

MARKER ASSISTED SEGREGATION ANALYSIS IN COMPLEX PEDIGREES

J.M. Henshall, B. Tier and R.J. Kerr

Animal Genetics and Breeding Unit¹, University of New England, Armidale, NSW, 2351

SUMMARY

The use of marker assisted segregation analysis is demonstrated on a simulated livestock pedigree. Where information on a linked marker is included in the analysis, the power to detect major gene effects increases. It is shown that it is not necessary to genotype all animals. When only the sires in the pedigree are genotyped, the results are almost as good as if all of the animals are genotyped. Importantly, when data from an unlinked marker are included in the analysis no major gene effects are found.

Keywords: Quantitative trait, genetic markers, complex pedigrees, segregation analysis.

INTRODUCTION

Quantitative Trait Locus (QTL) detection experiments in livestock species commonly involve either genotyping a number of unrelated individuals or genotyping a number of closely related individuals. In each case the assumption or existence of linkage disequilibrium between the marked locus and the causal locus is required. Therefore, unrelated animals may be genotyped where it is expected that the genotyped locus is extremely close to the causal locus, while genotyping closely related animals allows more loosely linked loci to be detected. Typically structured designs are used to detect linked loci, requiring large numbers of fullsib or halfsib progeny. Often these must be bred specifically for the experiment, and phenotypes and genotypes recorded.

In the major livestock species, some traits of economic importance are commonly recorded for large numbers of stud or nucleus animals. Unless an animal is genotyped, or is in the same contemporary group as a genotyped animal, its records are not commonly used in QTL detection. An exception to this occurs when records are used to infer the presence of QTL in complex pedigrees without using markers, as in segregation analysis. Methods that use all available marker and phenotype data for a complex pedigree are not commonly in use, but have the obvious advantage of maximising the use of existing data.

The ability to use linked marker information has recently been incorporated into the "Gene Detective", a segregation analysis program (Tier and Henshall 2001). In this paper the use of markers in segregation analysis for the purpose of QTL detection (as opposed to QTL exploitation) is demonstrated, and the effect of using some different genotyping strategies compared.

METHODS

The Gene Detective uses a Markov Chain Monte Carlo algorithm to perform segregation analysis on complex pedigrees. Models can include multiple traits (continuous or categorical) with direct and

¹ AGBU is a joint institute of NSW Agriculture and The University of New England

maternal polygenic effects, with or without a QTL effect. QTL genotype samples are obtained by applying a Metropolis-Hastings acceptance-rejection step (Metropolis *et al.* 1953; Hastings 1970) to sampled descent graphs. The likelihood used in the Metropolis-Hastings step is derived from the phenotypes (adjusted for current estimates of fixed and polygenic effects), the current QTL genotype sample and the current QTL effect vector sample. If identity-by-descent (IBD) data at a marker locus are available, then the method is easily extended to incorporate these data, by including the IBD probabilities, adjusted for a recombination rate, in the likelihood used in the Metropolis-Hastings step.

To demonstrate the method, analyses of a simulated population were performed. Data were simulated to roughly resemble a cattle pedigree structure. The base population comprised 25 sires and 1000 dams. Then followed eight breeding cycles, in each of which 1000 calves were born to 1000 cows. For each cycle after the first, approximately 50% of sires and 50% of dams were replaced with animals chosen at random from the latest crop of calves.

A trait with error variance $V_e = 4.0$, additive variance $V_a = 1.0$ and QTL variance $V_q = 1.0$ was simulated, with records available on all animals apart from the base. The QTL was simulated as having two alleles, with no dominance and allele frequencies of 0.5 in the base population. This requires that the difference between homozygous animals be 2.83. A marker, immediately adjacent to the QTL, but in linkage equilibrium with the QTL was also simulated. While having a marker this close to the QTL may be an ideal situation, having a marker haplotype including the QTL is quite feasible, and provides a similar level of information to a marker immediately adjacent to the QTL. Another marker, unlinked to the QTL was also simulated. Each marker locus had eight alleles, each at an allele frequency of 0.125 in the base population. IBD probabilities for the marker loci were estimated using the GEIC algorithm (Henshall *et al.* 1999, 2001), 100 independent descent graphs for each marker locus were obtained, with IBD probability estimated as the unweighted mean incidence.

Analyses were carried out without marker information, and using either the linked or the unlinked marker locus, with all animals, all sires, or just the last crop of calves treated as genotyped. Samples were obtained using the Gene Detective until the Markov chain was judged to have converged. This generally required fewer samples when more genotypic information was present.

RESULTS AND DISCUSSION

The results are presented in Table 1. Variance components were estimated as the mean of the last $n/4$ samples where n is the total number of samples for that analysis. The standard deviations of the samples for the variance components were less than 0.2 for all but V_a and V_q when the analysis did not use marker information, and V_a when the analysis included marker information from sires for the unlinked marker.

For all genotyping strategies the error variance is well estimated. Without marker information the QTL is too small to detect in this pedigree, and most of the QTL variance is attributed to the additive variance. This also occurs when information from the unlinked marker is included in the analysis. In fact, the amount of variance attributed to the QTL is less when information from the unlinked marker is used than when no marker information is used. This is regardless of the proportion of animals

genotyped. This suggests that it may be possible to exclude regions of the chromosome from further consideration on the basis of a relatively small number of genotypings.

Table 1. Estimated variance components and QTL effects under various genotyping regimes, for both a linked and an unlinked marker. V_e is the error variance, V_a the additive (polygenic) variance, V_q the variance due to the QTL, a is the magnitude of the QTL effect measured as the difference between homozygous classes, with standard deviation of the QTL effect samples (sd), and with n_{gen} the number of animals genotyped

		V_e	V_a	V_q	a		
Simulated		4.00	1.00	1.00	2.83		
Genotyped	Marker					(sd)	n_{gen}
0%		3.88	1.69	0.19	0.93	(0.76)	0
100%	linked	4.17	0.72	0.98	2.78	(0.11)	9025
100%	unlinked	4.10	1.63	0.04	0.51	(0.21)	9025
last crop ^A	linked	3.89	1.44	0.58	2.14	(0.18)	1000
last crop ^A	unlinked	4.11	1.66	0.02	0.28	(0.23)	1000
sires only ^B	linked	3.96	0.90	0.97	2.76	(0.22)	127
sires only ^B	unlinked	3.84	1.81	0.11	0.68	(0.58)	127

^A Only progeny born in the last year genotyped

^B Only animals used as sires genotyped

When information from the linked marker is included in the analysis, the power to detect the QTL increases, and the effect and variance due to the QTL are generally well estimated. An exception to this occurs when only the last crop of calves is genotyped. In this case, although the power to detect the QTL is increased, both QTL effect and the variance due to the QTL are underestimated. This may be due to fact that the genotypic data is somewhat removed from the bulk of the phenotypic data under this genotyping strategy, with the information from the ungenotyped individuals partially swamping that from the latest group of calves. It is also difficult to infer the genotypes of many ancestors from the genotypes of this set. For this type of data set it may be preferable to perform standard halfsib interval mapping, or perhaps to discard some of the earliest phenotypic data.

When only sires are genotyped, the additive variance and QTL effect and variance are well estimated despite sires being only 1.4% of the pedigree. Under this genotyping strategy all animals with phenotypes have a parent with a genotype record, and most also have two grandparents with genotype records. This appears to be a very powerful design for detecting QTL with relatively few genotypings required. An additional advantage is that the risk involved in selecting heterozygous sires to parent halfsib families for interval mapping experiments is avoided. Unfortunately, in the pedigrees available in most livestock species today, most of the ancestral sires will not be available for genotyping. However, where genotypes can be obtained for ancestral sires, they may have a value of many times that of a non-breeding animal currently in the herd.

The genotyping strategies described here are examples of what may be possible. For a particular data set, the availability of DNA from ancestral animals and the size of existing full and half-sib families may determine the optimum genotyping strategy. This may or may not involve the generation of additional progeny, and this decision might be made after initial screening of the existing population to detect heterozygous sires.

CONCLUSION

In this paper the use of marker assisted segregation analysis has been demonstrated. It is clear that good results can be obtained when only a small proportion of recorded animals are genotyped, especially if genotypes for ancestral sires can be obtained. This has the potential to significantly reduce the cost of performing QTL detection experiments if traits are already recorded on complex pedigrees, and if genotypes can be obtained for sires which are no longer in use, for example from stored semen. For the methods described here, all genotypic and all phenotypic information can be used in the analysis. The results presented here strongly suggest that breeding organisations should be preserving the DNA of widely used parents for future use in QTL detection.

REFERENCES

- Hastings, W.K. (1970) *Biometrika* **57**: 97.
Henshall, J.M., Tier, B. and Kerr, R.J. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**: 329.
Henshall, J.M., Tier, B. and Kerr, R.J. (2001) *Genetical Research* (Accepted).
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) *J. Chem. Phys.* **21**: 1087.
Tier, B. and Henshall, J.M. (2001) *Genet. Sel. Evol.* (Accepted).