Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Fine mapping of multiple interacting quantitative trait loci using combined linkage disequilibrium and linkage information*

LEE Sang Hong†, VAN DER WERF J.H. Julius

(*School of Rural Science and Agriculture, University of New England, Armidale, New South Wales 2351, Australia*)

†E-mail: slee38@une.edu.au

Received Sept. 13, 2007; revision accepted Sept. 16, 2007

Abstract: Quantitative trait loci (QTL) and their additive, dominance and epistatic effects play a critical role in complex trait variation. It is often infeasible to detect multiple interacting QTL due to main effects often being confounded by interaction effects. Positioning interacting QTL within a small region is even more difficult. We present a variance component approach nested in an empirical Bayesian method, which simultaneously takes into account additive, dominance and epistatic effects due to multiple interacting QTL. The covariance structure used in the variance component approach is based on combined linkage disequilibrium and linkage (LDL) information. In a simulation study where there are complex epistatic interactions between QTL, it is possible to simultaneously fine map interacting QTL using the proposed approach. The present method combined with LDL information can efficiently detect QTL and their dominance and epistatic effects, making it possible to simultaneously fine map main and epistatic QTL.

Key words: Quantitative trait loci (QTL), Combined linkage disequilibrium and linkage (LDL) information, Epistatic effects
doi:10.1631/jzus.2007.B0787 **Document code:** A **CLC number:** Q78; TP31

INTRODUCTION

In general, there are multiple quantitative trait loci (QTL), and their intra- (dominance) and inter-locus interaction (epistasis) underlying phenotypes of a complex trait. If such gene interactions are ignored in a QTL analysis, estimated QTL positions and effects will not be accurate and precise. For considering complicate epistatic interactions in QTL analysis, variance component approaches based on linkage information have been widely used in natural and outbred populations (Mitchell *et al*., 1997; Blangero *et al*., 2000; 2001; Purcell and Sham, 2004). However, the use of linkage information alone may limit the power to detect dominance and epistasis, and mapping resolution based on linkage information is not high.

Linkage disequilibrium (LD) in addition to linkage information (Meuwissen and Goddard, 2001) can give useful extra information about additive, dominance and epistatic effects for small genomic regions. Moreover, a variance component approach nested in an empirical Bayesian method makes it possible to simultaneously map multiple QTL as shown in Lee and van der Werf (2006). Therefore, simultaneous fine mapping of multiple epistatic QTL within a small region would be possible.

The aim of this study is to investigate how much the mapping resolution improves when considering QTL interactions in fine mapping of a complex trait. The posterior QTL density is estimated using an empirical Bayesian approach based on combined linkage disequilibrium and linkage (LDL) information (Lee and van der Werf, 2006) with three different statistical models (additive model, additive and dominance model, and additive, dominance and epistasis model).

* Project supported by the International Pig Improvement Company (PIC) and Sheep Genomics, Australia

MATERIALS AND METHODS

Mixed linear model

A vector of observed phenotypes for the individuals in a pedigree is a linear function of fixed effects, polygenic effects, additive and dominance effects due to n QTL, epistatic interaction among the QTL, and residuals errors. The model can be written as (Cockerham, 1954):

$$\begin{aligned}\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{Zu}+\sum_{i=1}^{n}(\boldsymbol{Za}_i+\boldsymbol{Zd}_i)\\+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\boldsymbol{Za}_i\boldsymbol{a}_j+\boldsymbol{Za}_i\boldsymbol{d}_j+\boldsymbol{Zd}_i\boldsymbol{a}_j+\boldsymbol{Zd}_i\boldsymbol{d}_j)+\boldsymbol{e},\end{aligned}\tag{1}$$

where $\boldsymbol{y}$ is a vector of phenotypic observations on the trait, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{u}$ is a vector of random polygenic effects for each animal, $\boldsymbol{a}_i$ and $\boldsymbol{d}_i$ are vectors of additive and non-additive random effects due to the ith putative QTL, $\boldsymbol{a}_i\boldsymbol{a}_j$, $\boldsymbol{a}_i\boldsymbol{d}_j$, $\boldsymbol{d}_i\boldsymbol{a}_j$ and $\boldsymbol{d}_i\boldsymbol{d}_j$ are vectors of epistatic interactions between the ith and jth putative QTL, and $\boldsymbol{e}$ is a vector of random residual errors. The random effects in the model ($\boldsymbol{u}$, $\boldsymbol{a}_i$, $\boldsymbol{d}_i$, $\boldsymbol{a}_i\boldsymbol{a}_j$, $\boldsymbol{a}_i\boldsymbol{d}_j$, $\boldsymbol{d}_i\boldsymbol{a}_j$, $\boldsymbol{d}_i\boldsymbol{d}_j$ and $\boldsymbol{e}$) are normally distributed with mean zero and variance $\boldsymbol{A}\sigma_{\boldsymbol{u}}^2$, $\boldsymbol{G}_i\sigma_{\boldsymbol{a}_i}^2$, $\boldsymbol{D}_i\sigma_{\boldsymbol{d}_i}^2$, $\boldsymbol{G}_i\boldsymbol{G}_j\sigma_{\boldsymbol{a}_i\boldsymbol{a}_j}^2$, $\boldsymbol{G}_i\boldsymbol{D}_j\sigma_{\boldsymbol{a}_i\boldsymbol{d}_j}^2$, $\boldsymbol{D}_i\boldsymbol{G}_j\sigma_{\boldsymbol{d}_i\boldsymbol{a}_j}^2$, $\boldsymbol{D}_i\boldsymbol{D}_j\sigma_{\boldsymbol{d}_i\boldsymbol{d}_j}^2$, and $\boldsymbol{I}\sigma_{\boldsymbol{e}}^2$, where σ^2 represents the variance of the component, $\boldsymbol{A}$ is a relationship matrix based on pedigree information, $\boldsymbol{G}_i$ and $\boldsymbol{D}_i$ are additive genetic and dominance relationship matrice at the ith putative QTL position, $\boldsymbol{G}_i\boldsymbol{G}_j$, $\boldsymbol{G}_i\boldsymbol{D}_j$, $\boldsymbol{D}_i\boldsymbol{G}_j$, and $\boldsymbol{D}_i\boldsymbol{D}_j$ are the Hadamard product of the additive genetic and dominance relationship matrices at the ith and jth putative QTL positions, and $\boldsymbol{I}$ is an identity matrix. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices for the fixed and random effects in the model. The associated variance covariance matrix ($\boldsymbol{V}$) of all observations from the modeled is:

$$\begin{aligned}\boldsymbol{V}=\boldsymbol{ZAZ}'\sigma_{\boldsymbol{u}}^2+\sum_{i=1}^{n}(\boldsymbol{ZG}_i\boldsymbol{Z}'\sigma_{\boldsymbol{a}_i}^2+\boldsymbol{ZD}_i\boldsymbol{Z}'\sigma_{\boldsymbol{d}_i}^2)+\boldsymbol{I}\sigma_{\boldsymbol{e}}^2\\+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\boldsymbol{ZG}_i\boldsymbol{G}_j\boldsymbol{Z}'\sigma_{\boldsymbol{a}_i\boldsymbol{a}_j}^2+\boldsymbol{ZG}_i\boldsymbol{D}_j\boldsymbol{Z}'\sigma_{\boldsymbol{a}_i\boldsymbol{d}_j}^2\\+\boldsymbol{ZD}_i\boldsymbol{G}_j\boldsymbol{Z}'\sigma_{\boldsymbol{d}_i\boldsymbol{a}_j}^2+\boldsymbol{ZD}_i\boldsymbol{D}_j\boldsymbol{Z}'\sigma_{\boldsymbol{d}_i\boldsymbol{d}_j}^2),\end{aligned}\tag{2}$$

where $\boldsymbol{Z}'$ is the transpose of $\boldsymbol{Z}$.

The matrices $\boldsymbol{G}$ and $\boldsymbol{D}$ are constructed based on identical by descent (IBD) probabilities that are estimated using LDL information (Meuwissen and Goddard, 2000; 2001). For IBD estimation, a Markov chain Monte Carlo (MCMC) method is used, which is robust and efficient especially for complex pedigree, many markers and missing genotypes (Lee *et al*., 2005). These $\boldsymbol{G}$ and $\boldsymbol{D}$ based on the IBD probabilities are incorporated into the QTL model selection in an empirical Bayesian approach (Casella, 2001; Lee and van der Werf, 2006).

Reversible jump Markov chain Monte Carlo (MCMC) for multiple QTL

The number of QTL n, the position of each QTL ρ_i (i=1,...,n) and the model parameters $\Theta=\{\sigma_{\boldsymbol{u}}^2,\ (\sigma_{\boldsymbol{a}_i}^2,\sigma_{\boldsymbol{d}_i}^2;\ i=1,...,n),\ (\sigma_{\boldsymbol{a}_i\boldsymbol{a}_{i+1}}^2,...,\sigma_{\boldsymbol{a}_i\boldsymbol{a}_n}^2,\sigma_{\boldsymbol{a}_i\boldsymbol{d}_{i+1}}^2,...,\sigma_{\boldsymbol{a}_i\boldsymbol{d}_n}^2,\ \sigma_{\boldsymbol{d}_i\boldsymbol{a}_{i+1}}^2,...,\sigma_{\boldsymbol{d}_i\boldsymbol{a}_n}^2,\sigma_{\boldsymbol{d}_i\boldsymbol{d}_{i+1}}^2,...,\sigma_{\boldsymbol{d}_i\boldsymbol{d}_n}^2;\ i=1,...,n-1),\ \sigma_{\boldsymbol{e}}^2\}$ are unknown. Therefore, they should be estimated using Eq.(3). The probability of estimated parameters given observed phenotypes is:

$$pr(n,\rho,\Theta\mid\boldsymbol{y})=\frac{pr(\boldsymbol{y}\mid n,\rho,\Theta)\,pr(n,\rho,\Theta)}{\sum pr(\boldsymbol{y}\mid n,\rho,\Theta)\,pr(n,\rho,\Theta)},\tag{3}$$

where $pr(\boldsymbol{y}|n,\rho,\Theta)$ is the likelihood of the observed phenotypes given the estimated parameters, $pr(n,\rho,\Theta)$ is the joint prior probability of the estimated parameters and the denominator is summed over the probabilities of all possible parameter states. Since the computation of the denominator is not feasible due to a large parameter state space, an MCMC approach can be used for solving this problem (Lee and van der Werf, 2006). In the process, the number of QTL and their positions are sampled from a proposal distribution. In a second step, residual maximum likelihood (REML) is used to estimate the model parameters for a given QTL model. The proposed model, variables and model parameters are accepted or rejected, according to the acceptance ratio from a reversible jump (RJ) MCMC from which the posterior QTL density is derived. Hence, a REML procedure is used nested within a Bayesian RJMCMC.

When considering epistatic interaction among QTL, the number of random effects increases, e.g., the number of random effects due to n QTL is n for

the additive model, $2n$ for the additive and dominance model and $2n+4n(n-1)/2$ for the full model. However, not all epistatic components have significant effects in the model. It may not be necessary to include non-significant epistatic effects that do not improve the model likelihood. For a given set of QTL within a RJMCMC step, each epistatic component is tested and if the likelihood is not improved, the epistatic component will be removed from the model.

Simulated data

One hundred generations of a historical population with effective size of 100 were simulated for 26 markers and 2 QTL in a 130 cM region. For the region from 10 to 20 cM and 110 to 120 cM as candidate regions, markers were densely positioned at 1 cM intervals. The potential QTL were simulated at 17.5 cM (QTL I) and 117.5 cM (QTL II). Unique alleles were assigned to QTL in generation 0, and one allele with moderate frequency (0.1~0.9) was randomly chosen to be the mutation in generation 100. In each generation, the number of male and female parents was 50 and their alleles were transmitted to descendents on the basis of Mendelian segregation using the gene-dropping method (MacCluer *et al*., 1986; Meuwissen and Goddard, 2000).

The number of alleles assumed at each marker locus was 4 in generation 0 and starting allele frequencies were all at 0.25. The marker allele was mutated at a rate of 4×10^{-4} per generation (Dallas, 1992; Weber and Wong, 1993; Ellegren, 1995), i.e. a new allele was introduced as a mutation. Therefore, this historical population would have an equilibrium distribution of alleles in all marker loci and would generate LD among closely linked regions. Note that pedigree and genotype information was deemed not available for these 100 generations. In generation 100 and afterwards, phenotypic values for individuals were simulated as:

$$\boldsymbol{y}=\boldsymbol{\mu}+\boldsymbol{a}_{\text{QTL I}}+\boldsymbol{a}_{\text{QTL II}}+\boldsymbol{d}_{\text{QTL I}}+\boldsymbol{d}_{\text{QTL II}}+\boldsymbol{i}_{\text{QTL I}\times\text{QTL II}}+\boldsymbol{e},$$

where $\boldsymbol{a}$ is additive QTL effects, $\boldsymbol{d}$ is dominance QTL effects, $\boldsymbol{i}$ is the effects due to interaction between QTL I and QTL II. The population mean ($\boldsymbol{\mu}$) was 100, values for residuals ($\boldsymbol{e}$) were from $N(0, \sigma_e^2)$ with $\sigma_e^2=50$. Two data sets were simulated. For the first data, the ratio of additive QTL variance over total phenotypic variance was 0 for the first QTL and 0.11 for the second QTL, the ratio of dominance QTL variance was 0 for both QTL, and the ratio of epistatic QTL variance was 0.2. For the second data, the ratio of additive QTL variance over total phenotypic variance was 0.15 for the first QTL and 0.06 for the second QTL, the ratio of dominance QTL variance was 0.08 for the first QTL and 0.1 for the second QTL, and the ratio of epistatic QTL variance was 0.09.

For QTL mapping results, the posterior QTL density was estimated in RJMCMC LDL mapping with additive effects only (additive model), additive and dominance effects only (additive and dominance model), or additive, dominance and epistasis effects (full model). In all cases, marker genotypes and phenotypes were available for the last 2 generations (200 animals) used for analyses.

RESULTS AND DISCUSSION

When there are complex interactions between QTL, the full model fitting additive, dominance and epistatic effects gives a higher mapping resolution than other reduced models. Fig.1 shows that reduced models fitting additive effects only or additive and dominance effects only without epistatic effects could not catch the signal of the first QTL (Figs.1a and 1b). This is probably due to the fact that the additive and dominance effects of the first QTL are negligible. However, the full model clearly maps the first and second QTL on the true position within a small region (Fig.1c).

Fig.2 shows that the model fitting additive effects only could not detect the second QTL because the second QTL has small additive effects (Fig.2a). However, when the model includes dominance, the second QTL can be detected (Fig.2b). When the model includes epistasis in addition to dominance and additive effects, the mapping resolution becomes higher (Fig.2c).

From the results, when there are complex interactions between QTL, the model considering additive, dominance and epistatic effects can properly capture the signals of the multiple QTL. The empirical Bayesian approach with LDL information (used in this study) can help to simultaneously map multiple interacting QTL within a small region. Therefore, the

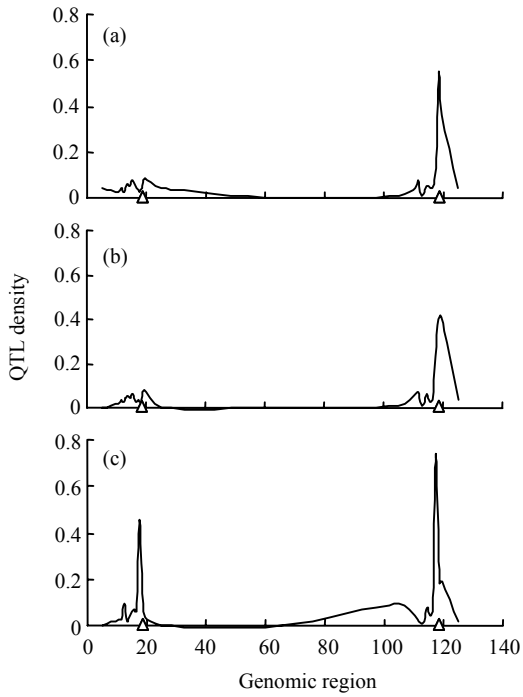

Fig.1 The posterior QTL density with additive model (a), additive and dominance model (b) and full model including additive, dominance and epistasis (c) when using the first data set

Triangle shows the true QTL positions. The ratio of additive variance over phenotypic variance is ~0 and ~0.11 for the first and second QTL respectively. The ratio of dominance variance over phenotypic variance is close to 0 for all QTL. The ratio of epistatic variance over phenotypic variance is 0.2

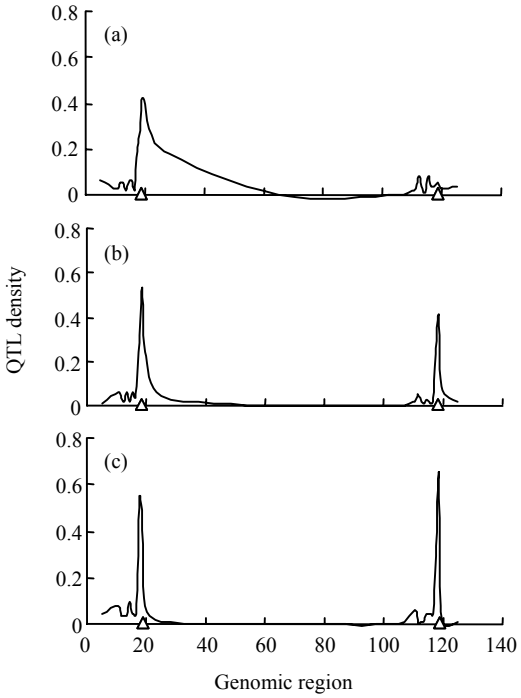

Fig.2 The posterior QTL density with additive model (a), additive and dominance model (b) and full model including additive, dominance and epistasis (c) when using the second data set

Triangle shows the true QTL positions. The ratio of additive variance over phenotypic variance is ~0.15 and ~0.06 for the first and second QTL respectively. The ratio of dominance variance over phenotypic variance is ~0.08 and ~0.1 for the first and second QTL respectively. The ratio of epistatic variance over phenotypic variance is 0.09

present approach would be an efficient tool to study complex traits.

ACKNOWLEDGEMENT

The authors thank the 3rd International Conference on Quantitative Genetics (ICQG) committee for giving an opportunity to present this study in the conference. The trip to ICQG was supported by the collaboration between University of New England and National Institute of Animal Science in Korea.

References

Blangero, J., Williams, J.T., Almasy, L., 2000. Quantitative trait locus mapping using human pedigrees. *Hum. Biol.*, **72**(1):35-62.

Blangero, J., Williams, J.T., Almasy, L., 2001. Variance component methods for detecting complex trait loci. *Adv. Genet.*, **42**:151-181.

Casella, G., 2001. Empirical Bayes Gibbs sampling. *Biostatistics*, **2**(4):485-500. [doi:10.1093/biostatistics/2.4.485]

Cockerham, C.C., 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, **39**(6): 859-882.

Dallas, J.F., 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome*, **3**(8):452-456. [doi:10.1007/BF00356155]

Ellegren, H., 1995. Mutation rates at porcine microsatellite loci. *Mamm. Genome*, **6**(5):376-377. [doi:10.1007/BF00364807]

Lee, S.H., van der Werf, J.H.J., 2006. Simultaneous fine mapping of multiple closely linked quantitative trait loci using combined linkage disequilibrium and linkage with a general pedigree. *Genetics*, **173**(4):2329-2337. [doi:10.1534/genetics.106.057653]

Lee, S.H., van der Werf, J.H.J., Tier, B., 2005. Combining the meiosis Gibbs sampler with the random walk approach for linkage and association studies with a general complex pedigree and multi marker loci. *Genetics*,

171(4):2063-2072. [doi:10.1534/genetics.104.037028]

MacCluer, J.W., VanderBerg, J.L., Raed, B., Ryder, O.A., 1986. Pedigree analysis by computer simulation. *Zoo Biology*, **5**(2):147-160. [doi:10.1002/zoo.1430050209]

Meuwissen, T.H.E., Goddard, M.E., 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*, **155**(1):421-430.

Meuwissen, T.H.E., Goddard, M.E., 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.*, **33**(6):605-634. [doi:10.1051/gse:2001134]

Mitchell, B.D., Ghosh, S., Schneider, J.L., Birznieks, G., Blangero, J., 1997. Power of variance component linkage analysis to detect epistasis. *Genetic. Epidemiology*, **14**(6):1017-1022. [doi:10.1002/(SICI)1098-2272(1997)14:6<1017::AID-GEPI76>3.0.CO;2-L]

Purcell, S., Sham, P.C., 2004. Epistasis in quantitative trait locus linkage analysis. *Behavior Genetics*, **34**(2):143-152. [doi:10.1023/B:BEGE.0000013728.96408.f9]

Weber, J.L., Wong, C., 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**(8):1123-1128. [doi:10.1093/hmg/2.8.1123]