QTL MAPPING USING LOGISTIC REGRESSION

Yuandan Zhang and Bruce Tier
Animal Genetics and Breeding Unit*
Universality of New England, Armidale NSW 2350

SUMMARY
This study compares logistic regression (LR) with maximum likelihood (ML) for mapping quantitative trait loci (QTL) in half-sib family with respect to genotyping schemes (full and selective), various levels of marker informativeness and marker interval. Under selective genotyping, the power of ML is limited in regions with low information contents. In this case, LR performed better than ML and provides a straightforward and robust solution to such situation.
Keywords: regression, quantitative trait loci, maximum likelihood, markers.

INTRODUCTION
There are many methods available for detecting QTL in half-sib populations when all individuals are genotyped (eg Zeng 1993, 1994; Kao *et al.* 1999*;* Kerr *et al*. 2005). These methods generally rely on good genetic information, that is, polymorphic markers spaced evenly at close intervals along the chromosome. This ideal case is uncommon in real genome scans, because of a lack of informative markers in some chromosomal regions. Selective genotyping is used to reduce the cost of genotyping and achieve a reasonable power of detecting QTL (Lebowitz *et al.* 1987). Low amounts of genetic information, selective genotyping or a spike in the phenotypic distribution can generate spurious QTL peaks (eg Broman 2003; Freenstra and Skovgaard 2004). Using simulated data, this study compares logistic regression (LR) with the maximum likelihood (ML) for detecting QTL with various levels of marker informativeness, marker interval and genotyping schemes (full or selective).

METHODS
Data. Twelve half-sib families (5 for selective genotyping and 7 for full genotyping) were simulated and analysed to examine the performance of LR and ML. One chromosome was considered for each family. Each progeny had a different dam. Genetic information in each family is shown in Table 1. Phenotypic residuals were drawn from a normal distribution with a mean of zero and variance of 10. The QTL effect (difference between two homozygous genotypes) was assumed to be one phenotypic standard deviation (3.16).

For selective genotyping (SA, SB, SC, SF, and SG cases), 100 individuals were simulated and the 25 individuals with highest and 25 with lowest phenotypic values were genotyped and analysed. Fifty individuals were simulated and genotyped in the full genotyping cases. For each case, 100 replicates were generated and analysed using both LR and ML. The chromosome was searched in steps of 2 cM. A chromosome wide significant threshold value was obtained using 800 or 1000 permutations across each test point for each analysed chromosome using LR or ML, respectively. The power was defined as a percentage of replicates in which a chromosome wide significance level of 5% was

* AGBU is a joint venture of NSW Department of Primary Industries and the University of New England.

achieved. The QTL position was identified by the test location with the highest test statistic. The average QTL position and effect over all replicates where a 5% chromosome wide significance was achieved were calculated as the QTL position and effect for each analysis method.

Table 1. Genetic information used in each family, number (N), position (cM) and information content range (MIC) of markers, minimum MIC with position in parenthesis. Simulated QTL position is also shown

Family[1]	N	Marker Position (cM)	QTL Position	MIC Range	Min MIC (cM)
A, SA	6	0, 26, 51, 77, 102, 128	64	0.9-1.0	0.75
B	6	0, 18, 30, 76, 102, 127	64	0.9-1.0	0.65(54)
SB	6	0, 9, 20, 76, 102, 127	64	0.9-1.0	0.49(48)
C	6	0, 26, 51, 97, 109, 127	64	0.9-1.0	0.65(74)
SC	6	0, 26 51, 108, 119, 128	64	0.9-1.0	0.49(80)
D	6	0, 5 , 56, 77, 103, 128	67	0.9-1.0	0.63(30)
E	6	0, 25, 51, 72, 123, 128	64	0.9-1.0	0.63(98)
F, SF	5	0, 81, 106, 132, 157	93	0.9-1.0	0.33(40)
SG	5	0, 81, 106, 132, 157	93	0.75-0.85	0.29(38)
G	6	0, 26, 51, 77, 128, 133	64	marker 4 0.5-0.6 others 0.9-1.0	0.44(96)

1 Family A – G for fully genotyping, SA, SB, SC, SF, SG for selectively genotyping.

Marker information content. Paternal haplotypes were estimated using the Lander-Green algorithm (Lander and Green 1987). Prior QTL transmission probabilities (TP) for each individual at test points were estimated using genotypes and the recombinant fragment of flanking markers, as described by Kerr *et al*. (2005). The variation of TP is used as an indicator of the marker information content (MIC) and is measured as 4 times of the TP variance, ranging from 0.0 to 1.0.

Statistic model. Assume a bi-allelic QTL with alleles Q and q contributes to the variance of a quantitative trait. The quantitative trait value y_i for individual i, being pre-adjusted for fixed effects and polygenic effects, can be related to the QTL by the model $y_i = \mu + \alpha x_i + e$, where α is the effect of the putative QTL, x_i is a probability if a progeny inherits allele Q from the sire and e is the residual. The logistic regression model is fitted and implemented as described by Dobson (2002). The estimated LR coefficient and the total variance were used to calculate the QTL effect α, as described by Henshall and Goddard (1999). The ML method is implemented as described by Kerr *et al.* (2005), the α value is estimated as described by Zeng (1994). In the case of selective genotype, the α values from ML are adjusted for the consequences of selective genotyping, using the method of Darvasi and Soller (1992). Empirical threshold values were determined using permutation tests (Churchill and Doerge 1994).

RESULTS AND DISCUSSION

The results are presented in Table 2. Generally, under selective genotyping, LR showed higher power to detect QTL and more accurate QTL location than did ML. LR conservatively estimated QTL effects, while ML highly overestimated these effects. In full genotyping, both LR and ML had similar power to detect QTL and revealed similar QTL positions. QTL effects estimated from LR were close to the simulated value while these effects from ML were slightly overestimated. The selective genotyping demonstrated a higher power in detecting QTL over the full genotyping using both LR and ML.

Table 2. Results of QTL analyses using Logistic Regression and Maximum Likelihood methods in the selective genotyping and full genotyping schemes. The power of each analysis, QTL position (cM) and effect (simulated at 3.16) are presented, with standard error of means in parenthesis

	Logistic Regression			Maximum Likelihood		
Family	Power[1]	Position	Effect	Power	Position	Effect
Selective Genotyping						
SA	96	63(1)	2.71(0.06)	90	62(1)	3.58(0.11)
SB	82	73(1)	2.71(0.06)	83	54(0)	4.81(0.04)
SC	90	54(1)	2.80(0.06)	88	72(0)	4.85(0.04)
SG	97	95(1)	2.51(0.05)	84	58(2)	4.75(0.05)
SF	94	95(1)	2.82(0.06)	74	57(1)	4.82(0.05)
Full Genotyping						
A	69	62(2)	3.24(0.06)	70	63(2)	3.72(0.07)
B	71	54(3)	3.23(0.07)	71	54(3)	3.81(0.07)
C	70	71(3)	3.21(0.06)	70	74(3)	3.82(0.07)
D	70	62(3)	3.23(0.06)	70	63(3)	3.76(0.07)
E	70	62(3)	3.24(0.06)	69	63(2)	3.73(0.07)
F	70	98(2)	3.18(0.06)	67	95(1)	3.81(0.08)
G	57	53(2)	3.16(0.07)	60	63(2)	4.11(0.10)

1 percentage of 100 replicates with a 5% chromosome wide significance level.

Selective Genotyping. When markers were highly informative and evenly spaced, both LR and ML displayed high power and accuracy for QTL locations and effects (SA). When an interval containing the QTL was extended to 57 cM, QTL positions assessed by both methods deviated from the true position. The position assessed by ML was close to the point of the lowest information content, while that from LR was in the same region but alway from the lowest point (SB & SC). When a large marker interval (up to 80 cM) occurred next to the interval containing the QTL, LR performed better

than ML in positioning QTL and power to detect QTL (SF). When marker information content was low (0.6 – 0.8, 0.29 at the lowest point) and with a large marker interval neighbouring the QTL region, LR also produced better estimates of QTL location and effect than those from ML (SG).

Chromosomal regions of low MIC frequently occur in QTL projects. Those regions are attributable to a lack of informative markers. ML can produce spurious peaks of test statistic in low information regions under selective genotyping. Freenstra and Skovgaard (2004) developed a 2-component mixture model to avoid such spurious peaks and applied to a back-cross design. To apply this method to F2 or half-sib design, a 3-component model is required and is yet to be developed. The LR method can overcome such problems and facilitate the efficient application of selective genotyping.

Under the full genotyping schemes, both LR and ML produced similar estimates for the parameters of interest. However, when MIC of a marker which anchored the QTL was low, QTL position from LR deviated from the true position by about 10 cM (G), while ML overestimated the QTL effects. With informative markers and at even spacing (A), both LR and ML showed similar power and revealed unbiased QTL locations and effects. Kerr *et al.* (2005) found when family size increased from 25 to 200, the power of ML could be enhanced from 34 up to 100. However, it is uncommon to identify such balanced cases in many genome scan studies.

CONCLUSION

Under selective genotyping, the power of the ML method is limited in low information regions. In this case, LR performed better than ML in detecting QTL. The LR method provides straightforward and robust solution to such situation.

REFERENCES

Broman, K. W. (2003) *Genetics* **163**:1169.
Churchill, G. A. and Doerge, R. W. (1994) *Genetics* **138**:963.
Darvasi, A. and Soller, M. (1992) *Theor. Appl. Genet.* **85**:353.
Dobson, A. J. (2002) "An Introduction to Generalized Linear Models" 2nd ed. Chapman and Hall, London
Feenstra, B. and Skovgaard, Ib. M. (2004) *Genetics* **167**: 959.
Henshall, J. M. and Goddard, M. E. (1999) *Genetics* **151**: 885.
Kao, C.-H., Zeng, Z.-B. and Teasdale, R. D. (1999) *Genetics* **152**: 1203.
Kerr, R. J., McLachlan, G.M. and Henshall J.M. (2005) *Genet. Sel. Evol*. **37** 83.
Lander, E. S. and Green, P. (1987) *Proc. Natl. Acad. Sci. USA* **84**:2363.
Lebowitz, R. J., Soller, M. and Beckmann, J. S. (1987) *Theor. Appl. Genet.* **73**:556.
Zeng, Z.-B. (1993) P*roc. Natl. Acad. Sci. USA* **90**:10972.
Zeng, Z.-B. (1994) *Genetics* **136**:1457.