Lecture 7: Interval mapping of QTL

Julius van der Werf

Single versus multiple markers

Association between a quantitative trait and genetic markers can be evaluated using single markers or multiple markers. When using one single marker, it is possible to make inference about the segregation of a QTL linked tot that marker. However, in *the case of single markers it is not possible to distinguish between size of a QTL effect and its position (relative to the marker)*. Also, single marker analyses have less power if the markers are far apart.

If two (or more) markers are jointly used in an analysis, there is a lot less confounding between the position and size of QTL effect, and there is more power in detecting a QTL, even if the markers are far apart. Inference about the QTL effect as well as the recombination rate between QTL and markers (i.e. position of QTL) is possible. The recombination rate between markers is usually assumed known.
Therefore mapping of a QTL therefore requires the use of multiple marker genotypes in the analysis. .

Determining associations between genetic markers and QTL with two markers

For two markers, the QTL probability given the marker genotype depends on more recombinations: those are the recombination rates between M1 and QTL (=r1), between M2 and QTL (=r2) and between M1 and M2 (=r12).
We consider again a half sib design where we know the sires marker genotype for two markers, the sire is heterozygous for the QTL and we know the marker-QTL phase.

TABLE 1

Parental genotype			M1 Q M2 m1 q m2	
Possible gametes			recombination?	Gamete probability
M1	Q	M2	no	(1-r1)(1-r2)/2
M1	q	M2	double: M1-q, q-M2	r1.r2/2
M1	Q	m2	yes: Q-m2	(1-r1)r2/2
M1	q	m2	yes: M1-q	r1(1-r2)/2
m1	Q	M2	yes: m1-Q	r1(1-r2)/2
m1	q	M2	yes: q-M2	(1-r1)r2
m1	Q	m2	double: m1-Q, Q-m2	r1.r2/2
m1	q	m2	no	(1-r1)(1-r2)/2

Assume now also (for simplicity) that we know which marker alleles came from the sire. We can now work out the expected difference between the paternal marker genotype-groups in the sire's progeny:

TABLE 2

Marker alleles obtained from sire group	QTL allele obtained from sire	frequency	Expected mean of progeny
M1M2	Q	(1-r1)(1-r2)/2	$\mu + \alpha$
M1M2	q	r1.r2/2	μ
M1m2	Q	(1-r1)r2/2	$\mu + \alpha$
M1m2	q	r1(1-r2)/2	μ
m1M2	Q	r1(1-r2)/2	$\mu + \alpha$
m1M2	q	(1-r1)r2	μ
m1m2	Q	r1.r2/2	$\mu + \alpha$
m1m2	q	(1-r1)(1-r2)/2	μ

α = average effect of allele substitution of Q (over q).

Some tedious algebra would give the following means for the possible paternal marker-haplotypes in progeny (sum of frequency * mean of group and divide by frequency of marker haplotype group)

TABLE 3. Expected means of different marker haplotypes.

Mean of M1M2-group:	$\frac{\frac{1}{2}(1-r1)(1-r2)(\mu+\alpha)+\frac{1}{2}r1.r2.\mu}{\frac{1}{2}(1-r12)}$ =	$\mu + (1-\frac{r1r2}{1-r12})\alpha$
Mean of M1m2-group:	$\frac{\frac{1}{2}(1-r1).r2.(\mu+\alpha)+\frac{1}{2}r1(1-r2)\mu}{\frac{1}{2}r12}$ =	$\mu + \frac{r2-r1r2}{r12}\alpha$
Mean of m1M2-group:	$\frac{\frac{1}{2}r1(1-r2)(\mu+\alpha)+\frac{1}{2}(1-r1).r2.\mu}{\frac{1}{2}r12}$ =	$\mu + \frac{r1-r1r2}{r12}\alpha$
Mean of m1m2-group:	$\frac{\frac{1}{2}r1.r2(\mu+\alpha)+\frac{1}{2}(1-r1)(1-r2)\mu}{\frac{1}{2}(1-r12)}$ =	$\mu + \frac{r1r2}{1-r12}\alpha$

The difference between the M1M2 and m1m2 haplotypes is now equal to .

$$[\ \mu + (1-\frac{r1r2}{1-r12})\alpha\ ] \ -[\ \mu + \frac{r1r2}{1-r12}\alpha\ ] = \ (1-\frac{2r1r2}{1-r12})\alpha$$

and as r1r2 is usually a small number, this difference is quite close to the actual QTL allelic effect (α). The coefficient for α in Table 3 in the last column is the probability of having inherited Q from the sire, conditional on (given the) the paternal marker haplotype. This is shown more explicit in Table 4.

TABLE 4. Probabilities for having inherited the paternal Q-allele of different marker haplotypes.

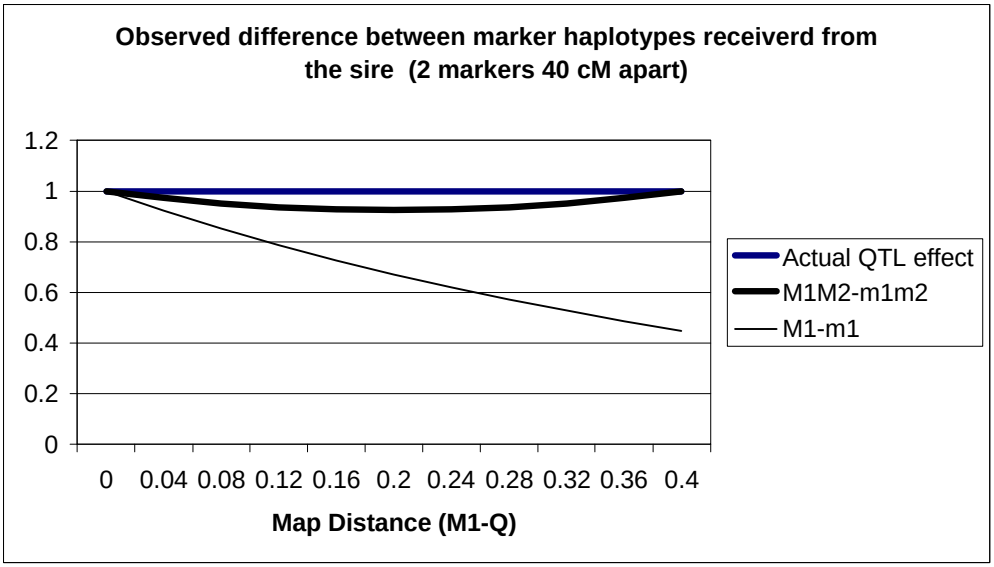
Prob(Q|M1M2)	$\frac{\frac{1}{2}(1-r1)(1-r2)}{\frac{1}{2}(1-r12}$	=	$(1-\frac{r1r2}{1-r12})$
Prob(Q|M1m2)	$\frac{\frac{1}{2}(1-r1)r2}{\frac{1}{2}r12}$	=	$\frac{r2-r1r2}{r12}$
Prob(Q|m1M2)	$\frac{\frac{1}{2}r1(1-r2)}{\frac{1}{2}r12}$	=	$\frac{r1-r1r2}{r12}$
Prob(Q|m1m2)	$\frac{\frac{1}{2}r1r2}{\frac{1}{2}(1-r12)}$	=	$\frac{r1r2}{(1-r12)}$

The following Table 5 gives an example of the probabilities of having inherited the Q-allele in a half-sib family, given the marker haplotypes (PQ|MiMj). The distance between the markers is 40 cM. The QTL location investigated is at 10 cM from M1. Haldane's mapping function is used to determine recombination rates based on these distances. Tables 1, 3 and 4 are used to determine probabilities. Table 3 is used to determine expected means of each marker type, assuming QTL genotypic means of 10 and 11 for qq and QQ, respectively. The dam population is assumed to have a q-frequency of 1. (comparable with a backcross design)

Table 5

Paternal	Probability of marker haplotypes			Qq mean =10 qq mean =9.5		Mean
Markertype	P(M1M2)	P(M1QM2)	P(Q|M1M2)	prob(Qq)	Prob(qq)	Expected
M1M2	0.362	0.352	0.972	0.972	0.028	9.986
M1m2	0.138	0.103	0.745	0.745	0.255	9.873
m1M2	0.138	0.035	0.255	0.255	0.745	9.627
m1m2	0.362	0.010	0.028	0.028	0.972	9.514

The following figure shows the difference between marker haplotype groups in progeny for a single marker (M-m) and for two markers (M1M2-m1m2), for different positions of the QTL relative to the M1.



The figure shows that the difference between the non-recombinant marker haplotypes is much less affected by the marker-QTL distance than the M1-m1 difference for the single markers. Moreover, the map position is now not confounded with the QTL effect. In a way, map position and QTL effect have become estimable with two markers.

The example shown here is based on half sib analysis. The interpretation of the genetic effect estimated depends on the constitution of the dam population, as shown in the previous chapter. If we want to estimate both a and d, we need a dam population that contributes both q and Q alleles, and where we can trace the inheritance from the dam. In other words, we need to identify also segregation from the dam. Choosing the dam population from a F1-cross of two extreme lines (extreme with respect to the putative QTL) would be the best choice.

Inbred lines have been used in QTL mapping to avoid uncertainty about the genetic effects estimated. However, in animal population, complete inbred lines (with marker-

and QTL alleles fixated) are hardly feasible, and possibly less relevant for QTL's to be used in practical applications.
In outbred populations, there is less certainty about the animals' QTL genotypes. Lack of design usually means that the marker genotypes are frequently not informative about paternal or maternal origin. In the next chapter, the advantages and disadvantages of different design will be discussed.
At this stage we can continue that for 'any' design, the QTL estimation is based on two steps

1) What is the probability that an individual has a certain QTL genotype (give the observed marker genotypes)
2) What is the estimated effect of this particular genotype on the individuals' phenotypes

The first step is much easier in well-defined experiments. The second step can be quantified either by using the likelihood principle, or by using regression (where the match is measured in terms of residual sums of squares).
We present the principle briefly here, and in Chapter 9 we will discuss in more detail these different methods.

Interval mapping

Maximum Likelihood
The term 'interval mapping' is used for estimating the position of a QTL within two markers (often indicated as 'marker-bracket'). Interval mapping is originally based on the maximum likelihood but there are also very good approximations possible with simple regression (see Chapter 9).

The principle is:

1) The Likelihood can be calculated for a given set of parameters (particularly QTL-effect and QTL position) given the observed data on phenotypes and marker genotypes.

2) The estimates for the parameters are those were the likelihood are highest.

3) The significance can be tested with a likelihood ratio test:

$$\text{LR} = -2\ln\frac{Max\_Likelihood(reduced\ \mathrm{mod}\ el)}{Max\_Likelihood(full\ \mathrm{mod}\ el)}$$

The reduced model refers to the null-hypothesis, e.g. "there is no QTL effect"

Using the log-likelihood: $LR = -2.(ln\_L_r - ln\_L)$

where ln_L is the $\log_e$ of the maximum likelihood.

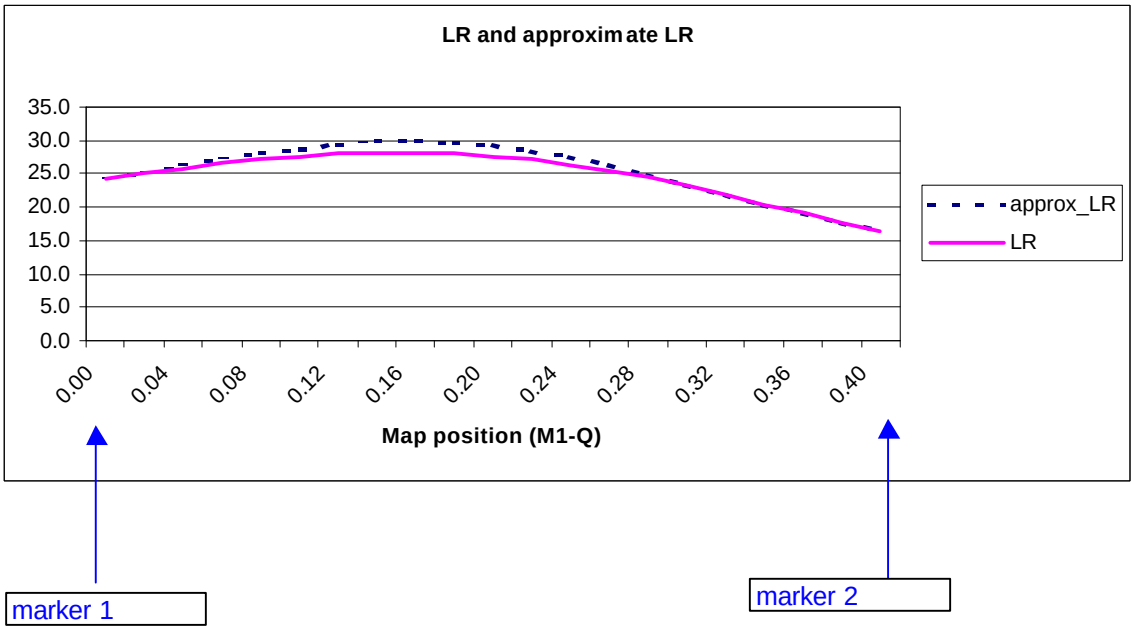
The evidence for a particular QTL at a particular chromosomal position can be displayed as a *likelihood map*, The LR-statistic is plotted against the map position of the QTL.

Lander and Botstein (1989) introduced first the concept of likelihood maps. The proposed to use the LOD-score as a test statistic. However, the LOD score is equal to a constant (1/4.61) time the LR test statistic, as shown:

The LOD score for a QTL at position c is:

$$\text{LOD(c)} = \log_{10} \frac{Max\_Likelihood(reduced \bmod el)}{Max\_Likelihood(full \bmod el, c)} = \frac{LR(c)}{2\ln 10} \approx \frac{LR(c)}{4.61}$$

The following figure shows a likelihood map for a marker bracket based on simulated data from one half sib family (backcross) with 300 progeny. The simulated QTL effect was 0.5 within-family standard deviations. The figure shows the true LR value based on ML, and the approximate LR (upper line) based on regression analysis.



Regression methods

Regression analysis is easier (standard software can be used) and usually much quicker than maximum likelihood, and in many cases, it is very similar. In Chapter 9 we will discuss the different methods in more detail. The basic idea is given here in the context of interval mapping

Regression on QTL probability, conditional on marker haplotypes.

For a given haplotype that was inherited from the sire, we can calculate the probability for having inherited the Q or the q allele. It seems therefore natural to regress phenotype on Q-probability. The model is

$$y = \mu + \alpha.x$$

where y is the observed phenotype
x is the probability of having inherited a paternal Q, given the observed marker genotype.

The coefficient in x is obtained as P(Q|Mi Mj) for a given QTL position. There are only 4 different x-values, one for each haplotype (Table 4). Note that different positions give different coefficients.

For each recorded animal, we can then give a predicted phenotype with this "QTL-model" which is equal to

$$\hat{y}_i = \hat{\mu} + \hat{a}.x_i$$

where the "hats" refer to estimated (predicted) values.

A model ignoring a QTL would predict each observation as

$$\hat{y} = \hat{\mu}_0$$

where $\hat{\mu}_0$ is typically the general progeny mean

Now let the total sum of squares (SST) be the sum (over animals) of $(\hat{y} - \hat{\mu}_0)^2$

and let the residual sum of squares (SSE) be the sum (over animals) of $(\hat{y}_i - \hat{\mu} - \hat{a}.x_i)^2$

Each map position will yield an SSE and the position with the lowest SSE is the most likely position.

A test statistic for this method is for an experiment with n observations is

$$LR = n \ln(\frac{SST}{SSE})$$

where n is equal to the number of observations. The LR stands for “Likelihood Ratio”, as this test statistic is approximately similar to the LR from maximum likelihood.
Haley and Knott (1992) have shown that this similarity. If there are more fixed effects in the model, the test statistic is calculated as

$$LR = n \ln\left(\frac{SSE_{reduced}}{SSE_{full}}\right)$$

Which is ration of the residual sums of squares in a model with the QTL (”full’) and a model without it (‘reduced’).

The information about a QTL is only dependent on the flanking markers. If the QTL lays outside the bracket, it will only depend the nearest marker. Likelihood maps can be constructed for neighboring marker brackets and they should exactly match up at each marker, and a map of multiple intervals M1-M2-M3....-Mk is smooth.

An example of interval mapping is given on the next page.

References

Haley, C.S. and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324

Lander, E.S. and D. Botstein.1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Example of QTL mapping by regression:
Data on 8 individuals with paternal marker haplotypes given. The probabilities are derived for different positions (dM1-Q is distance between marker 1 and QTL), with further the same assumptions as in this chapter (see Table 5).

[X]		yhat	y	markers
1.0000	1.0000	50.3656	50.9813	M1M2
1.0000	1.0000	50.3656	49.9813	M1M2
1.0000	1.0000	50.3656	50.7500	M1m2
1.0000	1.0000	50.3656	49.7500	M1m2
1.0000	0	50.1344	50.7500	m1M2
1.0000	0	50.1344	49.7500	m1M2
1.0000	0	50.1344	50.5187	m1m1
1.0000	0	50.1344	49.5187	m1m1

dM1-Q	SST	SSE	LR
0	2.2139	2.1070	0.3961

[X]		yhat	y
1.0000	0.9718	50.4321	50.9813
1.0000	0.9718	50.4321	49.9813
1.0000	0.7451	50.3446	50.7500
1.0000	0.7451	50.3446	49.7500
1.0000	0.2549	50.1554	50.7500
1.0000	0.2549	50.1554	49.7500
1.0000	0.0282	50.0679	50.5187
1.0000	0.0282	50.0679	49.5187

dM1-Q	SST	SSE	LR
0.1	2.2139	2.0455	0.6331

[X]		yhat	y
1.0000	0.9718	50.4321	50.9813
1.0000	0.9718	50.4321	49.9813
1.0000	0.2549	50.1554	50.7500
1.0000	0.2549	50.1554	49.7500
1.0000	0.7451	50.3446	50.7500
1.0000	0.7451	50.3446	49.7500
1.0000	0.0282	50.0679	50.5187
1.0000	0.0282	50.0679	49.5187

dM1-Q	SST	SSE	LR
0.3	2.2139	2.0455	0.6331

[X]		yhat	y
1.0000	0.9625	50.4813	50.9813
1.0000	0.9625	50.4813	49.9813
1.0000	0.5000	50.2500	50.7500
1.0000	0.5000	50.2500	49.7500
1.0000	0.5000	50.2500	50.7500
1.0000	0.5000	50.2500	49.7500
1.0000	0.0375	50.0187	50.5187
1.0000	0.0375	50.0187	49.5187

dM1-Q	SST	SSE	LR
0.2	2.2139	2.0000	0.8129

[X]		yhat	y
1.0000	1.0000	50.3656	50.9813
1.0000	1.0000	50.3656	49.9813
1.0000	0	50.1344	50.7500
1.0000	0	50.1344	49.7500
1.0000	1.0000	50.3656	50.7500
1.0000	1.0000	50.3656	49.7500
1.0000	0	50.1344	50.5187
1.0000	0	50.1344	49.5187

dM1-Q	SST	SSE	LR
0.4	2.2139	2.1070	0.3961