Lecture 11: Multiple trait models for QTL analysis

Julius van der Werf

Multiple trait mapping of QTL 99
Increased power of QTL detection.....99
Testing for linked QTL vs pleiotropic QTL 100
Multiple trait analysis using regression.....101
Multiple trait analysis using logistic regression.....101
References 103

Multiple trait mapping of QTL

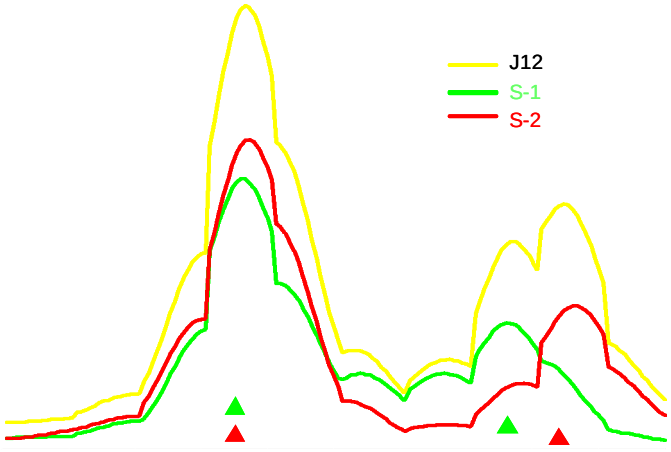
Jiang and Zeng (1995) have proposed a multiple trait version of the composite interval mapping. Their method is based on maximum likelihood, and requires special programs for analysis. The authors should considerable increase in power when using information from two correlated traits.

Most QTL detection studies comprise phenotypic data on multiple traits. Joint use of data from multiple traits in QTL analysis has two advantages: increased power and testing of models regarding the genetic correlation between two traits.

Increased power of QTL detection

Multiple traits that are correlated can add information to each other. To some extent, two measurements on correlated traits are somewhat like repeated measurements. Therefore, information from correlated traits can reduce the effect of error variance, therefore making it easier (more powerful) to detect QTL. Not only the power of QTL detection is increased, also the precision of the QTL map position is better.

Illustration of increased power from using joint analysis of two traits (J12) over single trait analysis (S1 and S2)
Jiang and Zheng, 1995



Jiang and Zeng (1995) also discussed the increased power from multiple trait analyses in relation to the correlation structure. In summary:

1. If the correlation between the traits (here: correlation between residual effects, this could be the sum of residual and polygenic effects) is zero, the joint test statistic is approximately the sum of the test statistics for the single traits

$$LR_j \approx LR_{S1} + LR_{S2} \qquad \text{if correlation} = 0$$

2. If the QTL is only affecting one of the two traits, say $\alpha_2 = 0$, then a joint analysis can increase the test statistic of detecting that trait, depending on the correlation (r) between the two traits.

$$LR_j \approx LR_{S1}/(1-r)^2 \geq LR_{S2}$$

3. The joint test statistic is equal or greater than the maximum of the single trait statistics.

$$LR_j \geq \text{maximum}[LR_{S1}, LR_{S2}]$$

4. $r\,\alpha_1\,\alpha_2 < 0$ (i.e r and $\alpha_1\,\alpha_2$ have different signs)

$$LR_j > LR_{S1} + LR_{S2}$$

This is the most favourable situation for using multiple traits analysis.

Testing for linked QTL vs pleiotropic QTL

When two QTL are found in the same region, when using single trait analysis, the question arises whether these are actually the same genes affecting both traits, or whether these are two separate QTL.
Unravelling this difference allows to better understand the nature of a genetic correlation between two traits. This would provide information concerning the possibility to break a unfavourable genetic correlation between two characters (in the case of linkage) or whether this is impossible (in the case of pleiotropism).
The test can be carried out with H_0: position 1 = position 2
H_1: position 1 ≠ position 2

Also other genetic models could be compared and tested (depending on design)
- Existence of epistasis (see Chapter 10)
- QTLs effecting one trait only vs effect on both traits

Maximum likelihood might be a bit laborious for multiple trait analyses, especially when comparing a range of genetic models.

Multiple trait analysis using regression

Moser (2000) has proposed a multiple trait regression approach and showed that again regression is very similar to maximum likelihood methods (at least in designed experiments).
As in single trait analysis, the approximate $LR \approx n\ ln(SSE_{reduced} / SSE_{full})$

Moser proposes to use for a multiple trait analysis

$$LR \approx n\ ln(|VE_{reduced}| / |VE_{full}|)$$

i.e. rather than the sum of squares of errors of a single trait analysis, he used the determinant of the matrix with residual sum of squares and sum of cross products of errors for two traits.
The advantage of the simple multiple trait regression method is that

1) permutation tests are feasible
2) a number of genetic models can bet tested and compared
   Moser (2000) used a genetic algorithm to efficiently find the most likely genetic model (as described in the previous chapter).

Multiple trait analysis using logistic regression

Henshall and Goddard (1999) proposed to use logistic regression for multiple trait QTL mapping. In fact, this method is also very useful for single trait analysis.

Logistic regression is used for traits where the response variable has a binomial distribution. Henshall and Goddard (1999) regressed, within half sib families, QTL genotype on phenotype. The QTL genotype refers to which QTL allele was received from the heterozygous sire (either Q or q). This is a 0/1 response with a probability, hence binomially distributed. Hence, rather than comparing phenotypic means for different marker genotype classes, they compared marker genotype classes for different phenotypes.

The main advantages of this method:

1) It is much simpler than maximum likelihood and standard software (like SAS) can be used, even for multi trait analyses. Maximum likelihood methods would be much more complex, as all data that was used in selection would have to be included in the analysis. Logistic regression however, is nearly equivalent than ML.

   Example: analysis of the traits Y and Z would require in SAS

```
proc logistic;
     model Q/n = Y Z
```

```
run;
```

The variable Q is the marker genotype (0 or 1) and n is the number of trials for each observation (=1)

2) The phenotypic observations can be subject to selection (as regression is not affected by regression on the 'x-variable'. Hence, logistic regression is a simple method that is applicable to data obtained from selective genotyping.

The principle of the method is as follows:

Let p = P(Q), i.e. probability of having inherited the Q-allele from the sire and assume genotype means of $\mu + \alpha$ and $\mu - \alpha$ for genotypes Q- and q- resp.
.
In single trait analysis, the logistic regression model is written as:

$$\log(\frac{p}{1-p}) = a + by$$

The QTL effect can be calculated as $a = \frac{-1+\sqrt{1+b^2\sigma^2}}{b}$

where σ^2 is the sum of the residual variance σ_e^2 and the QRTL variance = $\alpha^{2.}$

In multiple trait analysis, the model is: $\log(\frac{p}{1-p}) = Y'\beta$

where Y and β are vectors. The vector of QTL effects is

$A = \frac{\Sigma\beta}{1+\sqrt{\beta'\Sigma\beta+1}}$ where Σ = V * AA' is the sum of the residual covariance matrix and the QTL covariance matrix.

If there is no recombination between marker and QTL, we can observe p. However, in case of recombination (r), p depends on r.

We can observe p if the marker is at the QTL (no recombination). Henshall and Goddard (1999) suggest that in case of recombination, the vector β can be estimated at each marker (as if it was the QTL), and the estimate for β at any position between two markers is obtained by linear interpolation. They also show how the log-likelihood can be calculated for any position of a QTL between two marked loci.

References

Henshall, J.M. and M.E. Goddard. 1999. Multiple trait mapping of quantitative trait loci after selective genotyping using logistic regression. Genetics 151:885-894.

Jiang, C., and Z-B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111-1117.

Moser, G. 2000. A simple regression method for multiple trait mapping of multiple QTL's. Submitted (Univ. of New England, Armidale)