Evaluation of a Fine-mapping Method Exploiting Linkage Disequilibrium in Livestock Populations: Simulation Study

JongJoo Kim* and **Frederic Farnir**[1]

School of Biotechnology, Yeungnam University, Gyeongsan, 712749, Korea

ABSTRACT : A simulation study was conducted to evaluate a fine-mapping method exploiting population-wide linkage disequilibrium. Data were simulated according to the pedigree structure based on a large paternal half-sib family population with a total of 1,034 or 2,068 progeny. Twenty autosomes of 100 cM were generated with 5 cM or 1 cM marker intervals for all founder individuals in the pedigree, and marker alleles and a number of quantitative trait loci (QTL) explaining a total of 70% phenotypic variance were generated and randomly assigned across the whole chromosomes, assuming linkage equilibrium between the markers. The founder chromosomes were then descended through the pedigree to the current offspring generation, including recombinants that were generated by recombination between adjacent markers. Power to detect QTL was high for the QTL with at least moderate size, which was more pronounced with larger sample size and denser marker map. However, sample size contributed much more significantly to power to detect QTL than map density to the precise estimate of QTL position. No QTL was detected on the test chromosomes in which QTL was not assigned, which did not allow detection of false positive QTL. For the multiple QTL that were closely located, the estimates of the QTL positions were biased, except when the QTL were located on the right marker positions. Our fine mapping simulation results indicate that construction of dense maps and large sample size is needed to increase power to detect QTL and mapping precision for QTL position. (**Key Words :** QTL, Fine Mapping, Mapping Precision, Detection Power)

INTRODUCTION

Detection of quantitative trait loci (QTL) exploiting linkage disequilibrium (LD) within experimental or commercial populations has been routinely practiced in farm animals (Kim et al., 2004; Kim et al., 2005), which provides a genetic framework for further researches, i.e. fine mapping, marker-assisted selection or introgression, and positional cloning of causal genes (http://www.animalgenome.org/). However, the number of identified genes responsible for the detected QTL is very limited (Dekkers, 2004), partly due to insufficient sample size and low mapping resolution. To improve power to detect QTL and mapping precision to localize genes in the QTL region, several fine-mapping methodologies were developed and applied to livestock mapping populations (Riquet et al., 1999; Farnir et al., 2002; Gautier et al., 2006).

Livestock populations, contrary to humans, have young and dynamic structures with a small number of founders, which enables efficient implementation of fine-mapping procedures without the need for very high density maps, because LD spreads over long distance, e.g. more than ten cM (Farnir et al., 2000; McRae et al., 2002; Harmegnies et al., 2006).

We developed a fine-mapping method by exploiting population-wide LD, i.e. "historic" recombinants and reported successful implementation to fine-map milk QTL locations in dairy cattle (Blott et al., 2003; Grisart et al., 2004; Schnabel et al., 2005). However, efficiency of mapping precision and accuracy using the method was not investigated under different mapping conditions. We herein report simulation results to evaluate power to detect QTL and mapping precision in a large half-sib population by applying our developed fine mapping method.

MATERIALS AND METHODS

Mapping populations

Data were simulated according to the pedigree structure based on the 22 paternal half-sib families for a total of

Table 1. Simulated positions and effects (proportion of phenotypic variance, % σ_p^2) for the QTL that were randomly assigned in chromosomes (BTAs), and Lods, most likely positions, and effects for the detected QTL under different mapping conditions

| QTL | BTA | Simulation | | Lod[ab] | | | | cM[ac] | | | | % σ_p^{2ad} | | | |
| | | | | N = 1,034 | | N = 2,068 | | N = 1,034 | | N = 2,068 | | N = 1,034 | | N = 2,068 | |
		cM	% σ_p^2	5 cM	1 cM	5 cM	1 cM	5 cM	1 cM	5 cM	1 cM	5 cM	1 cM	5 cM	1 cM
1	2	16.8	6.1	4.4	5.9	6.8	9.7	15.0	19.5	17.5	17.5	6.7	7.7	3.5	5.2
2	3	19.0	2.8	2.2	4.8	-	8.3	25.0	18.5	-	19.5	-	3.3	-	2.9
3	4	30.0	0.9	-	-	-	-	-	-	-	-	-	-	-	-
4	4	34.5	2.2	-	1.4	3.5	2.7	-	34.5	37.5	33.5	-	-	4.6	-
5	4	42.9	0.5	-	-	-	-	-	-	-	-	-	-	-	-
6	4	54.3	0.9	-	-	3.1	2.6	-	-	52.5	56.5	-	-	-	-
7	5	13.3	4.9	2.2	4.1	8.1	4.3	17.5	23.5	27.5	12.5	-	4.3	7.4	3.2
8	5	35.9	7.2	2.7	3.2	7.7	10.6	37.5	38.5	22.5	36.5	-	5.1	7.8	4.5
9	10	39.2	5.9	2.4	2.8	4.6	8.2	52.5	43.5	42.5	41.5	-	-	3.8	6.4
10	10	68.0	3.0	2.0	3.4	6.6	11.6	62.5	67.5	62.5	67.5	-	-	3.4	14.2
11	12	52.4	11.7	3.7	4.9	12.5	15.9	47.5	51.5	47.5	51.5	6.0	5.6	6.9	6.3
12	14	64.2	3.8	-	-	5.5	7.3	-	-	62.5	70.5	-	-	5.9	5.5
13	15	12.3	0.5	5.4	2.4	4.7	-	22.5	17.5	17.5	-	12.8	-	2.5	-
14	15	67.8	11.7	6.1	9.5	8.2	13.2	67.5	67.5	75.0	67.5	8.7	10.5	8.0	6.6
15	16	59.1	3.8	-	-	-	2.6	-	-	-	54.5	-	-	-	-
16	16	64.9	1.9	-	-	2.9	2.2	-	-	72.5	62.5	-	-	-	-
17	17	43.6	2.3	2.4	3.9	4.1	5.6	32.5	39.5	32.5	39.5	-	6.4	4.0	4.0
18	17	77.4	0.5	-	-	-	4.0	-	-	-	63.5	-	-	-	1.5
19	18	67.7	15.5	7.2	9.0	18.2	24.4	72.5	65.5	72.5	68.5	8.3	6.0	7.4	9.0
20	20	70.5	0.3	-	-	-	-	-	-	-	-	-	-	-	-

[a] N = 1,034, the total number of progeny in the 22 paternal half-sib families. 5 cM, interval length between adjacent markers.

[b] Lod value of 3.5 is the threshold for QTL detection at 5% genome-wise level.

[c] QTL position at which the test statistic was maximized.

[d] Proportion of phenotypic variance due to the QTL, $[100 \times (2\sigma_H^2/[2\sigma_H^2 + \sigma_A^2 + \sigma_E^2])$, where σ_H^2 is the estimated variance for the QTL haplotype effect, σ_A^2 for the polygenic background, and σ_E^2 for residual environmental effect.

1,034 progeny in a Dutch black-and-white dairy cattle population (Farnir et al., 2000). Twenty autosomes were generated for all founder individuals in the pedigree, and four alleles for each marker were generated with different allele frequencies, such that the average heterozygosity was 0.7 that was observed in the Dutch cattle population. A number of QTL explaining a total of 70% phenotypic variance were generated and randomly assigned across the whole chromosomes, assuming linkage equilibrium between the markers. Relative positions and effects for the simulated QTL are displayed in Table 1. Founder chromosomes were then allowed to segregate within the pedigree, including recombination at a rate determined by the genetic distance between adjacent markers, and descended 14 generations down to the individuals of the last generations in the pedigree. Different marker map densities were assigned; 5 cM or 1 cM marker intervals, and different numbers of progeny (N = 1,034, or N = 2,068) were generated using the same pedigree structure, i.e. the same number of sires. Phenotype for each progeny was determined according to the sum of effects of alleles inherited from the ancestors in the pedigree and a random residual value was obtained from a normal distribution for the environmental effect.

QTL analysis

The mapping models utilized in the simulation study are described in detail in Kim and Georges (2002) and Blott et al. (2003). Briefly, the following procedures were applied for QTL detection: 1) marker linkage phase of the sires and their progeny were determined, 2) identity-by-descent (IBD) probabilities for all pair-wise combinations of sire (SC) and maternal chromosomes (MC) of progeny were computed using a coalescent model, 3) a dendrogram representing genetic relationships between all haplotypes was generated by applying a hierarchical clustering algorithm, 4) the haplotype cluster that gave the maximum solution in the fitted REML model was determined.

The mixed linear model that includes LD information among haplotypes is:

$$y = Xb + Z_h h + Z_u u + e$$

where y is the vector of phenotype records of all offspring, b is a vector of fixed effects which in this study reduces to the overall mean, X is an incidence matrix relating fixed effects to individual offspring, which in this study reduces to a vector of ones. h is the vector of random QTL effects corresponding to the defined haplotype clusters.
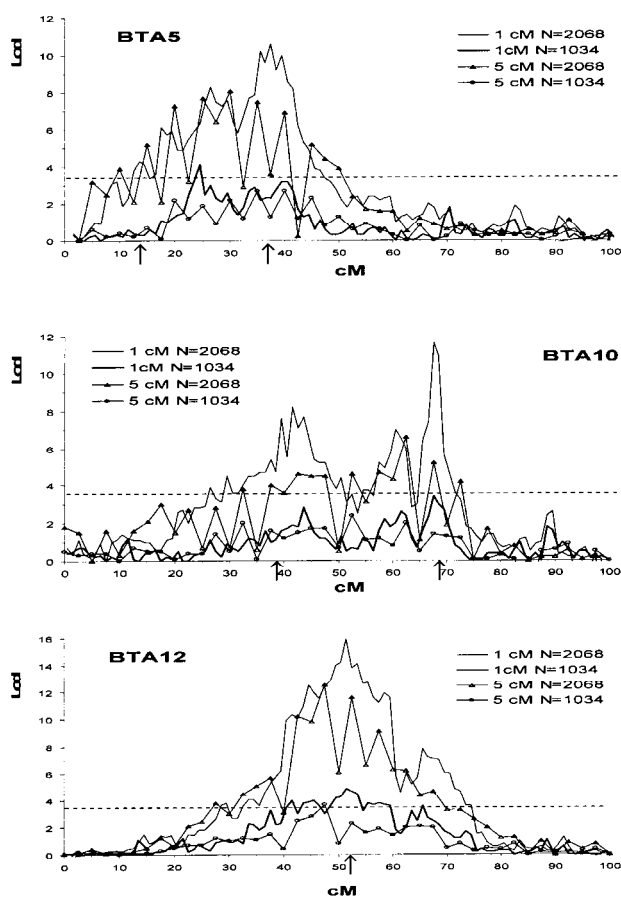
Figure 1. Lod profiles for the simulated QTL on bovine chromosomes (BTAs) 5, 10, and 12. Arrows indicate simulated QTL positions. Horizontal dotted lines are 5% genome-wide threshold values. 1 cM and 5 cM indicate interval distance between adjacent markers. N = 1,034 and N = 2,068 indicate the total number of progeny across the paternal halfsib families.

Z_h is an incidence matrix relating haplotype clusters to individual offspring. In Z_h, a maximum of three elements per line can have non-zero value: "1" in the column corresponding to the cluster to which the MC haplotype belongs, "λ_p" and "ρ_p" in the columns corresponding respectively to the haplotype clusters of the "right" and "left" SC. If either of the SC and/or MC belongs to the same cluster, the corresponding coefficients are added (Kim and Georges, 2002). Z_u is a diagonal incidence matrix relating individual polygenic effects to individual offspring, and **u** is the vector of the random additive polygenic effects. **e** is the vector of random sapling errors. The null model with no QTL is the reduced model without QTL haplotype effect, such that:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_u u} + \mathbf{e}$$

Test statistics, i.e. likelihood ratio test (LRT), were obtained by comparing the likelihoods between the QTL

model versus no QTL model, at each map position, i.e. the mid-point in marker interval, to generate Lod (0.22*LRT) score profiles. Previous simulation results showed that the LRT considering a whole genome scan with no QTL followed chi-squared distributions with 2 d.f. with Bonferroni correction for 174 independent traits, which corresponded to 3.5 Lod at 5% genome-wise (GW) significance level (Kim and Georges, 2002).

RESULTS AND DISCUSSION

Table 1 presents simulated positions and effects of the 20 QTL that were randomly assigned on chromosomes and test statistics (Lods), and estimated positions and effects for the QTL that were detected at the 5% GW significance level. Generally, power to detect QTL was high, i.e. high Lod scores for the QTL with at least moderate size, e.g. between 10-20% of phenotypic variance (σ_p^2). However, when QTL effect was small, e.g. less than 5% σ_p^2, power to detect QTL decreased, which was more pronounced when sample size was small and marker map density was low (Table 1). No QTL was detected on the test chromosomes in which QTL was not assigned, which did not allow detection of false positive QTL (results not shown).

There was a general tendency for power of detecting QTL to increase with marker map density and sample size. However, the latter contributed much more significantly to power of detecting QTL, e.g. for the QTL on BTA12, the Lod value was 3.7 when sample size was small (N = 1,034) and marker interval was 5 cM. The QTL was detected with 12.5 Lod, when the number of progeny was doubled given the same map density. However, the QTL was detected with 4.9 Lod, with a more dense map (1 cM marker interval) given the same sample size (Table 1; Figure 1). Increasing marker map density, however, enabled more precise location of QTL. For the QTL that was located at 52.4 cM in BTA12, the most likely position was estimated at 51.5 cM under a 1 cM marker interval map, while it was estimated at 47.5 cM under the 5 cM marker interval map (Table 1; Figure 1). These results are consistent with the reports of Lee and van der Werf (2005), or Meuwissen and Goddard (2000), in which increasing map density from 1 cM up to 0.25 cM marker interval yielded more precise QTL location.

Sometimes locations for multiple QTL were poorly estimated, when the QTL were closely located on the same chromosome. For the two QTL on BTA5 that were located at 13.3 cM and 35.9 cM, respectively, multiple distinct QTL peaks (positions) were observed in the region between 15 cM and 45 cM for the simulated data with N = 2068 and 1 cM marker interval, and the most likely position in this region was at 27.5 cM, the mid point of the two QTL positions, when the simulated sample with N = 2,068 and 5

cM marker interval was used (Table 1; Figure 1). However, when QTL was located on the right marker position, e.g. the QTL at 68.0 cM in BTA10 that was detected using the sample with N = 2,068 under the 1 cM marker interval map, the power of detecting QTL increased dramatically compared to the sample with N = 2,068 under the 5 cM marker interval map where the QTL marker was not included in the map, and the position estimate was unbiased, even when another QTL was closely located at 39.2 cM in the same chromosome (Figure 1).

Generally, there was limited accuracy in estimating QTL effect, i.e. proportion of phenotypic variance for the detected QTL (Table 1). This tendency may be partly due to imperfect clustering of haplotypes according to the alternate QTL alleles responsible for the phenotype variation, as well as to insufficient map density between flanking markers and QTL (Kim and Georges, 2002).

In conclusion, the results of this simulation study of exploiting LD information in livestock populations suggest that, in general, to increase power to detect QTL and mapping precision, it is necessary to construct a more dense map as well as utilizing a larger sample size.

REFERENCES

Blott, S., J.-J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir and W. Coppieters. 2003. Molecular dissection of a QTL: a phenylalanine to tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genet. 163:253-266.

Dekkers, J. C. M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. J. Anim. Sci. 82: E313-328E.

Lee, S. H. and J. H. J. van der Werf. 2005. The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree. Genet. 169:455-466.

Farnir, F., W. Coppieters, J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar and M. Georges. 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. 10:220-227.

Farnir, F., B. Grisart, W. Coppieters, J. Riquet, P. Berzi, N. Cambisano, L. Karim, M. Mni, S. Moisio, P. Simon, D. Wagenaar, J. Vilkki and M. Georges. 2002. Simultaneous mining of linkage and linkage disequilibrium to fine-map QTL in outbred half-sib pedigrees: revisiting the location of a QTL with major effect on milk production on bovine chromosome 14. Genet. 161:275-287.

Gautier, M., R. R. Barcelona, S. Fritz, C. Grohs, T. Druet, D. Boichard, A. Eggen and T. H. E. Meuwissen. 2006. Fine mapping and physical characterization of two linked quantitative trait loci affecting milk fat yield in dairy cattle on BTA26. Genet. 172:425-436.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J.-M. Frere, W. Coppieters and M. Georges. 2004. Genetic and functional confirmation of the causality of the *DGAT1 K232A* quantitative trait nucleotide in affecting milk yield and composition. P.N.A.S. 101:2398-2403.

Harmegnies, N., F. Farnir, F. Davin, I. Geerts, N. Buys, M. Georges and W. Coppieters. 2006. Measuring the extent of linkage disequilibrium in commercial pig populations. Anim. Genet. 37:225-231.

Kim, J. J. and M. Georges. 2002. Evaluation of a new fine-mapping method exploiting linkage disequilibrium: a case study analysing a QTL with major effect on milk composition on bovine chromosome 14. Asian-Aust. J. Anim. Sci. 15:1250-1256.

Kim, T.-H., B.-H. Choi, D.-H. Yoon, E.-W. Park, J.-T. Jeon, J.-Y. Han, S.-J. Oh and I.-C. Cheong. 2004. Identification of quantitative trait loci (QTL) affecting teat number in pigs. Asian-Aust. J. Anim. Sci. 17:1210-1213.

Kim, T.-H., B.-H. Choi, H.-K. Lee, H. S. Park, H. Y. Lee, D. H. Yoon, J. W. Lee, J.-T. Jeon, I.-C. Cheong, S.-J. Oh and J.-Y. Han. 2005. Identification of quantitative trait loci (QTL) affecting growth traits in pigs. Asian-Aust. J. Anim. Sci. 18:1524-1528.

McRae, A. F., J. C. McEwan, K. G. Dodds, T. Wilson, A. M. Crawford and J. Slate. 2002. Linkage diesequilibrium in domestic sheep. Genet. 160:1113-1122.

Meuwissen, T. H. E. and M. E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genet. 155:421-430.

Meuwissen, T. H. E., A. Karlsen, S. Lien, I. Olsaker and M. Goddard. 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genet. 161:373-379.

Riquet, J., W. Coppieters, N. Cambisano, J. Arranz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, M. Mni, P. Simon, J. F. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack and M. Georges. 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc. Natl. Acad. Sci. USA 96:9252-9257.

Schnabel, R. D., J.-J. Kim, M. S. Ashwell, T. S. Sonstegard, C. P. van Tassell, E. E. Conner and J. F. Taylor. 2005. Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene. P.N.A.S. 102:6896-6901.