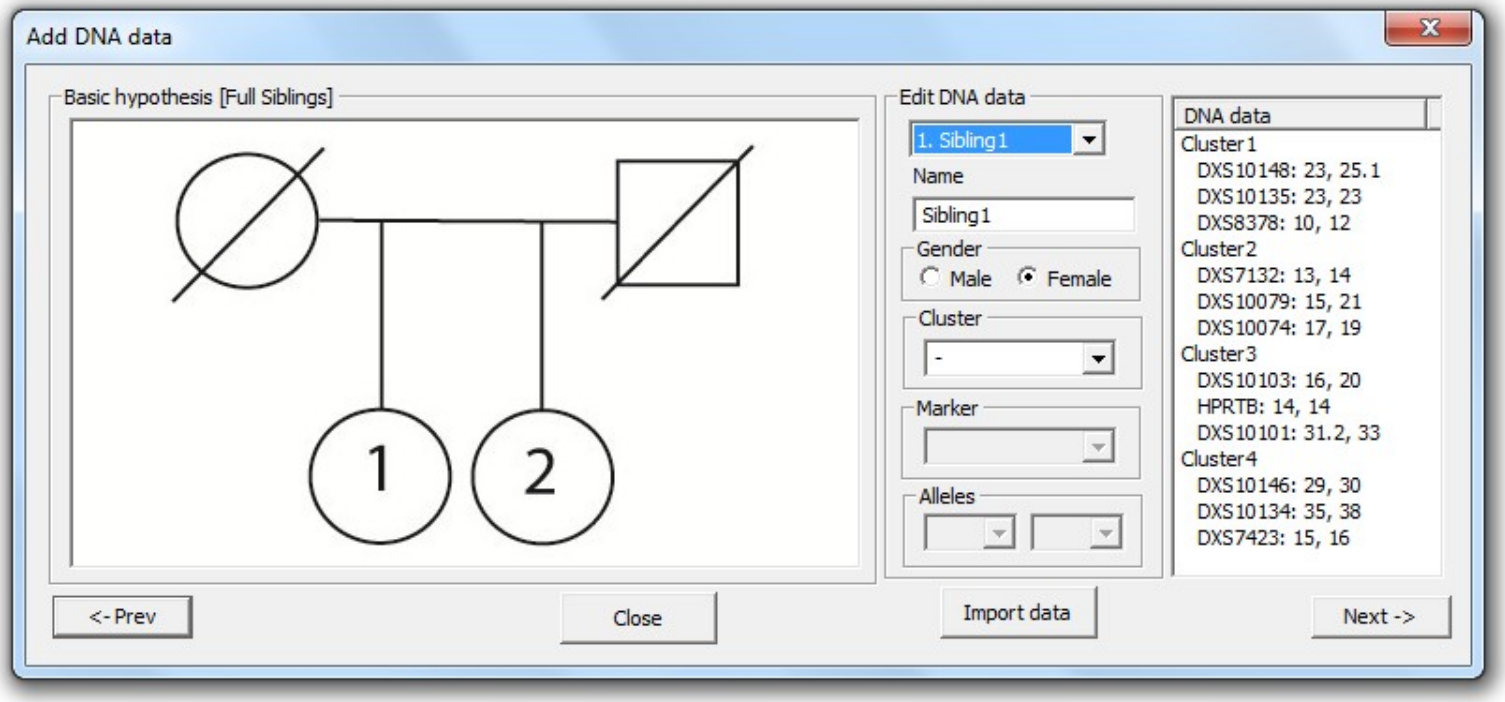
FamLinkX – At a glance

- Select alternative hypotheses



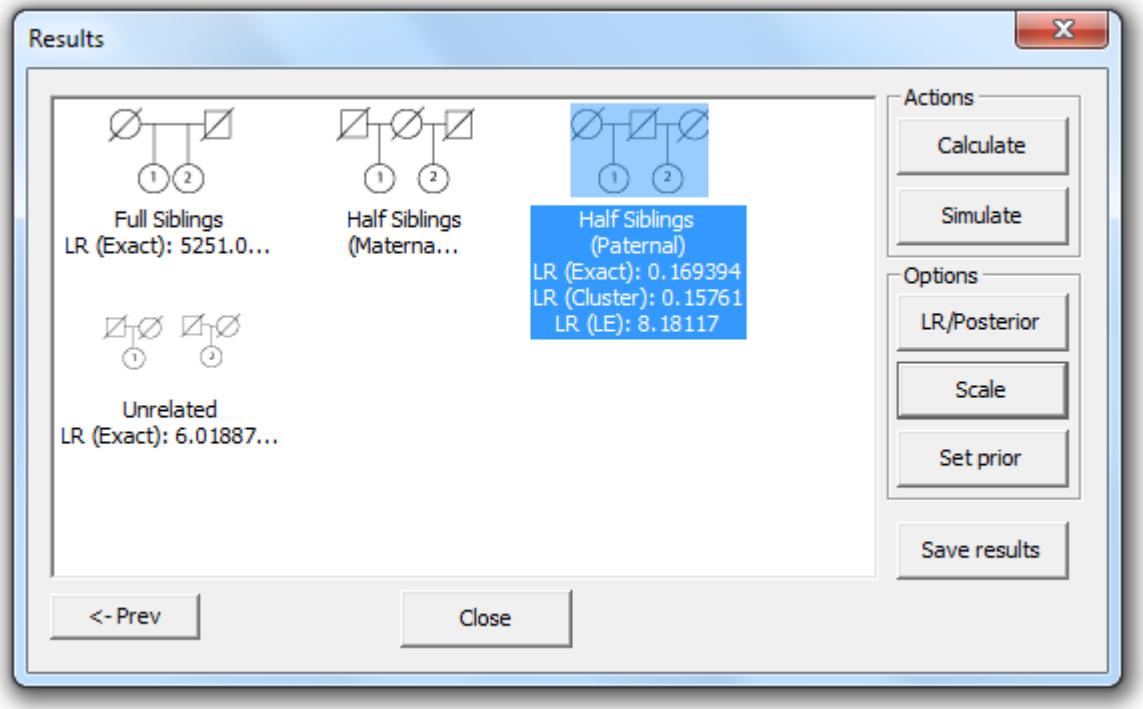
FamLink – At a glance

- Define DNA data



FamLink – At a glance

- Calculate likelihoods



- We display three computation methods

FamLinkX – Creating the database

– Size of the database?
  - Depend on the cluster size
– Include only males
  - Why?
– Input format for FamLinkX

ClusterID	Marker1	Marker2	... HaploCounts
1	13	20	10
2	13	21	11
.			
.			
.			

FamLinkX – Creating the database

– Estimation of updated haplotype frequencies
– The model

$$H_i = \frac{c_i + \lambda\, p_i}{C + \lambda}$$

Hi = Updated haplotype frequency
ci = Counts for haplotype *i*
C = Total number of haplotypes
pi = Expected haplotype frequency
λ = Prior weight given to expected frequency

– If λ=0, only observed haplotypes have a nonzero frequency.
– If λ=large, all haplotypes have a frequency.

FamLinkX – Questions?

EXERCISES