Model-based linkage analysis Part 2

26.10.2009

Yurii Aulchenko,
Erasmus MC Rotterdam

Outline

- Simple LOD score computations
- Heterogeneity analysis

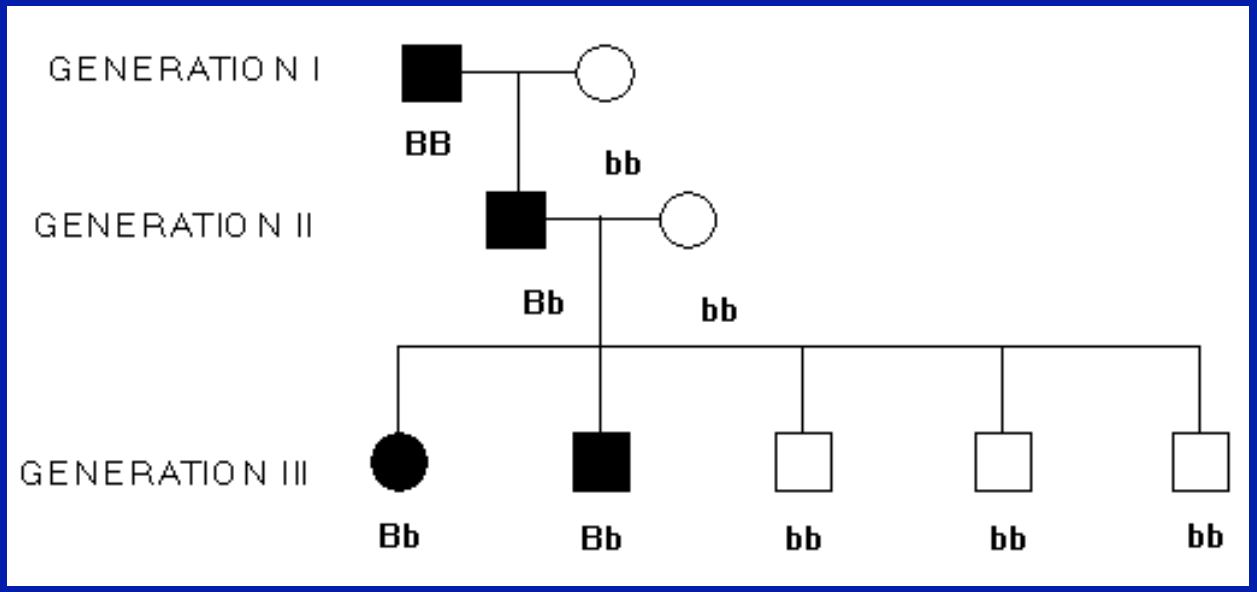
LOD score computations

Computation of LOD is not possible by hand and special software is to be used

We can still compute LODs by hand for some simple situations:

- Simple pedigree structures
- One completely informative marker in analysis
- Simple model of inheritance
- Absolute linkage vs. no linkage

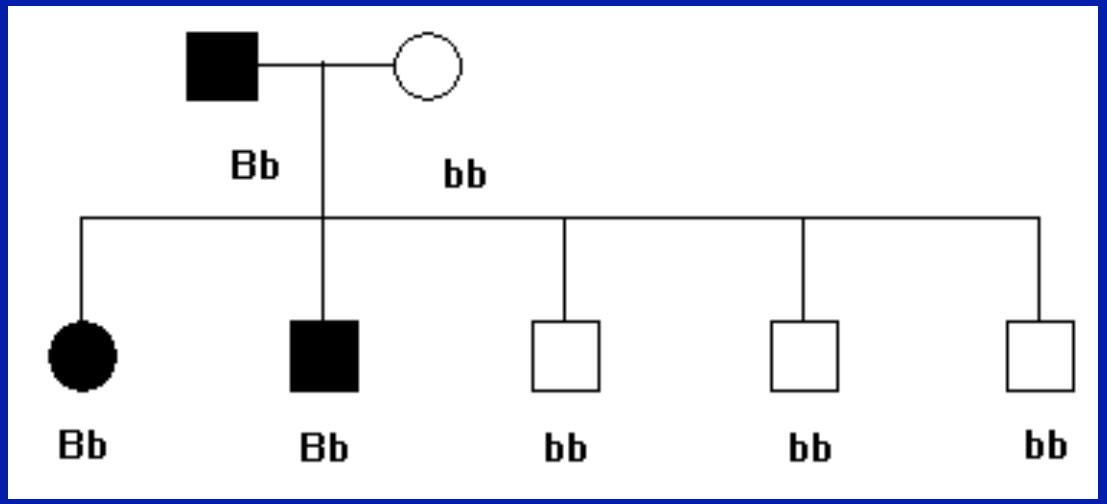
Dominant model, phase known



- If disease locus is linked with marker **B**, then the $P_1=1$
- If not, $P_0= (½)^5 = 2^{-5}$, where 5 is number of sibs
- LOD = $\log_{10}(P_1/P_0) = \log_{10}2^5 = 5 \log_{10}2 = 5*0.3 = 1.5$

- For N sibs LOD = N * 0.3

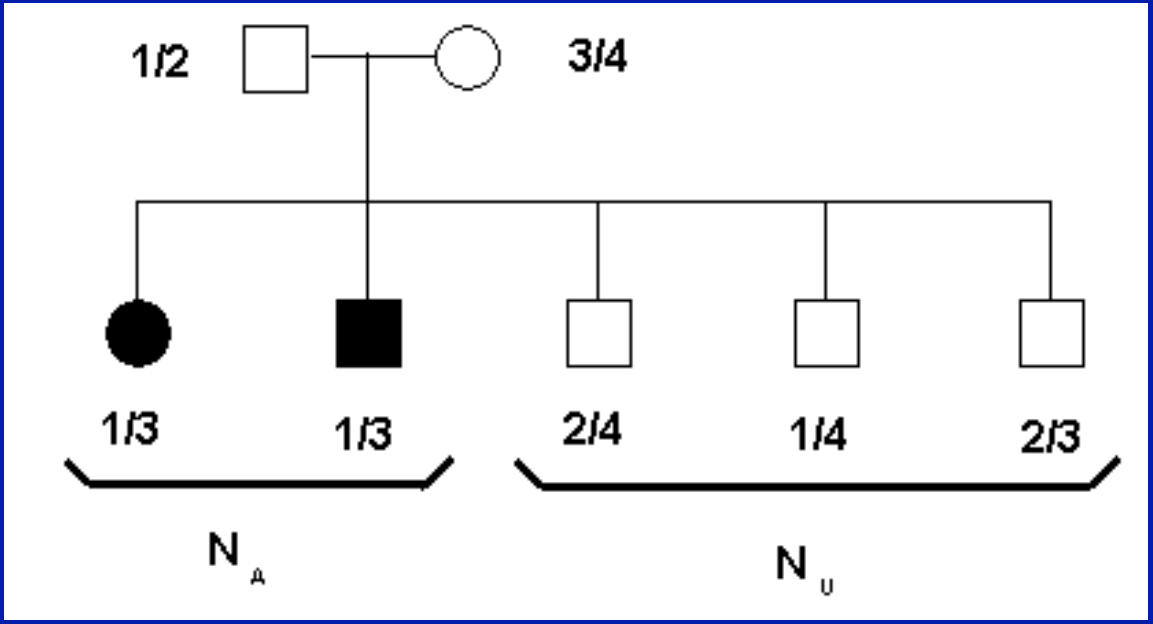
Dominant model, phase not known



- It can be either B or b linked with disease allele in the affected parent
- Under alternative $P_1 = ½$, under null $P_0 = (½)^5$
- LOD = $\log_{10}(P_1/P_0) = 4 \log_{10}2 = 4\ 0.3 = 1.2$

- For N sibs, LOD = (N-1) * 0.3

Recessive model



- N_A is number of affected
- N_U is number of unaffected

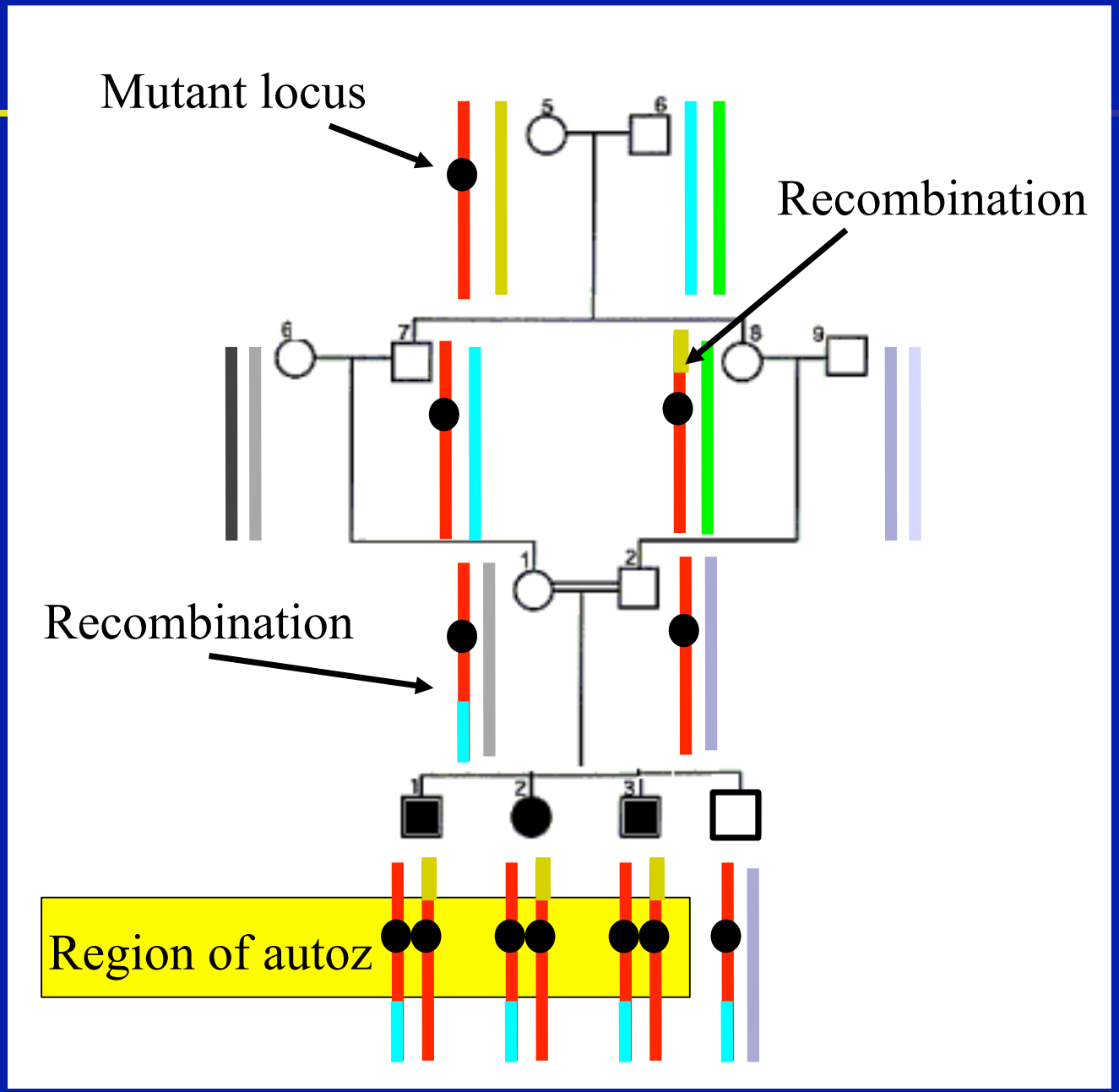
- $LOD = (N_A - 1) \log_{10} 4 + N_U \log_{10} 4/3 =$
  $= (N_A - 1)\ 0.6 + N_U\ 0.125$

Rationale for Homozygosity Mapping

Consider rare recessive allele model

In consanguineous families, the affected are most likely to be autozygous for the disease allele inherited from one of the founders

So will be the marker alleles surrounding the disease locus



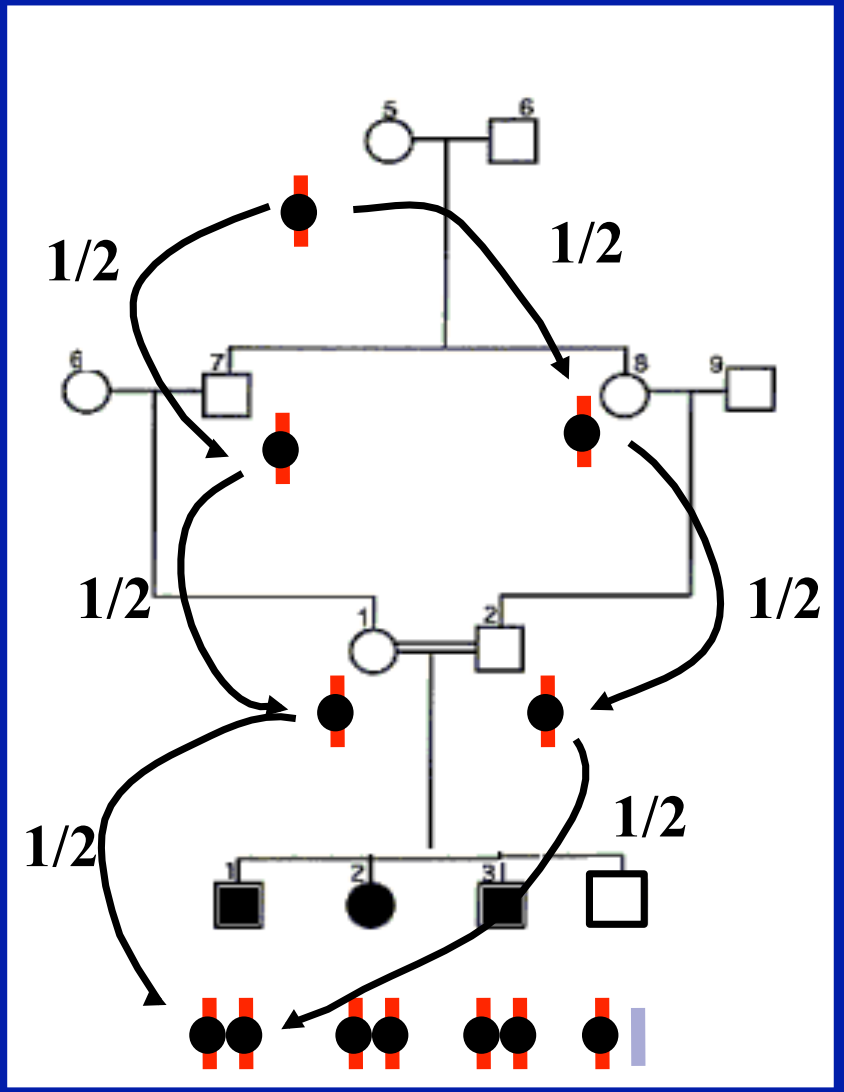
Likelihood of the data

Under alternative $P_1 = 1$

Under the null:

- Probability of HBD for a first offspring is $F = (½)^4$
- Probability of HBD for the next is ¼
- Probability of not HBD is ¾

$P_0 = F * ¼^{NA-1} * ¾^{NU}$



LOD score and comparison

- $LOD = \log_{10}1/F + (N_A - 1)\log_{10}4 + N_U \log_{10}4/3$
- LOD = 1.2 + 2 * 0.6 + 1 * 0.125
- For the above pedigree, LOD = 2.53

To reach LOD≥3

Dominant model

- Nuclear family with ≥11 offspring (at least one affected)

Recessive model

- Nuclear family, 6 affected offspring

- Cousin marriage, 4 affected offspring
- Second-cousin marriage, affected 3 offspring
- Two second-cousin marriages with 1 affected offspring

Outline

- Simple LOD score computations
- **Heterogeneity analysis**

Heterogeneity

Allelic heterogeneity:

- the same phenotype may be determined by different mutations in the same locus
- Linkage analysis **IS NOT** sensitive to that

Locus heterogeneity:

- the same phenotype may be determined by mutations in different genes
- Linkage analysis **IS** sensitive to that

Locus heterogeneity

Imagine the disease may be result of a rare mutation

- Mostly in locus 1 (located at say 1p) but also
- In locus 2 (located at say 17q)
- In locus 3 (located at say 12p)
- …

- Thus families where the disease is due to mutation in 1p will show linkage to 1p (such families will be call "linked to 1p")
- Other families will show negative LOD at 1p

Modeling heterogeneity

When studying specific location, assume that some proportion of families (α) exhibit the disease because of a mutation in this locus = "linked" (and [1-α] are not)

Probability under alternative:

- $P_1(\alpha) = P_{i=1,2,\ldots,K} [\alpha P_1(i) + (1 - \alpha) P_o(i)]$
- $P_0 = P_0(1) * P_0(2) * \ldots * P_0(K)$

Heterogeneity LOD:

$$HLOD(\alpha) = \log_{10}[\alpha P_1(1)/P_0(1) + (1 - \alpha)] + \ldots + \log_{10}[\alpha P_1(K)/P_0(K) + (1 - \alpha)]$$

Which families are linked?

Given the data, what is the posterior probability that a family segregates for a mutations in this region?

Selecting the families for further mutation screening

Posterior probability of linkage

Given

- Some region of interest
- Maximum likelihood estimate of α (proportion of linked families)
- LOD score for a particular family i

$$LOD(i) = \log_{10} P_1(1)/P_0(1)$$

Posterior probability of linkage of i^{th} family is

$$\alpha\, 10^{LOD(i)} / [\alpha\, 10^{LOD(i)} + (1 - \alpha)]$$