UNBIASED ESTIMATES OF VARIANCES DUE TO QTL DETECTED BY SELECTIVE GENOTYPING

A.J. Chamberlain[1], Lu.T. Duc[1], P.J. Bowman[2], M.E. Goddard[1,2]
[1] Institute of Land and Food Resources, University of Melbourne, Victoria 3000
[2] Victorian Department of Primary Industries, Victorian Institute of Animal Science, Attwood 3049

SUMMARY
Selective genotyping increases the power to detect QTL but leads to biased estimates of their effects on the selected trait and on correlated traits. We describe two methods to obtain unbiased estimates of the genetic variance due to a QTL and illustrate these with an analysis of milk production data from a large QTL mapping experiment in dairy cattle. The first method corrects estimates obtained by regression using only the genotyped daughters of a sire. The correction is based on the ratio of variance in the genotyped daughters to variance in all daughters. Unbiased estimates for the QTL effect on correlated traits can be obtained by expressing these traits as a linear combination of the selected trait and traits that are uncorrelated with the selected trait. The second method uses a variance component analysis and includes all the daughters of each sire.

INTRODUCTION
It has been demonstrated that marker assisted selection (MAS) can result in increased rates of genetic gains (Gimelfarb and Lande, 1994; Lande and Thompson, 1990; Mackinnon and Georges, 1998; Meuwissen and Goddard, 1996). These gains are particularly great when the traits being selected have low heritability, the proportion of the genetic variance explained by the QTL is high, and where selection is able to occur before records for the phenotype become available or selection is done on one sex while phenotypes are recorded in the other sex. Efficiency of MAS is dependant on the precision of estimating the QTL position and allelic effects (Montaldo and Meza-Herrara, 1998). In particular, if the variance due to a QTL is overestimated, it will be given too much emphasis in estimation of breeding values and the accuracy of selection will be reduced.

The daughter design (Weller et al., 1990) has been used successfully for detecting QTL in dairy cattle populations. However this design requires very large numbers of half-sibs to create enough power to detect QTL, which results in costly genotyping. Selective genotyping is a cost saving strategy, which involves genotyping only the sire's daughters from the phenotypic extremes of a particular trait. It has been shown that the power to detect QTL affecting the trait of interest is increased (Lander and Botstein, 1989) and that it is only necessary to genotype the upper and lower 25% of the population to achieve maximum power from the design (Darvasi and Soller, 1992). However, including only genotyped animals in the analysis causes the QTL allelic effects for the selected trait to be overestimated (Georges et al., 1995; Xu and Vogl, 2000). Estimates of QTL effects for traits correlated to the selected trait are also biased (Bovenhuis and Spelman, 2000). This overestimation of allelic effects would result in decreased response to MAS. Methods have been developed which correct for this bias (Darvasi and Soller, 1992; Lander and Botstein, 1989; Muranty and Goffinet, 1997) but they assume strict truncation selection of animals to be genotyped. This paper presents a method to correct for bias caused by selective genotyping that does not assume truncation selection.

QTL detection in a half-sib design is commonly based on least-squares (LS) analysis and in fewer cases variance component (VC) analysis. In an LS approach the QTL is treated as a fixed effect and the QTL effect estimated within families. Alternatively the QTL is considered as a random effect in a VC approach and the proportion of the genetic variance attributed to the QTL estimated across the population (Visscher et al., 1998). Recently Lu Tong Duc (2003) has shown that the VC method can give unbiased estimates of the QTL variance despite selective genotyping, provided that the ungenotyped daughters of the sire are included in the analysis. Therefore the aim of this study was to compare an LS approach with a new method for correcting for bias caused by selective genotyping to a VC approach which uses ungenotyped animals in the analysis to estimate the QTL variance.

MATERIALS AND METHODS

Sample Collection and Genotyping. Semen samples for six Holstein-Friesian sires was supplied by Genetics Australia Co. Op. These were selected because they were influential to the Australian breeding program and had large numbers of lactating daughters. For each sire blood samples from approximately 100 high Australian Selection Index (ASI) half-sib daughters and 100 low ASI half-sib daughters were collected and DNA extracted, resulting in 1221 daughter samples in total. ASI is a profit index calculated as $ per cow per year = (3.8 x Protein ABV) + (0.9 x Fat ABV) - (0.048 x Milk ABV). All animals were genotyped for 8 microsatellite markers on one chromosome.

Phenotype Data. Australian Breeding Values (ABV's), were collected for both genotyped daughters and ungenotyped daughters from the Australian Dairy Herd Improvement Scheme (ADHIS) on 6 traits, protein yield (kg), fat yield (kg), milk yield (kg), protein percentage, fat percentage and ASI, and these were 'deregressed' to obtain a corrected phenotype ie a phenotypic record corrected for all fixed effects and for the ABVs of the sire and dam of each cow.

Statistical Analysis.

An underlying model was used in both LS and VC analysis $y_{ij} = m + s_i + q_i + e_{ij}$

where y_{ij} = corrected phenotypes, μ = mean, s_i = effect of the ith sire, q_i = the additive effect of the QTL, e_{ij} = error ~N(0, σ^2/R) where R is the reliability of the ABV contributed by the cows own records and σ^2 is the genetic variance.

Least squares: Only genotyped animals were included in the LS analysis, which was done using QTL Express (Seaton et al., 2002). The model used was $y_{ij} = m + s_i + x_{ij} q_i^* + e_{ij}$

where x_{ij} = the probability that offspring ij inherits allele 1 from their sire based on marker data, q^* = the gene substitution effect of QTL allele 1 from sire i, e = error which is assumed ~N(0, σ^2/R). The QTL variance is calculated as $s_q^{2*} = (MSreduced - MSfull)/(\overline{R}v(x))$

where MS is the residual mean squares of the reduced and full models, $\overline{R}$ is the average reliability and v(x) is the variance of x, ie., it depends on how informative the markers are. Selective genotyping causes these estimates of QTL effects (q*) to be biased. For the selected trait, ASI, unbiased estimates of the QTL effect (q) can be obtained by q = q*/w, where w = σ^{2*}/σ^2, σ^{2*} = within-sire ASI variance of genotyped daughters, σ^2 = within-sire ASI variance of all daughters. Similarly, the QTL variance

calculated from the genotyped cows only (σ_q^{2*}) is inflated by w^2 and so unbiased estimates are obtained by $\sigma_q^2 = \sigma_q^{2*} / w^2$

QTL effects for traits correlated with ASI are also biased, but uncorrelated traits should have no bias. Protein% and Fat% are almost uncorrelated with ASI and so estimates of QTL effects and variances do not need correction. ASI is a function of milk, fat and protein yields so, once effects of QTL on ASI, protein% and fat% are known it is possible to calculate QTL effects on milk, protein and fat yields by

Milk = 10.5ASI – 1970P% - 624F%
Protein = 0.288ASI – 3.3 P% - 17.6F%
Fat = 0.408ASI – 91.3 P% + 41 F%

This correction for selective genotyping is similar in principle to the method of Henshall and Goddard (1999).

Variance Component: This was conducted with and without ungenotyped animals included in the analysis. The statistical model is $y = m + s + q + e$

where y = corrected phenotypes, μ = mean, s = effect of the sire, q = additive genetic effect of the QTL, e = error. s, q and e are normally distributed as follows: $s \sim N(0, I\sigma_s^2)$ and $q \sim N(0, G\sigma_q^2)$, $e \sim N(0, \sigma^2/R)$, where G is the relationship matrix for the effects of the QTL and is based on marker information. $G_{ij}*\sigma_q^2$ is the covariance between the QTL effects in cows i and j. Consequently $G_{ii} = 2$ because each cow inherits 2 unrelated QTL allele's. For two typed daughters of the same bull G_{ij} = the probability that they inherit the same QTL allele from their sire. For an ungenotyped daughter $G_{ij} = 0.5$ among daughters of the same bull where one or both are ungenotyped. $G_{ij}=0$ if i and j are daughters of different bulls. G was calculated using Gibbs sampling. σ_s^2 and σ_q^2 were then estimated by restricted maximum likelihood using ASREML (Gilmour et al., 2001) at each marker location and at the midpoint between markers, treating the QTL as a random effect.

RESULTS AND DISCUSSION

Table 1 presents the results from the point on the chromosome with the highest likelihood of a QTL being present. Selective genotyping increased the variance of ASI in the typed cows to 1.67 times that in all daughters of the same sires. There was no increased variance in fat% and protein% as expected since these traits are uncorrelated with ASI. Consequently the unbiased estimates of QTL effects (q) are much less than the effects estimated directly from the typed cows (q*) and the variance due to the QTL ($2\sigma_q^{2*}$) is even more severely overestimated. For ASI, the variance estimated by VC (195.45) seems to have been underestimated, for reasons that seem to be particular to this data set.

This QTL explains approximately 15% (368.72/2067) of the genetic variance for ASI. This is not a particularly large proportion but, in our experience, it is typical of the larger QTL affecting milk, fat or protein yield. The gene substitution effect of about 20 ASI units is still of considerable economic value. However if the uncorrected estimate of variance (1027.80) was used in marker assisted selection, EBVs calculated with the benefit of the markers might well be less accurate than EBVs calculated without the inclusion of marker data.

Table 1. Estimates of QTL effects and variances for both LS and VC analysis.

Within sire s^2	**Protein %**	**Fat %**	**ASI**		**Protein kg**		**Fat kg**		**Milk yield**	
All cows	0.0202	0.0907	2067		210		514		440279	
Typed cows	0.0188	0.0910	3451		299		524		484302	
w	0.9307	1.0033	1.6696		1.4238		1.0195		1.1000	
QTL effects			**q***	**q**	**q***	**q**	**q***	**q**	**q***	**q**
1	0.0204	-0.1314	37.95	22.73	12.95	8.79	10.45	2.02	429.03	280.49
2	-0.1256	-0.1552	-29.04	-17.39	-4.89	-1.86	-7.88	-1.99	67.44	161.66
3	-0.0770	-0.0590	-46.31	-27.74	-12.09	-6.70	-16.68	-6.71	-302.06	-102.74
4	0.0187	-0.0929	-26.86	-16.09	-5.77	-3.06	-17.95	-12.08	-232.05	-147.78
5	0.0189	0.1116	-27.04	-16.20	-10.59	-6.69	-9.73	-3.76	-458.63	-276.94
6	0.0852	0.2307	43.90	26.29	8.37	3.23	21.72	12.41	152.74	-35.74
LS 2 s_q^2	0.0017	0.0086	1027.80	368.72	74.07	36.54	154.64	148.79	61184	50566
VC 2 s_q^2			764.16	195.45						

REFERENCES

Bovenhuis, H., and Spelman, R. J. (2000). *J. Dairy Sci.* **83,** 173-180.

Darvasi, A., and Soller, M. (1992). *Theor. Appl. Genet.* **85,** 353-359.

Lu Tong Duc (2003). *In* "Mapping quantitative trait loci using linkage and linkage disequilibrium with selective genotyping". M. Agr. Sci. thesis, University of Melbourne.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X. Y., Womack, J. E., and Hoeschele, I. (1995). *Genetics* **139,** 907-920.

Gilmour, Cullis, Welham, and Thompson. (2001). "ASREML Reference Manual."

Gimelfarb, A., and Lande, R. (1994). *Genetical Research* **64,** 127-136.

Henshall, J. M., and Goddard, M. E. (1999). *Genetics* **151,** 885-894.

Lande, R., and Thompson, R. (1990). *Genetics* **124,** 743-756.

Lander, E. S., and Botstein, D. (1989). *Genetics* **121,** 185-199.

Mackinnon, M. J., and Georges, M. A. J. (1998). *Livestock Production Science* **54,** 229-250.

Meuwissen, T. H. E., and Goddard, M. E. (1996). *Genet. Sel. Evol.* **28,** 161-176.

Montaldo, H. H., and Meza-Herrara, C. A. (1998). *Electronic Journal of Biotechnology* **1,** 83-89.

Muranty, H., and Goffinet, B. (1997). *Biometrics* **53,** 629-643.

Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M., and Visscher, P. M. (2002). *Bioinformatics* **18,** 339-340.

Visscher, P. M., Haley, C. S., Heath, S. C., Muir, W. J., and Blackwood, D. H. R. (1998). *Am. J. Med. Genet.* **81,** 465-465.

Weller, J. I., Kashi, Y., and Soller, M. (1990). *J. Dairy Sci.* **73,** 2525-2537.

Xu, S. Z., and Vogl, C. (2000). *Heredity* **84,** 525-537.