ANALYSIS OF LINKAGE BETWEEN GENETIC MARKERS AND QTL USING REML: THE EFFECT OF SELECTION ON PARAMETER ESTIMATES

B.E. Clarke[1,2], J.A.M. van Arendonk[3] and B.P. Kinghorn[1]

[1]Department of Animal Science, University of New England, Armidale, NSW 2351

[2]CSIRO Division of Animal Production, Armidale, NSW 2350

[3]Department of Animal Breeding, Wageningen Agricultural University, 6700 AH Wageningen, The Netherlands

SUMMARY

The effects of various forms of selection in a granddaughter design on the detection of linkage between a quantitative trait locus (QTL) and genetic markers were examined. A restricted maximum likelihood method was used which utilised information from flanking markers and accounted for unlinked QTL. Selection of grandsires had little effect on the detection of linkage and parameter estimates. However, selection of sires within grandsire families had a large effect on linkage detection and resulted in large decreases in estimated additive genetic variance. This is due to the generation of a negative covariance between polygenic and QTL effects in the selected individuals.

Keywords: Restricted maximum likelihood, linkage analysis, quantitative trait locus

INTRODUCTION

Linkage between genetic markers and quantitative trait loci (QTL) may be detected using a restricted maximum likelihood (REML) technique. This method relies on the use of flanking marker brackets and involves use of mixed model equations which incorporate genetic marker information (van Arendonk *et al.* 1994). Derivative-free REML techniques, which are developed for animal breeding applications (Smith and Graser 1986; Graser *et al.* 1987), can be used to detect QTL-marker linkage and to estimate position of the QTL relative to the bracketing markers and the size of the QTL effect.

Detection of linkage between genetic markers and QTL requires experimental populations in which animals are genotyped for the marker loci and phenotypic observations are recorded. The granddaughter design (Weller *et al* 1990), is a popular design in dairy cattle for linkage detection experiments because of the reduced number of genotypings required compared to the alternative daughter design. Most analyses of QTL-marker linkage detection assume that no selection has occurred on any of the animals. In practice, however, the granddaughter design is actually a good example of where selection can severely bias experimental samples (Mackinnon and Georges 1992).

Mackinnon and Georges (1992) found that the proportion of animals selected had the most influence on the power to detect linkage when the effects of within family selection were studied. Their analyses involved stochastic simulation which ignored polygenic variance. The purpose of

this study was to examine the effects of selection at various points in a granddaughter designed population on the detection of QTL-marker linkage using the REML detection method.

METHOD

An outbred population was simulated with the assumption of a QTL in linkage equilibrium with all genetic markers in the base population. Linkage disequilibrium was generated within families through linkage between genetic markers and the QTL. The granddaughter design (Weller *et al.* 1990) was used with a number of grandsires having several half sib offspring (called sires), each of which was mated to several unrelated dams to produce half sib grandoffspring. Marker genotypes were obtained for the grandsires and sires but not for the grandoffspring or granddams.

Phenotypic expression as recorded in progeny of the sires is due to the effect of the QTL, a random normal polygenic effect and a random normal residual effect. In matrix notation the mixed model equations (from van Arendonk *et al.* 1994) can be written as:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z+A_u^{-1}\alpha_u & Z'W \\ W'X & W'Z & W'W+G_v^{-1}\alpha_v \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}$$

where: **X, Z** and **W** are known incidence matrices, **y** is a vector of phenotypic observations, **b** is a vector of fixed effects, **u** is a vector of random additive genetic effects due to polygenic effects, i.e. unlinked QTL not linked to the genetic markers, **v** is a vector of random additive gametic effects at the marked QTL, **e** is a vector of random residual effects, $\mathbf{A_u}$ is the numerator relationship matrix for unlinked QTL, $\mathbf{G_v}$ is the gametic relationship matrix for the marked QTL and $\alpha_u = \sigma_e^2 / \sigma_u^2$, $\alpha_v = \sigma_e^2 / \sigma_v^2$.

Following van Arendonk *et al.* (in preparation) the likelihood can be constructed as:

$$\log L = -\tfrac{1}{2}(\text{const} + N_e \log \sigma_e^2 + N_A \log \sigma_a^2 + N_G \log \sigma_v^2 + \log|A_u| + \log|G_v| + \log|C^*| + y'Py/\sigma_e^2)$$

where: $\mathbf{C^*}$ is the coefficient matrix with **X** replaced with a full rank submatrix $\mathbf{X^*}$, **N** is the number of observations, $\mathbf{NF^*}$ is the rank of **X**, $\mathbf{N_A}$ is the number of animals, $\mathbf{N_G}$ is the number of gametic effects, $N_e = N - NF^* - N_A - N_G$, and **P** is a matrix: $P = V^{-1} - V^{-1}X^*(X^{*'}V^{-1}X^*)^{-1}X^{*'}V^{-1}$, where: $V = (ZAZ'r + I)$ and $r = \sigma_a^2 / \sigma_e^2$.

From Graser *et al.* (1987) the error variance can be estimated directly from the residual sum of squares in a univariate analysis with identical and independently distributed errors as: $\sigma_e^2 = y'Py/(N - NF^*)$. The term y'Py depends on the other parameters to be estimated, and its value can be determined through the use of Gaussian elimination. The determinant of $\mathbf{C^*}$ can also be obtained during the elimination process. Instead of maximising *logL* as a function of the four parameters σ_e^2, σ_u^2, σ_v^2, and recombination rate (θ) these can be reparameterised to σ_e^2, α_u, α_v and θ. Estimation is then carried out in two steps - first the likelihood is maximised with respect to α_u, α_v and θ and secondly σ_e^2 is obtained. A likelihood ratio (LR) is calculated to confirm detection

of linkage between genetic markers and putative QTL. This is the ratio of *logL* (assuming linked QTL) over *logL* maximised assuming that the QTL is unlinked to the genetic markers.

The simulated pedigree involved 50 grandsires, 40 sires and 50 daughters per sire. A QTL was positioned 75% of the way along the marker bracket of length 0.3 Morgans. The QTL explained 10% of the total additive genetic variance in a trait with a heritability of 0.3, including QTL effect. 250 replicate populations were simulated and results averaged over all populations. Selection took place among the grandsires, granddams and sires. Grandsires and sires were progeny tested and selected on their daughter's performance. With no selection 50 grandsires were chosen at random from a base of 2000, and with selection they were chosen from 2000 (2.5% selected) or 5000 (1% selected). Granddam selection involved 2000 granddams selected on the basis of their own phenotypes from a base of 20000. Selection of sires was on the average performance of their daughters and was at 50%. Number of grandsires and sires included in the analysis was the same for all selection strategies.

RESULTS

Linkage between flanking markers and a putative QTL was detected in the simulations involving no selection, as evident by a high LR. The estimates of other parameters from the maximisation of *logL* were close to the simulated values (Table 1). The effect of increased selection intensity among the grandsires (Selection Options 2 and 3) was minor. Results of selection at different points within a granddaughter design show that the most marked changes in LR and parameter estimates were found when selection was carried out in the sires (Selection Options 5 and 6).

Table 1. Estimated parameters and variance components under different selection options

Parameter	True	Selection Options*					
Estimates	Values	1	2	3	4	5	6
Likelihood	-	5.8989	5.4454	5.4190	6.0701	1.3795	1.5303
ratio		(0.2596)	(0.2199)	(0.2490)	(0.2260)	(0.1245)	(0.1135)
Heritability	0.30	0.2997	0.2924	0.2916	0.2734	0.0523	0.0775
		(0.0007)	(00008)	(0.0007)	(0.0007)	(0.0004)	(0.0004)
QTL effect	0.10	0.1043	0.0970	0.0950	0.1063	0.1159	0.1162
		(0.0030)	(0.0026)	(0.0029)	(0.0026)	(0.0063)	(0.0071)
QTL	0.75	0.7286	0.7407	0.7300	0.7653	0.6425	0.6893
position		(0.0171)	(0.0183)	(0.0180)	(0.0162)	(0.0174)	(0.0221)
σ_e^2	70.0	69.9997	70.6464	70.7124	72.1866	88.7724	88.8382
		(0.0654)	(0.0664)	(0.0666)	(0.0649)	(0.0379)	(0.0408)

*Selection Options: 1 = no selection, 2 = grandsire selection (2.5%), 3 = grandsire selection (1.0%), 4 = grandsire (2.5%) and granddam (10%) selection, 5 = grandsire (2.5%), granddam (10%) and sire (50%) selection and 6 = sire (50%) selection only.

Variance components can be calculated according to the following relationships: $\sigma_a^2 = 2\sigma_v^2 + \sigma_u^2$, $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$, $\sigma_v^2 = 0.5*(\text{QTL effect})*\sigma_a^2$, $\sigma_u^2 = (1 - \text{QTL effect})*\sigma_a^2$ and $h^2 = \sigma_a^2 / \sigma_p^2$. The small effect of grandsire selection reflects that most of the information for estimating heritability comes from

within grandsire families and not from variation between grandsire families (which is reduced). This also explains why selection of sires has a much more severe effect on the heritability estimate.

DISCUSSION

The REML detection method was able to identify QTL linked to the genetic markers under a model involving no selection. With a single marker, van Arendonk *et al.* (1993) observed differences in separately estimating position and effect of a QTL when information from a single marker was used. In the present paper, this problem is solved by using information from a marker bracket. The advantages of the REML method over other detection methods such as regression (Weller *et al.* 1990; Haley and Knott 1992) and maximum likelihood (Lander and Botstein 1989) is that it is able to account for genetic variation due to QTL not linked to the genetic markers. The regression of Meuwissen and Goddard (1996) will handle unlinked QTL, but has not yet been tested for QTL linkage mapping. Markov Chain Monte Carlo techniques, such as Gibbs sampling, are also able to fit these polygenic effects, but are more computationally demanding.

Factors affecting the ability of linkage to be detected in the presence of selection were selection intensity and the point of selection within the pedigree. In all practical designs grandsires and granddams are selected and this selection cannot be avoided in collecting information. The results in Table 1 show that these forms of selection do not have a big impact on the results. Selection within grandsire families, however, has a major impact. This form of selection can be avoided by collecting DNA from all sires which are progeny tested.

Results from the selection of sires within grandsire families illustrate the expected effect (Bulmer, 1971) of truncation selection to generate a negative covariance between the QTL and polygenic effects. That is, within family (or sire) selection causes variation due to polygenes to be confounded with QTL genetic effects. The QTL variance and LR clearly show that in the presence of within family selection large underestimates of variance are given in linkage detection analysis.

ACKNOWLEDGMENTS

This work was carried out with the assistance of a Wageningen Agricultural University Fellowship, financial support was also provided by a CSIRO/UNE Postgraduate Research Scholarship.

REFERENCES

Bulmer, M.G. (1971) *Am. Nat.* **105**: 201
Graser, H-U., Smith, S.P. and Tier, B. (1987) *J. Anim. Sci.* **64**:1362
Haley, C.S. and Knott, S.A. (1992) *Heredity* **69**:315
Lander, L.S. and Botstein, D. (1989) *Genetics* **121**:185
Mackinnon, M.J. and Georges, M.A.J. (1992) *Genetics* **132**:1177
Meuwissen, T.H.E. and Goddard, M.E. (1996) *Genet. Sel. Evol.* **28**: 161
Smith, S.P. and Graser, H-U. (1986) *J. Dairy Sci.* **69**:1156
van Arendonk, J.A.M., Tier, B. and Kinghorn, B.P. (1993) *Proc 17th Int. Congr. Genet., Birmingham, UK* (192)
van Arendonk, J.A.M., Tier, B. and Kinghorn, B.P. (1994) *Genetics* **137**:319
Weller, J.I., Kashi, Y. and Soller, M. (1990) *J. Dairy Sci.* **73**:2525