QTL DETECTION USING MULTIPLE REGRESSION ON AVERAGE TRANSMISSION PROBABILITIES

J. M. Henshall[1] and M. E. Goddard[2]

[1] Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351
[2] Institute of Land and Food Resources, University of Melbourne, Parkville, Vic 3052

SUMMARY

A method of applying multiple regression to marker data is described. Phenotype is regressed on QTL transmission probability for regions of the genome. The QTL transmission probabilities combine the information from multiple linked markers, and are likely to be more uniform in the proportion of informative offspring than single markers.

Keywords: QTL, multiple regression

INTRODUCTION

Experiments to map quantitative trait loci (QTL) commonly compare the genotype at many marker loci with a smaller number of traits. QTL are mapped to positions between markers using maximum likelihood (Lander and Botstein 1989) or regression (Haley and Knott 1992) methods. Not uncommonly there are chromosomal regions in which QTL are suspected to map and the aim of the experiment is to confirm the existence of the QTL and their segregation in the current families. Such experiments use a number of markers from the relevant region of the chromosome but, unless the number of animals used is very large, they have little power to estimate the position of the QTL amongst the markers. Analysis methods which incorporate the information from multiple markers to estimate QTL effect and significance, without inferring location relative to the markers, may be ideal for such experiments.

A multiple regression of phenotype on marker genotype has been considered for QTL, and has some advantages over single marker or marker bracket methods. The method is quick, multiple unlinked QTL are accounted for, problems with multiple testing are reduced, and computer software is readily available. However, a major disadvantage with the multiple regression approach to mapping QTL is the requirement that the independent variable, marker genotype, be available for all records included in the analysis. For marker genotype to be available for all markers for all animals, a marker must be fully informative for all animals. While this is theoretically possible for designed experiments, it does impose severe restrictions on which markers can be used. One method of avoiding the requirement that all animals are fully informative for all markers is to regress on a function of marker genotype, such as the probability of a particular marker allele being inherited. However, markers are unlikely to be equally informative, so markers may be identified as significant or not significant based on the proportion of informative animals.

* AGBU is a joint institute of NSW Agriculture and The University of New England

In this article, a method of combining information from a number of linked markers is described, to produce a QTL transmission probability (QTP). The multiple regression of phenotype on QTP is proposed as a method that retains the advantages of multiple regression, while minimising the disadvantages caused by uninformative markers.

METHOD

With interval mapping, for each animal, the observed marker transmission, assumed marker and QTL locations and the mapping function determine a QTL transmission probability at points (loci) on the chromosome. An alternative approach is to use a single QTL transmission probability for a region of the chromosome containing a number of linked markers. Such a QTP is easily obtained by integrating the mapping function over the region, and dividing by the map length. With Haldane's mapping function (Haldane 1919) this integral is best expressed in a mixture of units of map distance (d) and recombination rate (r). Let a sire have marker genotype *AaBb* for two linked markers, where A and B are on one chromosome of the homologous pair, and a and b on the other. Offspring can inherit AB, Ab, aB, or ab. If a QTL is located between the markers, then let q be the probability that the QTL allele on the AB chromosome was inherited by an offspring. Then, with Haldane's mapping function, and assuming that all locations within the bracket are equally likely for the QTL, and that $r > 0$,

$$(q \mid AB) = \frac{1}{2} + \frac{1}{2(1-r)d}$$

$$(q \mid ab) = \frac{1}{2} - \frac{1}{2(1-r)d}$$

$$(q \mid Ab) = (q \mid aB) = \frac{1}{2}$$

Additional probabilities are required for the situation where one of the markers is uninformative.

$$(q \mid A-) = (q \mid -B) = \frac{1}{2} + \frac{r}{2d}$$

$$(q \mid a-) = (q \mid -b) = \frac{1}{2} - \frac{r}{2d}$$

For chromosome regions containing more than two markers, the transmission probabilities for marker brackets can be averaged, weighted by the map lengths of the brackets. Given a QTP for a region, maximum likelihood or regression approaches can be used to estimate QTL effects and significance levels. If markers are available for multiple regions, on more than one chromosome, then all markers can be combined in a single analysis by performing multiple regression on the QTP. As each QTP incorporates information from multiple markers, missing genotypes and unbalanced marker information should be less of a problem than with ordinary multiple regression.

Test Data To demonstrate the method, a single sire, halfsib QTL detection experiment was simulated. 60 markers on 20 chromosomes were simulated for 100 progeny of the sire. A record was simulated for each halfsib with phenotypic variance of 1.0, of which approximately 20 % was

due to QTL for which the sire was heterozygous. Each of the chromosomes contained one QTL, with the variance due to the QTL on the first chromosome 0.1, the second 0.05, the third 0.025 and so on, down to the tenth chromosome with QTL variance 0.00019. For chromosomes 11 to 20 the QTL variance was zero. All QTL-marker pairs were simulated with a recombination rate of 0.05. A third marker was located with a recombination rate of 0.1 from one of the two flanking markers. All markers were informative half the time, but within markers, alleles were not equally informative. The marker allele linked to the positive (negative) QTL effect was simulated to be informative for 50 % (50 %), 90 % (10 %) and 30 % (70 %) of offspring. The QTL was located between the first two markers on even numbered chromosomes, and the last two markers on odd numbered chromosomes.

Two analysis methods were applied to the simulated data. For both analyses, SAS procedure REG (SAS 1990) was used, with the stepwise option selected. In the first analysis, the phenotype was regressed on QTP, with QTP obtained as described above. Some examples of the QTP are displayed in Table 1. In the second analysis, the phenotype was regressed on marker genotype, with marker genotype coded −1 or +1 for the two sire marker alleles, and 0 where the marker was uninformative. From the output from SAS procedure REG, the most significant marker on each chromosome was identified. Both analyses were performed on 200 simulated data sets.

Table 1. Examples of observed marker genotype and estimated QTL transmission probability (QTP). For each animal, the genotype at three linked markers A, B, and C is shown with a dash where the marker is uninformative

Animal	Markers	QTP
1	---	0.5000
2	--C	0.9033
3	A-C	0.9919
4	-b-	0.0519
5	ab-	0.0270
6	--c	0.0967
8	-BC	0.9730

RESULTS AND DISCUSSION

A summary of the results obtained appears in Table 2. When a nominal $p<0.05$ significance level was used in the regression on markers, apparently significant (ie false positive) QTL were detected on chromosomes 11 – 20 in much more the 5 % of the simulations. When a nominal $p<0.005$ significance level was used the actual type 1 error rate was close to 5 % (on a chromosome-wide basis). In the regression on QTP, use of a nominal $p<0.05$ significance level resulted in approximately 5 % false positive QTL detected on chromosomes 11-20. When using the more stringent significance level in the marker regression, the power to detect QTL on chromosomes 1-4 was less than the power of the regression on QTP. Neither method had a high power to detect the small QTL on chromosomes 5-10. The estimated size of the QTL effect, in the replicates in which a significant effect was observed, overestimated the true effect, especially in the method using regression on the markers.

From these results it appears that a stepwise multiple regression of phenotype on QTP may be useful in the detection of QTL in livestock. A logical extension to the method is to perform interval mapping on one region, using the results from the multiple regression to adjust for QTL effects in other regions. The method would then be similar to those of Zeng (1994) and Jansen (1994), but should be less affected by unequal proportions of informative animals for each marker.

Table 2. Results from stepwise regression analysis of 60 markers on 20 chromosomes. Phenotype was regressed on QTP, and on single markers with the most significant marker on each chromosome stored (Markers). Shown are the percenrtage of replicates that were significant at the 0.05 or 0.005 level, and the mean estimated effect (a) for significant replicates

Chrom		QTP		Markers			
-osome	Effect	%sig(0.05)	a	%sig(0.05)	a	%sig(0.005)	a
1	0.4472	96.0	0.469	95.0	0.569	71.5	0.638
2	0.3162	78.5	0.363	84.5	0.460	52.0	0.525
3	0.2236	49.5	0.327	65.5	0.435	36.5	0.502
4	0.1581	27.5	0.318	46.5	0.328	22.5	0.382
5	0.1118	16.5	0.294	36.0	0.317	17.5	0.422
6	0.0791	17.5	0.249	38.5	0.273	12.5	0.300
7	0.0559	11.5	0.182	38.5	0.099	14.0	0.104
8	0.0395	8.5	0.155	34.0	0.064	16.0	0.084
9	0.0280	6.5	0.107	33.0	0.068	13.0	0.153
10	0.0198	7.0	0.234	31.0	0.061	10.5	0.041
11	0.0000	6.0	0.097	24.5	0.065	8.0	0.051
12	0.0000	3.5	0.048	18.0	0.026	8.5	0.084
13	0.0000	7.0	-0.006	33.5	0.015	16.5	-0.022
14	0.0000	6.5	-0.112	32.5	-0.027	10.0	-0.148
15	0.0000	5.5	-0.089	30.0	-0.020	12.0	-0.093
16	0.0000	6.0	-0.135	25.0	-0.028	11.0	-0.039
17	0.0000	8.0	0.004	28.0	0.009	7.0	0.066
18	0.0000	8.0	-0.038	28.5	-0.007	12.5	0.033
19	0.0000	5.5	0.039	26.5	0.034	8.0	0.315
20	0.0000	7.5	0.017	30.5	-0.058	14.5	0.001

REFERENCES

Haldane, J. B. S. (1919) *J. Genet.* **8**:299
Haley, C. S. and Knott, S. A. (1992) *Heredity* **69**:315
Jansen, R. C. (1994) *Genetics* **138**:871
Lander, E. S. and Botstein, D. (1989) *Genetics* **121**:185
SAS (1990) "SAS/STAT User's Guide, Version 6", 4th ed. SAS Institute, Cary
Zeng, Z. (1994) *Genetics* **136**:1457