A simplified method for estimation of QTL effects through marker-trait association

BY YUEFU LIU AND PRAMOD K. MATHUR

Canadian Centre for Swine Improvement, Ottawa, Ontario, Canada

Introduction

There are various models for QTL effect estimation:

- Mixture model (Lander and Botstein 1989),
- Regression model (Haley and Knott 1992), and
- Gametic model (Fernando and Grossman 1989).

The gametic model is applicable to the data of general pedigree and can be analyzed with conventional statistical tools. It was proposed for marker-assisted genetic evaluation (Fernando and Grossman 1989) and also used in QTL mapping (Grignola et al. 1996; Zhang et al. 1998).

A drawback of the model is its computing demand due to the large number of QTL effects, especially if non-additive effects are included for refining QTL mapping or marker-assisted genetic evolution. This study extended the gametic model to include dominance effects and suggested computing simplifications for the model.

Covariance between individuals due to QTLs

$$V = A_1\sigma^2_{A_1} + D_1\sigma^2_{D_1} + A_2\sigma^2_{A_2} + D_2\sigma^2_{D_2} + \ldots + (A_1 \# A_2)\sigma^2_{A_1A_2}$$
$$+ (A_1 \# D_2)\sigma^2_{A_1D_2} + (D_1 \# A_2)\sigma^2_{D_1A_2} + (D_1 \# D_2)\sigma^2_{D_1D_2} + \ldots$$

where

$$A_q = \left\{\frac{1}{2}[P(Q_i^1 \equiv Q_j^1 \mid M) + P(Q_i^1 \equiv Q_j^2 \mid M) + P(Q_i^2 \equiv Q_j^1 \mid M) + P(Q_i^2 \equiv Q_j^2 \mid M)]\right\}_{N\times N}$$

$$D_q = \left\{P(Q_i^1 \equiv Q_j^1 \mid M)P(Q_i^2 \equiv Q_j^2 \mid M) + P(Q_i^2 \equiv Q_j^1 \mid M)P(Q_i^1 \equiv Q_j^2 \mid M)\right\}_{N\times N}$$

QTL identity probabilities, e.g. $P(Q_i^1 \equiv Q_j^2 \mid M)$ can be calculated conditional on linked marker and pedigree information according to Wang et al. (1995) and Liu et al. (2002). The gametic relationship matrix conditional on marker information is

$$G_v = \left\{\begin{pmatrix} P(Q_i^1 \equiv Q_j^1 \mid M) & P(Q_i^1 \equiv Q_j^2 \mid M) \\ P(Q_i^2 \equiv Q_j^1 \mid M) & P(Q_i^2 \equiv Q_j^2 \mid M) \end{pmatrix}\right\}_{N\times N}$$

Gametic model with QTL effects and its simplification

Fernando and Grossman (1989) proposed a gametic model for marker-assisted genetic evaluation:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{Wv} + \mathbf{e}$$

where $\mathbf{W} = \mathbf{ZM}$ and $\mathbf{M} = \mathbf{I}_N \otimes [1 \;\; 1]$, $\mathbf{I}_N$ is an identity matrix of dimension equal to the number of individuals to be evaluated (N) and $\otimes$ stands for Kronecker product. The covariance matrices for residual polygenic effects, QTL gametic effects and model residual are $\mathbf{A}\sigma_a^2$, $\mathbf{G}_v\sigma_v^2$ and $\mathbf{I}\,\sigma_e^2$, respectively. The number of mixed model equations for a gametic model is $N(2m+1)$ for N animals and m QTLs.

The gametic effects of individual i at QTL q can be merged into QTL additive effects: $\mathbf{a}_q = \mathbf{Mv}$. The gametic model becomes additive QTL effect model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{Za}_q + \mathbf{e}$$

with the covariance matrix for QTL additive effects equal to

$$V(\mathbf{a}_q) = \mathbf{M}\mathbf{G}_v\mathbf{M}'\sigma_v^2 = 2\mathbf{A}_q\sigma_v^2 = \mathbf{A}_q\sigma_{\mathbf{a}_q}^2$$

since $\mathbf{A}_q = \frac{1}{2}\mathbf{M}\mathbf{G}_v\mathbf{M}'$ (Jamrozik and Schaeffer 1991). $\mathbf{A}_q$ can be converted from gametic relationship matrix $\mathbf{G}_v$ since

$$r_{ij} = \frac{g_{2(i-1)+1,\ 2(j-1)+1} + g_{2(i-1)+1,\ 2(j-1)+2} + g_{2(i-1)+2,\ 2(j-1)+1} + g_{2(i-1)+1,\ 2(j-1)+2}}{2}$$

according to Liu et al. (2002), where r_{ij} is an element of $\mathbf{A}_q$ at row i column j, and $g_{2(i-1)+1,\,2(j-1)+2}$ stands for the element of $\mathbf{G}_v$ at row $2(i-1)+1$ column $2(j-1)+2$.

For genetic evaluation, it is sometimes enough to know the total breeding value. The model can be further simplified as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}_t + \mathbf{e}$$

by letting $\mathbf{a}_t = \mathbf{a} + \mathbf{a}_q$.

Accordingly, $V(a_t) = \mathbf{A}_t\sigma_{a_t}^2 = \mathbf{A}\sigma_a^2 + \mathbf{A}_q\sigma_{a_q}^2$.

Therefore

$$\mathbf{A}_t = \mathbf{A}\frac{\sigma_a^2}{\sigma_{a_t}^2} + \mathbf{A}_q\frac{\sigma_{a_q}^2}{\sigma_{a_t}^2}$$

Dominance model with QTL effects

The model with dominance QTL effects can be expressed as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}(\mathbf{a} + \mathbf{d} + \mathbf{a}_q + \mathbf{d}_q) + \mathbf{e} \qquad (1)$$

where $\mathbf{a}$ and $\mathbf{d}$ stand for additive and dominance effects of residual polygenes while $\mathbf{a}_q$ and $\mathbf{d}_q$ for additive and dominance effects at QTL q. The covariance matrices for $\mathbf{a}$, $\mathbf{d}$, $\mathbf{a}_q$, $\mathbf{d}_q$ and $\mathbf{e}$ are assumed to be $\mathbf{A}\sigma_a^2$, $\mathbf{D}\sigma_d^2$, $\mathbf{A}_q\sigma_{a_q}^2$, $\mathbf{D}_q\sigma_{d_q}^2$ and $\mathbf{I}\sigma_e^2$, where $\mathbf{D}$ and $\mathbf{D}_q$ is the dominance relationship matrix for residual polygenic effects (Smith and Maki-Tanila 1990) and QTL effects. The element of $\mathbf{D}_q$ at row i and column j can be calculated from $\mathbf{G}_v$ (Liu et al. (2002):

$$d_{ij} = g_{2(i-1)+1,\ 2(j-1)+1} \times g_{2(i-1)+2,\ 2(j-1)+2} + g_{2(i-1)+1,\ 2(j-1)+2} \times g_{2(i-1)+2,\ 2(j-1)+1}.$$

For the purpose of genetic evaluation, model (1) can be simplified as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}(\mathbf{a}_t + \mathbf{d}_t) + \mathbf{e} \qquad (2)$$

where $\mathbf{a}_t = \mathbf{a} + \mathbf{a}_q$ and $\mathbf{d}_t = \mathbf{d} + \mathbf{d}_q$.

$V(\mathbf{a}_t) = \mathbf{A}\sigma_a^2 + \mathbf{A}_q\sigma_{a_q}^2$ and $V(\mathbf{d}_t) = \mathbf{D}\sigma_d^2 + \mathbf{D}_q\sigma_{d_q}^2$.

Therefore

$$\mathbf{A}_t = \mathbf{A}\frac{\sigma_a^2}{\sigma_{a_t}^2} + \mathbf{A}_q\frac{\sigma_{a_q}^2}{\sigma_{a_t}^2} \text{ and } \mathbf{D}_t = \mathbf{D}\frac{\sigma_d^2}{\sigma_{d_t}^2} + \mathbf{D}_q\frac{\sigma_{d_q}^2}{\sigma_{d_t}^2}.$$

Computing simplification for dominance models

With dominance effects of polygenes and QTLs, the number of mixed model equations increase considerably. For model (1), the mixed model equations are

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z} & \mathbf{X'Z} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda_a & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{D}^{-1}\lambda_d & \mathbf{Z'Z} & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{A}_q^{-1}\lambda_{a_q} & \mathbf{Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{D}_q^{-1}\lambda_{d_q} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \\ \hat{\mathbf{a}}_q \\ \hat{\mathbf{d}}_q \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \end{bmatrix}$$

In addition, there is no simple method so far for inverting matrix $\mathbf{D}$ and $\mathbf{D}_q$. Fortunately, the computing strategy developed by Schaeffer (2003) can be applied to simplify the computation. In the equations, subtractions of the equation 2 from equations 3, 4 and 5 result in the following formulae, respectively:

$\hat{\mathbf{d}} = \frac{\lambda_a}{\lambda_d}\mathbf{D}\mathbf{A}^{-1}\hat{\mathbf{a}}$, $\hat{\mathbf{a}}_q = \frac{\lambda_a}{\lambda_{a_q}}\mathbf{A}_q\mathbf{A}^{-1}\hat{\mathbf{a}}$ and $\hat{d}_q = \frac{\lambda_a}{\lambda_{d_q}}D_q\mathbf{A}^{-1}\hat{\mathbf{a}}$.

Therefore, solving the equations can be replaced by an iterative procedure of solving a much smaller set of mixed model equations

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda_a \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{pmatrix}$$

and evaluating these three formulae based on the solution from this reduced mixed model equations.

Similarly, the analysis of model (2) can be replaced by the iterative procedure of solving a mixed model equations

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_t^{-1}\lambda_{a_t} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_t \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{pmatrix}$$

and evaluating formula

$$\hat{\mathbf{d}}_t = \frac{\lambda_{a_t}}{\lambda_{d_t}}\mathbf{D}_t\mathbf{A}_t^{-1}\hat{\mathbf{a}}_t .$$

Discussion

This study modified the gametic model to include additive and dominance QTL effects. A computing strategy was proposed for data analysis based on Schaeffer's development (2003). The methods can be readily extended to include epistatic effects. Including non-additive effects in the model can increase the accuracy of QTL mapping, and provides the possibility to refine breeding value estimation by removing non-additive effects. The method that parameterizes QTL effects and residual polygenic effects separately can be used in QTL mapping. Marker-assisted genetic evaluation allows using simpler models and computing procedures by merging polygenic and QTL effects.

References

Fernando and Grossman 1989 Genet. Sel. Evol. 21: 467-477; Grignola et al. 1996 Genet. Sel. Evol. 28: 479-490; Haley and Knott 1992 Heredity 69: 315-324; Lander and Botstein 1989 Genetics 121: 185-199; Liu et al. 2002 Genet. Sel. Evol. 34: 657-678; Schaeffer 2003 J. Anim. Bred. Genet. 120:394-402; Wang et al. 1995 Genet. Sel. Evol. 27: 251-272; Zhang et al. 1998 Genetics 149: 1959-1973.