

## EFFICIENT DESIGNS FOR FINE-MAPPING OF QUANTITATIVE TRAIT LOCI USING LINKAGE DISEQUILIBRIUM AND LINKAGE

**S.H. Lee and J.H.J. van der Werf**

Department of Animal Science, University of New England, Armidale, NSW

### SUMMARY

This study showed that common half sib designs can contain sufficient linkage disequilibrium (LD) information for fine-mapping of quantitative trait loci (QTL). A design of two large half sib families was applied to a hypothetical population with an effective size ( $N_e$ ) of 1000 simulated for 6000 generations. 70 ~ 75 % of replicates could position QTL within 0.75 cM when mutation age was more than 200 generations in mapping using multi-allelic markers. When  $N_e$  was linearly decreasing (1000 to 100) in the last 50 generations, the accuracy of QTL mapping was decreased to only 50 to 70 % of replicates that could position QTL within 0.75 cM. Provided that QTL allele segregates in the dam population, it is advantageous to use the same half sib design in linkage mapping and fine-mapping making gene mapping cost effective in livestock populations.

**Key-words:** quantitative trait loci, fine-mapping, designs.

### INTRODUCTION

In many mapping studies, it would be desirable to use fine-mapping to reduce the size of the confidence interval at the genomic region containing quantitative trait loci (QTL) to a few cM. Recently, a variance component (VC) method using combined linkage disequilibrium (LD) and linkage information has been considered as a promising approach for fine-mapping. The method has proven to result in a mapping resolution accurate enough to narrow down the QTL confidence interval to a few cM (Meuwissen and Goddard 2001).

In mapping studies, efficient design and family structure is required to make gene mapping cost effective. However, the efficiency of different designs for fine-mapping using combined LD and linkage information has hardly been reported. Half sib designs are often used for coarse QTL mapping in outbred populations. The question is open whether such designs contain sufficient information for fine-mapping. Besides the design of the experiment, other properties of the population used in the study may be important for fine-mapping. For example, the effective size ( $N_e$ ) has an important effect on the degree of LD. Hayes *et al.* (2002) have also shown that LD patterns are affected by whether the population size has effectively increased (as in humans) or effectively decreased (as in most livestock) in recent times. Also, the age of the favourable QTL mutation may be relevant for the efficiency of LD mapping as it will affect the LD pattern of marker haplotypes surrounding the QTL.

The aim of this study is to investigate the efficiency of various experimental designs for fine-mapping of QTL. Several hypothetical situations with varying  $N_e$  and various mutation ages (MA) are used to test the

usefulness of existing designs for linkage mapping and proposed designs in livestock for fine scale mapping.

## **MATERIALS AND METHODS**

**Simulation study.** The first part of the simulation was designed to generate a variety of populations differing in effective population size and the length of the population history ( $t$ ). In each generation, the number of male and female parents was equal, and their alleles were transmitted to descendants based on Mendelian segregation using the genedropping method (MacCluer *et al.* 1986). Unique numbers were assigned as alleles chosen to be mutant in a given generation (depending on mutation age). In the last generation, one of the surviving mutant alleles was chosen and treated as the favourable QTL allele. The marker alleles were mutated at a rate of  $4 \times 10^{-4}$  per generation. In the bi-allelic marker model, the existing allele was substituted by a mutated locus whereas in the multi-allelic markers model, a new allele was added.

The second part of the simulation model was designed to create a variety of family structures with recorded data sets through varying the number of sires, dams and offspring. The sires and dams were randomly selected in the last generation ( $t$ ) of the first part of the simulation. Descendants in generation  $t+1$  were given a phenotypic record and pedigree was only known for these animals (i.e. animals from generations  $t$  were considered unrelated base animals).

Marker genotypes for all animals were available for animals from generation  $t$  and  $t+1$ . For a fair comparison between experimental designs, phenotypic value was only available for a fixed number of progeny in generation  $t+1$ . Phenotypic values were simulated as  $y = \mu + q + u + e$ . The mean of population ( $\mu$ ) was 100. QTL effect ( $q$ ), polygenic effect ( $u$ ) and residuals ( $e$ ) were normally distributed with variances 25, 25 and 50, respectively.

*Effect of family structure on efficiency of fine-mapping.* Various experimental designs for fine-mapping of QTL were investigated. Mutation occurred at generation 0. An effective population size of 100 was assumed for 100 generations. At generation 101, full sib and half sib families were generated. The number of families were 64, 32, 16, 8 or 2 with a total number of progeny of 128 for each design (i.e. 2, 4, 8, 16 and 64 progeny per family). Ten markers were positioned at 1 cM intervals. The proportion of replicates (out of 50) positioning the QTL within 3cM of the true location was determined for each design.

*Properties of the population used for LD mapping.* In a second part of the study, properties of the population used for fine-mapping were evaluated to determine their efficiency for LD mapping. Populations were simulated of varying effective size and age of the mutation. Initially, a population with an effective size of 1000 was simulated for 6000 generations with various mutation ages (i.e.  $t = 6000$ ). The mutation age describes the period since a mutation occurred. Mutations occurred at the 2000<sup>th</sup>, 4000<sup>th</sup>, 5000<sup>th</sup>, 5500<sup>th</sup>, 5800<sup>th</sup> or the 5900<sup>th</sup> generation, respectively. At generation  $t+1$ , two half sib families of size 64 were generated. Ten multi-allelic markers were positioned at every 0.25 cM (or 1 cM) intervals.

For each mutation age, the proportion of 50 replicated simulations positioning the QTL within 0.75 cM (or 3 cM) of the true location was estimated. Also, a population with linearly decreasing  $N_e$  with the various mutation ages was tested:  $N_e$  decreased linearly from 1000 to  $N_e=100$  over the last 50 generations.

**Analysis of simulation data sets.** The phenotypic observations are modelled as  $y = Xb + Z_1u + Z_2q + e$ , where  $y$  is a vector of observations,  $b$ ,  $u$ ,  $q$  and  $e$  are vectors of fixed effects, random polygenic effects for each animal, random effects due to QTL and residuals. The random effects ( $u$ ,  $q$  and  $e$ ) are assumed to be normally distributed with mean zero and variance  $S_u^2$ ,  $S_q^2$  and  $S_e^2$ .  $X$ ,  $Z_1$  and  $Z_2$  are design matrices for the effects in  $b$ ,  $u$  and  $q$ , respectively. The variance covariance matrix of all observations is defined as

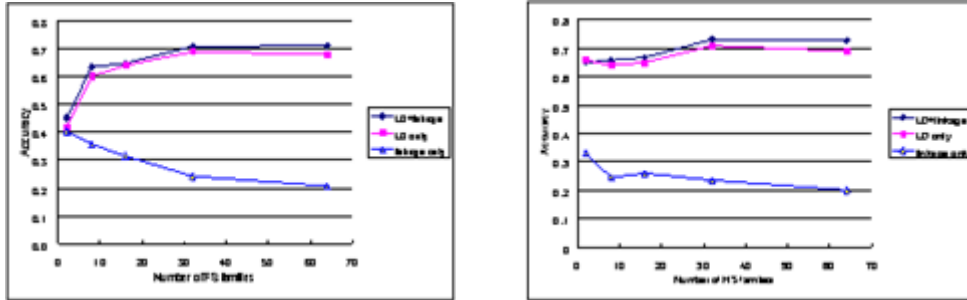
$$V = Z_1AZ_1'S_u^2 + Z_2GZ_2'S_q^2 + IS_e^2$$

where  $A$  is the numerator relationship matrix,  $G$  is the gametic relationship matrix (GRM) whose elements are identity by descent (IBD) probabilities between individuals at a putative QTL, and  $I$  is an identity matrix. Using the genedropping method and a recursive algorithm (Wang *et al.* 1995), the GRM was constructed based on combined LD and linkage (Meuwissen and Goddard 2001). QTL position and variance components were estimated by average information restricted maximum likelihood (AIREML, Johnson and Thompson 1995).

## RESULTS AND DISCUSSION

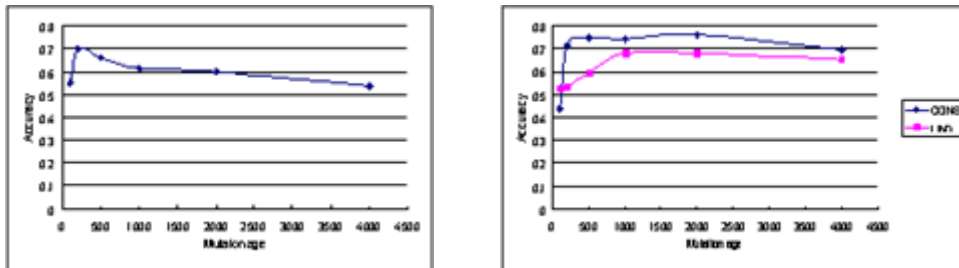
**Efficient designs for fine-mapping of QTL.** For combined LD and linkage mapping, many families of small size provided higher mapping resolution than few families of large size. The proportion of replicates positioning the QTL within 3 cM was 0.71 (full sib families) and 0.73 (half sib families) in 64 families each with 2 individuals compared to 0.43 (full sib) and 0.65 (half sib) in 2 families each with 64 individuals (Figure 1). The overall mapping accuracy decreases slightly with LD information only, and it decreases considerably with linkage information only.

It is apparent that, with a half sib design, few families of large size can give a reasonable mapping accuracy. This is likely due to the fact that in half sib designs, there is substantial LD information in the dam population which can be used. Note that the number of base dams is constant in the different half sib structures. This result implies that common half sib designs such as few families with large size which are often used in linkage mapping can also have sufficient LD information for fine-mapping.



**Figure 1. Accuracy of QTL mapping (as proportion of replicates with position estimate within 3 cM of true value) depending on number of families (fixed number of individuals in each design is 128) using LD, linkage or a combined method for estimation.**

**Effective population size and mutation age.** Figure 2 (left) shows the proportion of replicates with the QTL positioned within 3 cM of the true location is 55 ~ 70 % with constant  $N_e=1000$  when multi-allelic markers are positioned every 1 cM. The accuracy is highest for 200 generations of mutation age (69.9 %) and then gradually decreases (66 % for  $MA=500$  ~ 53.8 % for  $MA=4000$ ). When there is a long time period since the mutation ( $> 200$  generations), the IBD region around the QTL is too small for the marker spacing of 1 cM. When marker spacing is 0.25 cM, overall accuracy is improved (70 ~ 75 %, Figure 2 (right)). After 200 generations since the mutation, the accuracy is stable (74.2 % for  $MA=1000$ ; 75.9 % for  $MA=2000$ ; 69.4 % for  $MA=4000$ ). When  $N_e$  was linearly decreased (LIND), overall accuracy was lower than that with constant  $N_e$  (Figure 2). This is probably due to fact that in the case of old mutation ( $MA>200$ ), some of the LD information from the original population was lost due to the decrease of effective population size.



**Figure 2. Proportion of replicates with QTL positioned within 3 cM (left) and 0.75 cM (right) of true position when mutation age is varied and ten multi-allelic markers are positioned every 1 cM (left) and 0.25 cM (right). CONS – constant  $N_e=1000$ , LIND – linearly decreasing 1000 to 100.**

### CONCLUSION

In the present study, we showed that a half sib design could be efficiently used for fine-mapping of QTL provided that the QTL segregates in the dam population. After the population has a certain degree of LD between the trait mutation and flanking markers ( $> 200$  generations since the mutation), a QTL can be positioned within 0.75 cM of the true location in 70 ~ 75 % of replicates with constant  $N_e=1000$ , 50 ~ 70 % with decreasing  $N_e$ . When the number of animals used for fine-mapping increases, the accuracy can be increased (result not shown). It is suggested that the same half sib design with the same animals used in linkage mapping should be considered for fine-mapping of QTL.

### REFERENCES

- Hayes, B.J., Visscher, P.M., McPartlan, H. and Goddard, M.E. (2002) *Genome Research* (accepted).
- Johnson, D.L. and Thompson, R. (1995) *J. Dairy. Sci.* **78**: 449.
- MacCluer, J.W., VanderBerg, J.L., Read, B. and Ryder O.A. (1986) *Zoo Biology* **5**: 147.
- Meuwissen, T.H.E. and Goddard, M.E. (2001) *Gen. Sel. Evol.* **33**: 605.
- Wang, T., Fernando, R.L., van der Beek, S., Grossman, M. and van Arendonk, J.A.M. (1995) *Gen. Sel. Evol.* **27**: 251.