Proc. Assoc. Advmt. Anim. Breed. Genet. Vol 14

MULTI-TRAIT ANALYSES TO DETECT MULTIPLE INTERACTING QUANTITATIVE TRAIT LOCI USING A GENETIC ALGORITHM

G. Moser and J.H.J. van der Werf

Department of Animal Science, University of New England, Armidale, NSW 2351

SUMMARY

Multiple trait mapping has not only proven to be more powerful than single trait analysis, but also allows testing of a number of biologically interesting hypotheses concerning the nature of genetic correlation between traits. A simple method is presented that combines multiple trait analysis and the generality of regression methods for fitting multiple and interacting QTL using an efficient search method based on a genetic algorithm.

Keywords: QTL, multiple traits, regression mapping, epistasis.

INTRODUCTION

Most QTL experiments include the measurement of multiple phenotypes. Several investigations have shown the advantages of multiple trait analyses (Korol *et al.* 1995; Jiang and Zeng 1995). These advantages include not only increased power to detect QTL and increased precision in mapping QTL, but also the ability to test a number of biologically interesting hypotheses concerning the nature of genetic correlation between traits.

Multiple trait analysis based on maximum likelihood methods is relatively complex and computationally demanding. Recently, simpler multiple trait QTL mapping methods based on least squares were developed (Moser 1998; Knott and Haley 2000). Moser and Van der Werf (2000) have shown that the least squares method gives very similar results to those obtained by multiple trait maximum likelihood.

Currently, most statistical methods for detecting multiple QTL such as composite interval mapping (CIM) (Zeng 1994) map individual QTL separately, since a multidimensional exhaustive search for several QTL is computationally not feasible. CIM combines interval mapping with multiple regression on selected markers to eliminate variation arising from other QTL elsewhere in the genome. A major limitation of CIM is that interactions between loci (epistasis) are ignored. A new approach to map multiple interacting QTL simultaneously was introduced by Carlborg *et al.* (2000) using a genetic algorithm (GA) as an efficient search strategy.

In this paper a method for the simultaneous mapping of multiple interacting QTL using information from multiple traits is presented. The approach uses a genetic algorithm as an efficient search strategy. We compare the method with the CIM approach. A simulation example of mapping multiple linked QTL for two correlated traits in a F_2 generation derived from a cross between outbred lines is presented to illustrate the method.

MATERIALS AND METHODS

Multiple trait tests. Following Moser (1998) the likelihood ratio test (LRT) in the multiple trait analysis using least squares can be approximated by:

$$LRT \approx n \log_e \frac{|V_{reduced}|}{|V_{full}|} \; ;$$

where $|V_{reduced}|$ is the determinant of the residual variance-covariance matrix of the reduced model, and $|V_{full}|$ the determinant of the model including the QTL. In the case of two traits, V is equal to

$$\begin{vmatrix} RSS_1 & RSP_{21} \\ RSP_{12} & RSS_2 \end{vmatrix} ;$$

where RSS_1 and RSS_2 are residual sums of squares of trait 1 and trait 2 respectively, and RSP_{12}, RSP_{21} are crossproducts of the residuals for trait 1 and trait 2. The residuals are obtained by performing a separate regression for each trait in turn and are then combined to form the joint test statistic. Hypotheses testing in multi-trait analysis can be performed as outlined by Jiang and Zeng (1995). Under a pleiotropic model a single QTL is at the same position in the two separate regressions on both traits. Fitting a single QTL at different positions for the two traits leads to a close linkage model.

One dimensional search for multiple QTL. Following Jiang and Zeng (1995) we have implemented CIM using the linear regression model for interval mapping and the multiple linear model to control for linked QTL. Selection of cofactors was based on single-trait analysis. Instead of the actual marker genotypes, QTL genotype probabilities at the marker locations flanking a putative QTL were used as cofactors. For the test pleiotropy *vs.* close linkage only cofactors outside the interval containing the putative closely linked QTL were fitted.

Simultaneous search for QTL using a genetic algorithm. The basic idea of a genetic algorithm is to have a population of solutions that improves relative to the objective through a "survival of the fittest" mechanism. As a stochastic function minimiser the Differential Evolution algorithm (DE) of Price and Storn (1997) was used. Briefly, the algorithm starts with a randomly generated population in which each individual is a vector containing values of the parameters of an objective function to be optimised. The parameters optimised by the genetic algorithm are the positions of the QTL. The objective function is equal to the denominator of the LRT. After initialisation, each vector is selected in turn as the target vector. DE generates new candidate solutions by adding the weighted difference between two randomly chosen population vectors to a third randomly chosen vector. If the resulting vector yields a higher LRT statistic than the target vector, the newly generated vector replaces the target vector in the next generation, otherwise the target vector is retained. The process is stopped when no improvement in fitness is observed in the last 500 evaluations of the fitness criteria.

Numerical example. One hundred replicates of a F_2 generation (n = 500) from a cross between outbred lines were simulated. Regression mapping was carried out using the method described by Haley *et al*. (1994). The genome consisted of one chromosome of length 150 cM with equally spaced markers in 10 cM intervals. Each marker had four alleles segregating at equal frequencies in the

Table 1. Simulated parameters and estimates of QTL positions and effects from composite interval mapping (CIM) and a genetic algorithm (GA) over 100 replicates of simulation using multiple trait analysis

		Additive- and Dominance effects			
	Position	Additive		Dominance	
QTL	(cm)	Trait1	Trait2	Trait1	Trait2
		Simulated parameter values of marginal QTL effects [A]			
1	32.0	-1.50	-1.60	0.90	1.00
2	87.0	-1.10		0.40	
3	94.0		1.20		0.60
4	128.0	1.30		-0.40	
		No epistasis fitted			
		Estimates by CIM [B]			
1	31.3 ± 2.1	-1.41 ± 0.40	-1.49 ± 0.39	1.07 ± 0.35	1.13 ± 0.37
2	87.3 ± 5.8	-1.05 ± 0.36		0.53 ± 0.38	
3	93.4 ± 5.4		1.34 ± 0.33		0.70 ± 0.38
4	129.1 ± 3.7	1.43 ± 0.36		-0.38 ± 0.38	
		Estimates by GA [B]			
1	31.5 ± 1.6	-1.39 ± 0.15	-1.47 ± 0.15	0.98 ± 0.21	1.09 ± 0.17
2	87.2 ± 3.6	-0.98 ± 0.19		0.48 ± 0.21	
3	93.8 ± 2.4		1.34 ± 0.15		0.68 ± 0.22
4	128.4 ± 2.6	1.34 ± 0.16		-0.41 ± 0.23	
		Epistasis fitted			
		Estimates by GA [B]			
1	31.7 ± 1.6	1.50 ± 0.25	-1.62 ± 0.23	0.90 ± 0.37	0.99 0.32
2	87.2 ± 4.0	-1.12 ± 0.32		0.39 ± 0.38	
3	93.4 ± 2.0		1.20 ± 0.25		0.57 ± 0.33
4	128.6 ± 3.1	1.34 ± 0.18		-0.41 ± 0.22	
		Interaction effects [B]			
AA		0.39 ± 0.28	0.42 ± 0.26		
AD		0.34 ± 0.32	0.40 ± 0.30		
DA		0.39 ± 0.36	0.42 ± 0.36		
DD		0.43 ± 0.48	0.46 ± 0.47		

[A]Epistatic effects (*aa*,*ad*,*da*,*dd*) were simulated between QTL1 and QTL2 for trait1 and between QTL1 and QTL3 for trait2 and accounted for less than 10% of the genetic variance.
[B]Estimates are means ± standard deviations.

founder breeds. Two pleiotropic QTL (QTL1, QTL4) affecting two traits and two closely linked QTL (QTL2, QTL3) each affecting a single trait were simulated with main effects given in Table 1. Epistatic effects were simulated between QTL1 and QTL2 for trait 1 and between QTL1 and QTL3 for trait 2 using the common model for two locus interactions (Mather and Jinks 1982). Epistatic effects were of magnitude 0.4 and accounted for less than 10% of the genetic variance. Broad-sense heritability was 0.4 and genetic and phenotypic correlation between the two traits was 0.54 and 0.35, respectively.

RESULTS AND DISCUSSION

Not accounting for interaction effects in the statistical model resulted in biased estimates of QTL effects for both methods, but estimates of the positions of QTL were close to the simulated values (Table 1). Estimates obtained with the GA method were more accurate. Accounting for epistasis in the GA method removed the bias of the estimates of main QTL effects, but estimates of the interaction terms were not very accurate. Sampling variances of estimates of main effects were also consistently larger when epistasis was included in the model. The percentage of replicates in which QTL2 and QTL3 were detected in different marker segments and the close linkage hypothesis was favoured over the pleiotropic hypothesis was low. Power to detect close linkage was 34% for CIM and 52% for GA.

A prerequisite for detecting epistasis is to simultaneously model QTL. As a multidimensional exhaustive search for multiple interacting QTL using information from several traits is computationally not feasible, a GA-based search method seems to be the method of choice in most situations. Even for the relatively small problem of mapping QTL for two traits in a 150 cM genome, using a genetic algorithm increased the computational efficiency by a factor of more than 10^6. Such a computationally efficient approach is required in the analyses of real data and when permutation testing and bootstrap methods are used to determine significance thresholds and confidence intervals.

In summary, a simple least squares method is presented that allows mapping of multiple and multiple interacting QTL for several traits simultaneously. The complexity of the method does not increase greatly with the number of parameters estimated, as it does in multiple trait maximum likelihood analysis. The method can easily be implemented for a wide spectrum of designs such as back- and intercrosses between inbred lines and in F_2- and half-sib families in outbred populations.

REFERENCES

Carlborg, Ö., Andersson, L. and Kinghorn, B. (2000) *Genetics* **155**: 2003.
Haley, C.S., Knott, S.A. and Elsen, J.-M. (1994) *Genetics* **136**: 1195.
Jiang, C. and Zeng, Z.-B. (1995) *Genetics* **140**: 1111.
Korol, A.B., Ronin, Y.I. and Kirzhner, V.M. (1995) *Genetics* **140**: 1137.
Knott, S.A. and Haley, C.S. (2000) *Genetics* **156**: 899.
Mather, K.M. and Jinks, J.L. (1982) "Biometrical Genetics" 3rd ed. Chapman & Hall, London.
Moser, G. (1998) *45th Annual meeting of the Genetics Society of Australia*, 9-12 October, Sydney.
Moser, G. and Van der Werf, J.H.J. (2000) *Book of Abstr. of the EAAP*, p5, 21-24, The Hague.
Price, K. and Storn, R. (1997) *Dr. Dobb's Journal*, **April**: 18.
Zeng, Z. -B. (1994) *Genetics* **136**: 1457.