Original article

The covariance between relatives conditional on genetic markers

Yuefu LIU[a]*, Gerald B. JANSEN[a], Ching Y. LIN[a,b]

[a] Department of Animal and Poultry Science, University of Guelph, Guelph, ON N1G 2W1, Canada
[b] Dairy and Swine Research and Development Centre, Agriculture and Agri-Food, Canada

Abstract – The development of molecular genotyping techniques makes it possible to analyze quantitative traits on the basis of individual loci. With marker information, the classical theory of estimating the genetic covariance between relatives can be reformulated to improve the accuracy of estimation. In this study, an algorithm was derived for computing the conditional covariance between relatives given genetic markers. Procedures for calculating the conditional relationship coefficients for additive, dominance, additive by additive, additive by dominance, dominance by additive and dominance by dominance effects were developed. The relationship coefficients were computed based on conditional QTL allelic transmission probabilities, which were inferred from the marker allelic transmission probabilities. An example data set with pedigree and linked markers was used to demonstrate the methods developed. Although this study dealt with two QTLs coupled with linked markers, the same principle can be readily extended to the situation of multiple QTL. The treatment of missing marker information and unknown linkage phase between markers for calculating the covariance between relatives was discussed.

covariance between relatives / molecular marker / QTL / transmission probability / relationship matrix

1. INTRODUCTION

Quantifying the resemblance between relatives is a fundamental issue in quantitative genetics. It is needed for estimating genetic parameters, predicting breeding values, planning mating schemes, QTL mapping and marker assisted genetic evaluation. The study of the correlation between relatives can be traced back to the beginning of the last century [29, 36]. Kempthorne [22] summarized the work on this topic up to Malecot's study [27]. Fisher [12] first studied the two-locus epistatic deviations and their effects on the covariance between relatives such as parents and descendants, fullsibs, uncles

---

and cousins. Cockerham [6,7] partitioned the two-locus epistatic variance into additive by additive, additive by dominance and dominance by dominance. Kempthorne [21,22] applied the analysis of factorial experiments to partition the genetic variance and studied the covariance between relatives in random mating populations [21,23], inbred populations [24] and a simple autotetraploid population [25]. Plum [31] formulated a recursive method for calculating the relationship and inbreeding coefficients. Cockerham [8] and Weir *et al.* [37] analyzed the influence of linkage on the covariance between relatives. The theory and computational algorithms for the correlation between relatives were well established in the early development of quantitative genetics.

The resemblance between relatives is attributed to gene transmission from the parents to the descendants so that the relatives share identical genes by descent with certain probabilities. Since the gene transmission between generations is not observable, the transmission probability of an allele is generally taken to be 0.5. Actually, the transmission of an allele from a parent to offspring follows an all-or-none pattern. With information from molecular markers, it becomes possible to track the transmission of a linked gene more precisely than by using pedigree data alone.

There have been several studies on the conditional covariance between relatives. Fernando and Grossman [11] developed a method for calculating the gametic covariance conditional on a single linked marker, assuming completely informative markers. Van Arendonk *et al.* [3] designed a computing procedure for the gametic relationship matrix given a single linked marker, which is valid when the parental origin of the offspring's alleles is known. Goddard [16] derived the conditional gametic covariance due to allelic effects in terms of genetic effects without using the concept of identity probabilities, where parental origins of marker alleles and linkage phases among markers are assumed to be known. However, the parental origin of the offspring's alleles is often unknown in real data analysis. Wang *et al.* [35] extended Fernando and Grossman's [11] method to accommodate situations where the parental origin of marker alleles can not be determined unequivocably. However, the method used to account for this biological uncertainty has been developed only for a single marker linked to a QTL. In QTL mapping for human populations, Fulker and Cardon [13,14] used a regression approach to approximate the IBD of QTL from the IBD of flanking markers. Their development is based on the method of Haseman and Elston [18] which considers the expected IBD of a locus as a linear function of the IBD of another linked locus. Kruglyak and Lander [26] developed a hidden Markov model to estimate the IBD states of a putative QTL using the probability distribution of the marker IBDs. This approach is more accurate than Fulker and Cardon's approximation [13,14], but is more complicated to compute. Xu and Gessler [38] made a compromise

between the two methods and proposed an approximate hidden Markov model to improve the computing speed at the expense of estimation accuracy. Almasy and Blangero [2] improved Fulker and Cardon's method [13, 14] in regard to the sib-pair approach of QTL mapping and developed a general framework of multipoint identity by descent. Pong-Wong *et al.* [32] combined the method of Haseman and Elston [18] for estimating identity by descent between sibs often used in human genetics and the method of Wang *et al.* [35] for general pedigree to derive a simple method for calculating the gametic identity-by-descent matrix of QTLs. Meuwissen and Goddard [28] developed a method of predicting gametic identity probability from marker haplotypes by a simplified coalescence process, assuming that the number of generations since the base population and effective population size are known. These studies on conditional identity measures of relatives have generally focused on the identity by descent due to allelic effects. The theory of conditional covariance due to non-additive effects has been little studied. Aside from the covariance due to allelic effects, the quantification of the conditional covariance components due to additive and non-additive effects is also frequently required to refine the statistical model for marker assisted analysis of quantitative traits.

This study aimed to develop a general theory for constructing the conditional covariance between relatives in the presence of additive, dominance and epistatic effects and to update the classical theory when both pedigree and marker data are available. The development relaxed the assumptions of previous studies and applied both single and flanking marker inferences with known or unknown parental origins of offspring's haplotypes.

2. THEORY

2.1. Notations

The notations used in this study basically follow those of Wang *et al.* [35]. Considering an individual i in the population, its lth QTL locus Q_i^l is bracketed by marker loci M_i^l and N_i^l. The recombination rate between M_i^l and Q_i^l is θ_1^l and between Q_i^l and N_i^l is θ_2^l. The recombination rate between the two markers M_i^l and N_i^l is θ^l. The homologous QTL alleles at locus l of individual i are denoted by Q_i^{l1} and Q_i^{l2}. The marker alleles are expressed as M_i^{l1} and M_i^{l2} at the flanking locus M_i^l, and N_i^{l1} and N_i^{l2} at the flanking locus N_i^l. The superscript l will be dropped for simplicity whenever a single QTL is considered. These symbols are random variables. For example, when an individual i has the genotype A_1A_2 at marker locus m, then $M_i^{m1} = A_1$ and $M_i^{m2} = A_2$. The symbol "$\equiv$" stands for the identity between alleles and the symbol "$\Leftarrow$" for the allelic transmission from a parent to a descendant.
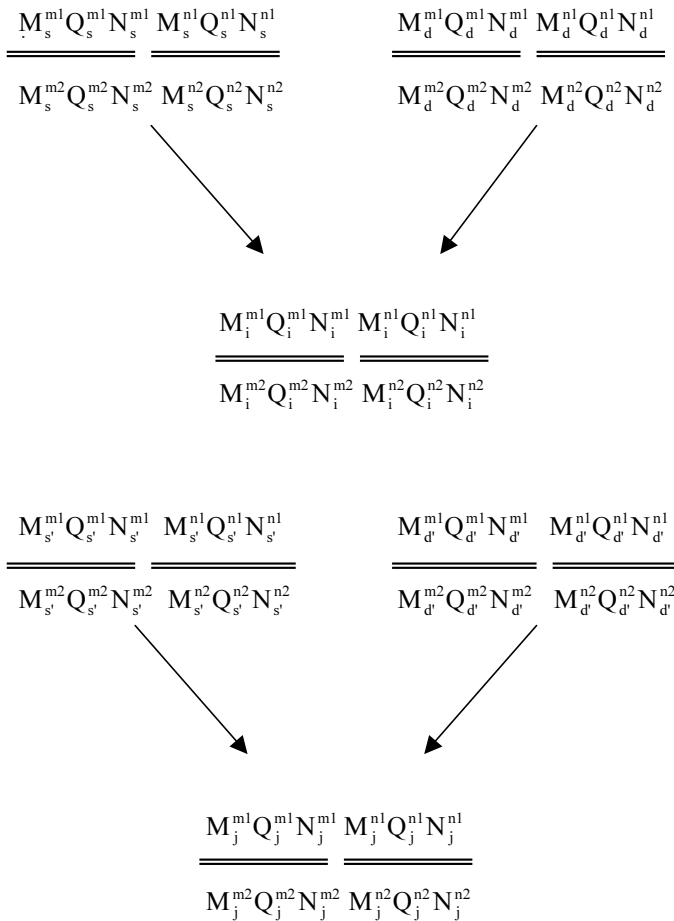
$$\underline{\underline{M_s^{m1}Q_s^{m1}N_s^{m1}}}\ \underline{\underline{M_s^{n1}Q_s^{n1}N_s^{n1}}}$$

$$M_s^{m2}Q_s^{m2}N_s^{m2}\ M_s^{n2}Q_s^{n2}N_s^{n2}$$

$$\underline{\underline{M_d^{m1}Q_d^{m1}N_d^{m1}}}\ \underline{\underline{M_d^{n1}Q_d^{n1}N_d^{n1}}}$$

$$M_d^{m2}Q_d^{m2}N_d^{m2}\ M_d^{n2}Q_d^{n2}N_d^{n2}$$

$$\underline{\underline{M_i^{m1}Q_i^{m1}N_i^{m1}}}\ \underline{\underline{M_i^{n1}Q_i^{n1}N_i^{n1}}}$$

$$M_i^{m2}Q_i^{m2}N_i^{m2}\ M_i^{n2}Q_i^{n2}N_i^{n2}$$

$$\underline{\underline{M_{s'}^{m1}Q_{s'}^{m1}N_{s'}^{m1}}}\ \underline{\underline{M_{s'}^{n1}Q_{s'}^{n1}N_{s'}^{n1}}}$$

$$M_{s'}^{m2}Q_{s'}^{m2}N_{s'}^{m2}\ M_{s'}^{n2}Q_{s'}^{n2}N_{s'}^{n2}$$

$$\underline{\underline{M_{d'}^{m1}Q_{d'}^{m1}N_{d'}^{m1}}}\ \underline{\underline{M_{d'}^{n1}Q_{d'}^{n1}N_{d'}^{n1}}}$$

$$M_{d'}^{m2}Q_{d'}^{m2}N_{d'}^{m2}\ M_{d'}^{n2}Q_{d'}^{n2}N_{d'}^{n2}$$

$$\underline{\underline{M_j^{m1}Q_j^{m1}N_j^{m1}}}\ \underline{\underline{M_j^{n1}Q_j^{n1}N_j^{n1}}}$$

$$M_j^{m2}Q_j^{m2}N_j^{m2}\ M_j^{n2}Q_j^{n2}N_j^{n2}$$

Figure 1. The marker and QTL genotypes for individuals i and j, and their respective parents s, d, and s', d'.

2.2. Genetic covariance components

If there are q loci controlling a quantitative trait, the classical formula for computing the covariance between genotypic values (g) of individuals i and j [21,22] is:

$$Cov(g_i, g_j) = \sum_{t=1}^{q}\sum_{s=0}^{t}(r_{ij})^{t-s}(u_{ij})^{s}\sigma^2_{A^{t-s}D^s} \tag{1}$$

under the assumption of no inbreeding and linkage equilibrium among loci. When there is only one locus ($q = 1$), formula (1) reduces to $Cov(g_i, g_j) = r_{ij}\sigma^2_A + u_{ij}\sigma^2_D$. In this notation, $\sigma^2_{A^2D^1}$ stands for σ^2_{AAD} while $\sigma^2_{A^1D^2}$ stands for

σ^2_{ADD}. Traditionally, the coefficients r_{ij} and u_{ij} are assumed to be identical for the q loci because the allelic transmission at each individual locus can not be traced. Considering only two loci, say m and n, the genetic covariance due to these two QTL loci can be written as:

$$Cov(g_i, g_j) = r_{ij}(\sigma^2_{A_m} + \sigma^2_{A_n}) + u_{ij}(\sigma^2_{D_m} + \sigma^2_{D_n}) + r_{ij}r_{ij}\sigma^2_{A_mA_n}$$
$$+ r_{ij}u_{ij}(\sigma^2_{A_mD_n} + \sigma^2_{D_mA_n}) + u_{ij}u_{ij}\sigma^2_{D_mD_n} \qquad (2)$$

where $\sigma^2_{A_m}$ and $\sigma^2_{A_n}$ are the additive variances of loci m and n, and $\sigma^2_{D_m}$ and $\sigma^2_{D_n}$ are dominance variances at the two loci. The epistatic variances for additive by additive, additive by dominance, dominance by additive and dominance by dominance between loci m and n are $\sigma^2_{A_mA_n}, \sigma^2_{A_mD_n}, \sigma^2_{D_mA_n}$ and $\sigma^2_{D_mD_n}$, respectively.

Information on the markers linked to QTL affecting a trait can be used to refine the covariance among relatives. Conditional on the marker information $\mathcal{M}$, the coefficients r_{ij} and u_{ij} at different loci may vary from locus to locus. Therefore, formula (2) needs to be rewritten as:

$$Cov(g_i, g_j|\mathcal{M}) = r^m_{ij}\sigma^2_{A_m} + r^n_{ij}\sigma^2_{A_n} + u^m_{ij}\sigma^2_{D_m} + u^n_{ij}\sigma^2_{D_n} + r^m_{ij}r^n_{ij}\sigma^2_{A_mA_n}$$
$$+ r^m_{ij}u^n_{ij}\sigma^2_{A_mD_n} + u^m_{ij}r^n_{ij}\sigma^2_{D_mA_n} + u^m_{ij}u^n_{ij}\sigma^2_{D_mD_n} \qquad (3)$$

where r^m_{ij}, r^n_{ij}, u^m_{ij} and u^n_{ij} are the additive and dominance relationship coefficients between individuals i and j at loci m and n, and $r^m_{ij}r^n_{ij}$, $r^m_{ij}u^n_{ij}$, $u^m_{ij}r^n_{ij}$ and $u^m_{ij}u^n_{ij}$ are the relationship coefficients of epistatic interactions between loci m and n. The relationship coefficients r^l_{ij} and u^l_{ij} ($l = m, n$) depend on the conditional probability of QTL allelic identities between individuals i and j:

$$r^l_{ij} = \frac{1}{2}[Pr(Q^{l1}_i \equiv Q^{l1}_j|\mathcal{M}) + Pr(Q^{l1}_i \equiv Q^{l2}_j|\mathcal{M})$$
$$+ Pr(Q^{l2}_i \equiv Q^{l1}_j|\mathcal{M}) + Pr(Q^{l2}_i \equiv Q^{l2}_j|\mathcal{M})] \qquad (4)$$
$$u^l_{ij} = Pr(Q^{l1}_i \equiv Q^{l1}_j|\mathcal{M})Pr(Q^{l2}_i \equiv Q^{l2}_j|\mathcal{M})$$
$$+ Pr(Q^{l1}_i \equiv Q^{l2}_j|\mathcal{M})Pr(Q^{l2}_i \equiv Q^{l1}_j|\mathcal{M}). \qquad (5)$$

This development refines the estimation of genetic covariance and its additive and non-additive components by using marker information that provides locus specific knowledge of QTL allelic transmissions. Therefore, tracing allelic transmission and assessing the conditional probability of QTL allelic identity between relatives are two fundamental issues in this study.

2.3. Conditional probability of QTL allelic identity by descent

For every pair of individuals i and j in a population, there are four possible QTL allelic identities: $(Q^1_i \equiv Q^1_j)$, $(Q^1_i \equiv Q^2_j)$, $(Q^2_i \equiv Q^1_j)$ and $(Q^2_i \equiv Q^2_j)$.

The probabilities of these identities can be inferred conditional on the marker information. Let matrix $\mathbf{P}_{ij}$ contain the probabilities of the four QTL allelic identities between individuals i and j:

$$\mathbf{P}_{ij} = \begin{pmatrix} Pr(Q_i^1 \equiv Q_j^1 | \mathcal{M}) & Pr(Q_i^1 \equiv Q_j^2 | \mathcal{M}) \\ Pr(Q_i^2 \equiv Q_j^1 | \mathcal{M}) & Pr(Q_i^2 \equiv Q_j^2 | \mathcal{M}) \end{pmatrix}.$$

The additive and dominance relationship coefficients between individuals i and j can be obtained from the four elements (p_{11}, p_{12} p_{21} and p_{22}) of $\mathbf{P}_{ij}$ according to formulae (4) and (5):

$$r_{ij} = \frac{1}{2}(p_{11} + p_{12} + p_{21} + p_{22})$$

$$u_{ij} = p_{11} p_{22} + p_{12} p_{21}.$$

Similarly, the QTL allelic identity matrices between i's parents s and d and j's parents s' and d' in the parental generation (Fig. 1) can be defined as $\mathbf{P}_{s s'}$, $\mathbf{P}_{s d'}$, $\mathbf{P}_{d s'}$ and $\mathbf{P}_{d d'}$. For a descendant i, there are eight possible ways to inherit the QTL alleles of the parents s and d. The conditional probabilities of QTL allelic transmission from parents to descendant i can be summarized in matrix $\mathbf{T}_i$:

$$\mathbf{T}_i = \begin{pmatrix} \mathbf{t}_1 & \mathbf{t}_2 \\ \mathbf{t}_3 & \mathbf{t}_4 \end{pmatrix} = \begin{pmatrix} Pr(Q_i^1 \Leftarrow Q_s^1 | \mathcal{M}) & Pr(Q_i^2 \Leftarrow Q_s^1 | \mathcal{M}) \\ Pr(Q_i^1 \Leftarrow Q_s^2 | \mathcal{M}) & Pr(Q_i^2 \Leftarrow Q_s^2 | \mathcal{M}) \\ Pr(Q_i^1 \Leftarrow Q_d^1 | \mathcal{M}) & Pr(Q_i^2 \Leftarrow Q_d^1 | \mathcal{M}) \\ Pr(Q_i^1 \Leftarrow Q_d^2 | \mathcal{M}) & Pr(Q_i^2 \Leftarrow Q_d^2 | \mathcal{M}) \end{pmatrix}$$

where the $\mathbf{t}$'s are all (2×1) column vectors. Similarly, QTL allelic transmission probabilities from parents s' and d' to descendant j can be defined in matrix $\mathbf{T}_j$.

The QTL allelic identity probabilities between individuals i and j, i.e. $\mathbf{P}_{ij}$, can be calculated as:

$$\mathbf{P}_{ij} = \mathbf{T}_i' \begin{pmatrix} \mathbf{P}_{sj} \\ \mathbf{P}_{dj} \end{pmatrix} \tag{6}$$

where

$$\mathbf{P}_{sj} = (\mathbf{P}_{js})' = \left(\mathbf{T}_j' \begin{bmatrix} \mathbf{P}_{s's} \\ \mathbf{P}_{d's} \end{bmatrix} \right)' = \begin{bmatrix} \mathbf{P}_{ss'} & \mathbf{P}_{sd'} \end{bmatrix} \mathbf{T}_j$$

and similarly

$$\mathbf{P}_{dj} = \begin{bmatrix} \mathbf{P}_{ds'} & \mathbf{P}_{dd'} \end{bmatrix} \mathbf{T}_j.$$

Substitution of $\mathbf{P}_{sj}$ and $\mathbf{P}_{dj}$ into formula (6) leads to

$$\mathbf{P}_{ij} = \mathbf{T}_i' \begin{pmatrix} \mathbf{P}_{ss'} & \mathbf{P}_{sd'} \\ \mathbf{P}_{ds'} & \mathbf{P}_{dd'} \end{pmatrix} \mathbf{T}_j. \tag{7}$$

Formula (7) corresponds to Falconer's [10] "basic rule" for calculating coancestry whereas formula (6) relates to the "supplementary rule". Computationally, formula (6) is more efficient than formula (7). Both (6) and (7) indicate that the QTL allelic identity probabilities in a population can be tabulated recursively from ancestors to descendants.

The same principle applies to the derivation of QTL allelic identity probabilities of individual i with itself. Letting $j = i$, $s' = s$ and $d' = d$ in formula (7), and replacing the marginal probabilities with conditional probabilities in $\mathbf{T}_j$ of formula (7) because the allelic transmission from parent to the first allele of offspring i is not independent of that to the second allele, the QTL allelic identity probabilities of individual i with itself ($\mathbf{P}_{ii}$) can be derived from formula (7) and take the following form:

$$
\mathbf{P}_{ii} = \begin{pmatrix} Pr(Q_i^1 \equiv Q_i^1 | \mathcal{M}) & Pr(Q_i^1 \equiv Q_i^2 | \mathcal{M}) \\ Pr(Q_i^2 \equiv Q_i^1 | \mathcal{M}) & Pr(Q_i^2 \equiv Q_i^2 | \mathcal{M}) \end{pmatrix}
$$

$$
= \begin{pmatrix} 1 & \dfrac{\mathbf{t}_1'\mathbf{P}_{sd}\mathbf{t}_4}{\mathbf{1}'\mathbf{t}_1} + \dfrac{\mathbf{t}_3'\mathbf{P}_{ds}\mathbf{t}_2}{\mathbf{1}'\mathbf{t}_3} \\ \dfrac{\mathbf{t}_4'\mathbf{P}_{ds}\mathbf{t}_1}{\mathbf{1}'\mathbf{t}_1} + \dfrac{\mathbf{t}_2'\mathbf{P}_{sd}\mathbf{t}_3}{\mathbf{1}'\mathbf{t}_3} & 1 \end{pmatrix} \tag{8}
$$

where $\mathbf{P}_{sd}$ and $\mathbf{t}$'s are as defined above and $\mathbf{1}' = (1\ \ 1)$. Matrix $\mathbf{P}_{ii}$ is always symmetric. When the parental origins of the two QTL alleles are known (e.g. Q_i^1 is from the father and Q_i^2 from mother), formula (8) simplifies to

$$
\mathbf{P}_{ii} = \begin{pmatrix} 1 & \mathbf{t}_1'\mathbf{P}_{sd}\mathbf{t}_4 \\ \mathbf{t}_4'\mathbf{P}_{ds}\mathbf{t}_1 & 1 \end{pmatrix}. \tag{9}
$$

In this situation, there is no dependence between the two events of allelic transmission. Therefore, formulae (6) and (7) can be directly applied to assess the QTL identity probabilities of an individual i with itself when parental origins of offspring's alleles are known. This explains why formula (8) of Van Arendonk *et al.* [3] works in the same way as the method of Wang *et al.* [35] when parental origins are known.

2.4. QTL allelic transmission probabilities

The parental origin of QTL alleles is usually unknown because the QTL allelic transmission is not directly observable. Therefore, the eight transmission probabilities of QTL alleles from parents s and d to descendant i ($\mathbf{T}_i$) have to be assessed based on marker alleles transmitted from parents s and d to the offspring i and genetic distances between QTL and markers. When two flanking markers are available, the transmission probability from QTL allele

k_p ($k_p = 1, 2$) of parent p ($p = s, d$) to allele k_i ($k_i = 1, 2$) of descendant i can be formulated as:

$$Pr(Q_i^{k_i} \Leftarrow Q_p^{k_p}|\mathcal{M}) = \sum_{k'_p=1}^{2} \sum_{k''_p=1}^{2} Pr(Q_i^{k_i} \Leftarrow Q_p^{k_p}|M_i^{k_i}N_i^{k_i} \Leftarrow M_p^{k'_p}N_p^{k''_p})$$

$$Pr(M_i^{k_i}N_i^{k_i} \Leftarrow M_p^{k'_p}N_p^{k''_p}|\mathcal{M})$$

where $Pr(Q_i^{k_i} \Leftarrow Q_p^{k_p}|M_i^{k_i}N_i^{k_i} \Leftarrow M_p^{k'_p}N_p^{k''_p})$ is the conditional probability given in the 5th column of Table I when $k_p = 1$ and in the 6th column when $k_p = 2$.

Matrix $\mathbf{T}_i$ can now be expressed in terms of marker allelic transmission probabilities, $\mathbf{S}_i$, and recombination rates between QTL and markers and between flanking markers:

$$\mathbf{T}_i = \mathbf{\Theta}\,\mathbf{S}_i \tag{10}$$

where

$\mathbf{\Theta} =$

$$\begin{pmatrix} \frac{(1-\theta_1)(1-\theta_2)}{1-\theta} & \frac{(1-\theta_1)\theta_2}{\theta} & \frac{\theta_1(1-\theta_2)}{\theta} & \frac{\theta_1\theta_2}{1-\theta} & 0 & 0 & 0 & 0 \\ \frac{\theta_1\theta_2}{1-\theta} & \frac{\theta_1(1-\theta_2)}{\theta} & \frac{(1-\theta_1)\theta_2}{\theta} & \frac{(1-\theta_1)(1-\theta_2)}{1-\theta} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{(1-\theta_1)(1-\theta_2)}{1-\theta} & \frac{(1-\theta_1)\theta_2}{\theta} & \frac{\theta_1(1-\theta_2)}{\theta} & \frac{\theta_1\theta_2}{1-\theta} \\ 0 & 0 & 0 & 0 & \frac{\theta_1\theta_2}{1-\theta} & \frac{\theta_1(1-\theta_2)}{\theta} & \frac{(1-\theta_1)\theta_2}{\theta} & \frac{(1-\theta_1)(1-\theta_2)}{1-\theta} \end{pmatrix}$$

and

$$\mathbf{S}_i = \begin{pmatrix} Pr(M_i^1N_i^1 \Leftarrow M_s^1N_s^1|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_s^1N_s^1|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_s^1N_s^2|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_s^1N_s^2|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_s^2N_s^1|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_s^2N_s^1|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_s^2N_s^2|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_s^2N_s^2|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_d^1N_d^1|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_d^1N_d^1|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_d^1N_d^2|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_d^1N_d^2|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_d^2N_d^1|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_d^2N_d^1|\mathcal{M}) \\ Pr(M_i^1N_i^1 \Leftarrow M_d^2N_d^2|\mathcal{M}) & Pr(M_i^2N_i^2 \Leftarrow M_d^2N_d^2|\mathcal{M}) \end{pmatrix}.$$

Note that M_i^1 and N_i^1 always stem from the same parent, so do M_i^2 and N_i^2.

When only a single linked marker is available, the situation simplifies to the case of Wang *et al.* [35]. Formula (10) is identical to formula (5) of Wang *et al.* [35] if their $\mathbf{B}$ matrix is transposed.

Table I. Conditional probabilities of QTL alleles of a descendant given marker haplotypes transmitted from parent p.

$M_p^{k'_p} N_p^{k''_p}$	$Pr(M_p^{k'_p} N_p^{k''_p})$	$Pr(M_p^{k'_p} Q_p^1 N_p^{k''_p})$	$Pr(M_p^{k'_p} Q_p^2 N_p^{k''_p})$	$Pr(Q_p^1 \mid M_p^{k'_p} N_p^{k''_p})$	$Pr(Q_p^2 \mid M_p^{k'_p} N_p^{k''_p})$
$M_p^1 N_p^1$	$\dfrac{1-\theta}{2}$	$\dfrac{(1-\theta_1)(1-\theta_2)}{2}$	$\dfrac{\theta_1\theta_2}{2}$	$\dfrac{(1-\theta_1)(1-\theta_2)}{1-\theta}$	$\dfrac{\theta_1\theta_2}{1-\theta}$
$M_p^1 N_p^2$	$\dfrac{\theta}{2}$	$\dfrac{(1-\theta_1)\theta_2}{2}$	$\dfrac{\theta_1(1-\theta_2)}{2}$	$\dfrac{(1-\theta_1)\theta_2}{\theta}$	$\dfrac{\theta_1(1-\theta_2)}{\theta}$
$M_p^2 N_p^1$	$\dfrac{\theta}{2}$	$\dfrac{\theta_1(1-\theta_2)}{2}$	$\dfrac{(1-\theta_1)\theta_2}{2}$	$\dfrac{\theta_1(1-\theta_2)}{\theta}$	$\dfrac{(1-\theta_1)\theta_2}{\theta}$
$M_p^2 N_p^2$	$\dfrac{1-\theta}{2}$	$\dfrac{\theta_1\theta_2}{2}$	$\dfrac{(1-\theta_1)(1-\theta_2)}{2}$	$\dfrac{\theta_1\theta_2}{1-\theta}$	$\dfrac{(1-\theta_1)(1-\theta_2)}{1-\theta}$

* θ_1 and θ_2 are the recombination rates of a QTL with marker loci M and N, respectively. θ is the recombination rate between two markers. Assume the parent p is in the coupling phase. $M_p^{k'_p} N_p^{k''_p}$ stands for $M_i^{k_i} N_i^{k_i} \Leftarrow M_p^{k'_p} N_p^{k''_p}$.

2.5. Marker haplotype transmission probability

Although marker genotypes can be observed through genotyping techniques, the parental origin of a descendant's haplotype is often uncertain. For example, if a descendant and its parents all have genotype A_1A_2 at a single marker, there is no way to ascertain which parent the descendant's haplotypes come from. Furthermore, when a parent is homozygous, it is impossible to determine which parental gamete a descendant's haplotype comes from. In this development, we trace all possible paths from parental gametes to a descendant's marker haplotype. Because the inference is always conditional on marker information, the notation for conditioning on marker information ($|\mathcal{M}$) will be dropped hereafter for ease of presentation.

The assessment of the marker haplotype transmission involves three steps. First, the transmission probabilities of each path from parental gametes to a descendant's haplotype needs to be quantified. For this, we need to infer which parent a descendant's haplotype comes from (parental origin), and which parental gamete type the descendant's haplotype originates from given the parental origin (gametic frequency). The probability of each transmission path is a probabilistic product of the parental origin and the gametic frequency given parental origin, following the Law of Compound Probability [5]. There are four mutually exclusive paths for each descendant's haplotype in a single marker case and eight in a flanking marker case. Second, we need to determine the probabilities of each descendant's haplotype given the transmission path from a parental gamete to the descendant's haplotype. This can be done by comparing the descendant's haplotype with the parental gametic type. Third, our purpose is to determine the probabilities of each transmission path from parental gametes to a descendant's haplotype given that the descendant's haplotype is observed. This requires calculating the reverse probability of each path given the observed haplotype of the descendant using the Bayes Theorem [5].

Consider the single marker case first. A marker haplotype $M_i^{k_i}$ of descendant i may come from the first or second paternal allele ($M_i^{k_i} \Leftarrow M_s^1$) or ($M_i^{k_i} \Leftarrow M_s^2$), or the first or second maternal allele ($M_i^{k_i} \Leftarrow M_d^1$) or ($M_i^{k_i} \Leftarrow M_d^2$). They are four mutually exclusive events. The transmission probability of each path above is a product of the probability of the parental origin of marker haplotype, $Pr(M_i^{k_i} \Leftarrow p)$ (for $p = s, d$), and the gametic frequency of parents given the parental origin:

$$Pr(M_i^{k_i} \Leftarrow M_p^{k_p}) = Pr(M_p^{k_p}|M_i^{k_i} \Leftarrow p)Pr(M_i^{k_i} \Leftarrow p).$$

There are two possible parental origins for $M_i^{k_i}$. It may be paternal, *i.e.* $Pr(M_i^{k_i} \Leftarrow s) = 1$ and $Pr(M_i^{k_i} \Leftarrow d) = 0$, or maternal, *i.e.* $Pr(M_i^{k_i} \Leftarrow s) = 0$ and $Pr(M_i^{k_i} \Leftarrow d) = 1$. When the parental origin can not be inferred, both $Pr(M_i^{k_i} \Leftarrow s)$ and $Pr(M_i^{k_i} \Leftarrow d)$ are assumed to be 0.5. The two probabilities

sum to one as expected. In the single marker case, the frequencies of parental gametes given parental origins are all 0.5.

For each realization of $M_i^{k_i}$, say $M_i^{k_i} = A_1$, we determine the probability of $M_i^{k_i} = A_1$ given that $M_i^{k_i}$ is transmitted from $M_p^{k_p}$, i.e. $Pr(M_i^{k_i} = A_1 | M_i^{k_i} \Leftarrow M_p^{k_p})$ for $p = s, d$ and $k_p = 1, 2$. Now we can update the four transmission probabilities of $M_i^{k_i} \Leftarrow M_p^{k_p}$ given $M_i^{k_i} = A_1$ using Bayes theorem:

$$Pr(M_i^{k_i} \Leftarrow M_p^{k_p} | M_i^{k_i} = A_1) = \frac{Pr(M_i^{k_i} \Leftarrow M_p^{k_p}) \, Pr(M_i^{k_i} = A_1 | M_i^{k_i} \Leftarrow M_p^{k_p})}{\displaystyle\sum_{p=s}^{d} \sum_{k_p=1}^{2} Pr(M_i^{k_i} \Leftarrow M_p^{k_p}) \, Pr(M_i^{k_i} = A_1 | M_i^{k_i} \Leftarrow M_p^{k_p})}$$

for $k_i = 1, 2$. Since $Pr(M_i^{k_i} = A_1 | M_i^{k_i} \Leftarrow M_p^{k_p})$ is unity when $M_i^{k_i}$ and $M_p^{k_p}$ are identical, and is zero when they are not, this formula can be rewritten as:

$$Pr(M_i^{k_i} \Leftarrow M_p^{k_p} | M_i^{k_i} = A_1) = \frac{Pr(M_i^{k_i} \Leftarrow M_p^{k_p}) \, I(M_i^{k_i} \equiv M_p^{k_p})}{\displaystyle\sum_{p=s}^{d} \sum_{k_p=1}^{2} Pr(M_i^{k_i} \Leftarrow M_p^{k_p}) \, I(M_i^{k_i} \equiv M_p^{k_p})} \quad (11)$$

where $I(M_i^{k_i} \equiv M_p^{k_p})$ is an indicator function, which is equal to one if $M_i^{k_i} \equiv M_p^{k_p}$, and zero otherwise.

In the case of flanking markers, there are eight mutually exclusive marker transmission paths for each haplotype $M_i^{k_i} N_i^{k_i}$ ($k_i = 1, 2$) of descendant i: $M_i^{k_i} N_i^{k_i} \Leftarrow M_s^1 N_s^1$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_s^1 N_s^2$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_s^2 N_s^1$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_s^2 N_s^2$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_d^1 N_d^1$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_d^1 N_d^2$, $M_i^{k_i} N_i^{k_i} \Leftarrow M_d^2 N_d^1$ and $M_i^{k_i} N_i^{k_i} \Leftarrow M_d^2 N_d^2$. Here, $M_i^{k_i}$ and $N_i^{k_i}$ always stem from the same parent. The probability of each transmission path from a parental gamete to a descendant's haplotype is:

$$Pr(M_i^{k_i} N_i^{k_i} \Leftarrow M_p^{k_p} N_p^{k'_p}) = Pr(M_p^{k_p} N_p^{k'_p} | M_i^{k_i} N_i^{k_i} \Leftarrow p) Pr(M_i^{k_i} N_i^{k_i} \Leftarrow p)$$

where $k_p = 1, 2$ and $k'_p = 1, 2$ index the pth parental alleles at marker locus M and N, respectively. The paternal and maternal gametic frequencies given parental origins are $(1 - \theta)/2$, $\theta/2$, $\theta/2$, and $(1 - \theta)/2$. In a similar way, the probabilities of parental origins of $M_i^{k_i} N_i^{k_i}$ can be expressed as $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow s) = 1$ and $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow d) = 0$ for the paternal origin, and $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow s) = 0$ and $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow d) = 1$ for the maternal origin. If the parental origin can not be inferred, it is assumed that both $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow s)$ and $Pr(M_i^{k_i} N_i^{k_i} \Leftarrow d)$ are 0.5. The updated probability of each path given that $M_i^{k_i} N_i^{k_i} = A_1 B_1$ are

observed can be expressed as:

$$Pr(M_i^{k_i} N_i^{k_i} \Leftarrow M_p^{k_p} N_p^{k_p'} | M_i^{k_i} N_i^{k_i} = A_1 B_1)$$

$$= \frac{Pr(M_i^{k_i} N_i^{k_i} \Leftarrow M_p^{k_p} N_p^{k_p'})\, I(M_i^{k_i} N_i^{k_i} \equiv M_p^{k_p} N_p^{k_p'})}{\sum\limits_{p=s}^{d} \sum\limits_{k_p=1}^{2} \sum\limits_{k_p'=1}^{2} Pr(M_i^{k_i} N_i^{k_i} \Leftarrow M_p^{k_p} N_p^{k_p'})\, I(M_i^{k_i} N_i^{k_i} \equiv M_p^{k_p} N_p^{k_p'})} \quad (12)$$

where $I(M_i^{k_i} N_i^{k_i} \equiv M_p^{k_p} N_p^{k_p'})$ is an indicator function.

To clarify the description above, the stepwise calculation of the transmission probabilities of each marker haplotype is given for an example pedigree in which descendant $\frac{A_1 B_2}{A_2 B_2}$ results from the father $\frac{A_1 B_1}{A_2 B_2}$ and the mother $\frac{A_1 B_2}{A_1 B_2}$. The elements of matrix $\mathbf{S}_i$, *i.e.* the probabilities for each marker allelic transmission path from parental gametes to the descendant's haplotypes given the descendant's haplotypes are observed ($M_i^1 N_i^1 = A_1 B_2$ and $M_i^2 N_i^2 = A_2 B_2$), are listed in Table II, together with the intermediate results, such as the probabilities of parental origins, probabilities of parental gametic frequencies and the probabilities of $M_i^{k_i} N_i^{k_i} = A_1 B_2$ and $M_i^{k_i} N_i^{k_i} = A_2 B_2$ given the marker transmission paths.

When the linkage phase can not be ascertained in the flanking marker situation, the marker transmission probability can be estimated by its expectation. For example, if only one individual's phase in the related individuals (including sire s, dam d and descendant i) is unknown, the expectation of $\mathbf{S}_i$ of descendant i,

$$E(\mathbf{S}_i) = \mathbf{S}_i^c\, p^c + \mathbf{S}_i^r\, p^r,$$

can be used to calculate the QTL transmission probabilities (matrix $\mathbf{T}_i$), where $\mathbf{S}_i^c$ and $\mathbf{S}_i^r$ are estimated from coupling and repulsion phases, respectively. The probabilities of the coupling phase (p^c) and repulsion phase (p^r) will be inferred based on the pedigree and marker information of parents, mates and offspring. In general, the expectation of $\mathbf{S}_i$ can be estimated as:

$$E(\mathbf{S}_i) = \sum_{f_s=c}^{r} \sum_{f_d=c}^{r} \sum_{f_i=c}^{r} \mathbf{S}_i^{f_s f_d f_i}\, p_s^{f_s}\, p_d^{f_d}\, p_i^{f_i}$$

since the unknown linkage phase can occur with the sire, dam and descendant, separately or simultaneously. $\mathbf{S}_i^{f_s f_d f_i}$ is calculated from the linkage phase combination of the sire, dam and descendant ($f_s, f_d, f_i = c$ or r).

3. COMPUTATIONAL PROCEDURE

Computationally, it is convenient to arrange the conditional QTL allelic identity probabilities between relatives in a gametic relationship matrix, $\mathbf{G}$. If

Table II. The stepwise calculation of the marker haplotype transmission probabilities for the example pedigree $\frac{A_1B_1}{A_2B_2} \times \frac{A_1B_2}{A_1B_2} \Rightarrow \frac{A_1B_2}{A_2B_2}$.

For descendant's haplotype $M_i^1 N_i^1 = A_1 B_2$:

Gamete $M_p^{kp} N_p^{kp}$	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_1B_2	A_1B_2	A_1B_2	A_1B_2
Elements of $\mathbf{S}_i$	s_{11}	s_{21}	s_{31}	s_{41}	s_{51}	s_{61}	s_{71}	s_{81}
Prob. of parental origin	$Pr(M_i^1 N_i^1 \Leftarrow s) = 0$				$Pr(M_i^1 N_i^1 \Leftarrow d) = 1$			
Gametic frequency	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$
$I(M_i^1 N_i^1 \equiv M_p^{kp} N_p^{kp})$	0	1	0	0	1	1	1	1
$Pr(M_i^1 N_i^1 \Leftarrow M_p^{kp} N_p^{kp} \vert A_1 B_2)$	0	0	0	0	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$

For descendant's haplotype $M_i^2 N_i^2 = A_2 B_2$:

Gamete $M_p^{kp} N_p^{kp}$	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_1B_2	A_1B_2	A_1B_2	A_1B_2
Elements of $\mathbf{S}_i$	s_{12}	s_{22}	s_{32}	s_{42}	s_{52}	s_{62}	s_{72}	s_{82}
Prob. of parental origin	$Pr(M_i^2 N_i^2 \Leftarrow s) = 1$				$Pr(M_i^2 N_i^2 \Leftarrow d) = 0$			
Gametic frequency	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$
$I(M_i^2 N_i^2 \equiv M_p^{kp} N_p^{kp})$	0	0	0	1	0	0	0	0
$Pr(M_i^2 N_i^2 \Leftarrow M_p^{kp} N_p^{kp} \vert A_2 B_2)$	0	0	0	1	0	0	0	0

* θ stands for the recombination rate between markers.

n individuals in a population are coded successively, $\mathbf{G}$ will consist of 2×2 submatrices as follows:

$$\mathbf{G} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \ldots & \mathbf{P}_{1n} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \ldots & \mathbf{P}_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{P}_{n1} & \mathbf{P}_{n2} & \ldots & \mathbf{P}_{nn} \end{pmatrix}$$

where the (2×2) $\mathbf{P}_{ij}$ submatrix can be calculated as

$$\mathbf{P}_{ij} = \mathbf{S}_i' \, \mathbf{\Theta}' \begin{pmatrix} \mathbf{P}_{sj} \\ \mathbf{P}_{dj} \end{pmatrix}$$

or

$$\mathbf{P}_{ij} = \mathbf{S}_i' \, \mathbf{\Theta}' \begin{pmatrix} \mathbf{P}_{ss'} & \mathbf{P}_{sd'} \\ \mathbf{P}_{ds'} & \mathbf{P}_{dd'} \end{pmatrix} \mathbf{\Theta} \, \mathbf{S}_j.$$

The submatrix $\mathbf{P}_{ij}$ between individuals i and j can be used in turn to compute the $\mathbf{P}$ submatrix of their descendants. Thus, the gametic relationship matrix $\mathbf{G}$ can

be calculated recursively, beginning with the uppermost left submatrix of **G** for founders of the population. The recursive calculation proceeds by processing individuals in the order of increasing birthdate. Once matrix **G** is calculated, the additive and dominance relationship coefficients r_{ij} and u_{ij} can be calculated from the (2×2) submatrix of **G** since matrix $\mathbf{P}_{ij}$ is identical to the submatrix of matrix **G**:

$$\mathbf{P}_{ij} = \begin{pmatrix} g_{2(i-1)+1,2(j-1)+1} & g_{2(i-1)+1,2(j-1)+2} \\ g_{2(i-1)+2,2(j-1)+1} & g_{2(i-1)+2,2(j-1)+2} \end{pmatrix}.$$

Letting $\mathbf{R} = \{r_{ij}\}$ be the additive relationship matrix and $\mathbf{U} = \{u_{ij}\}$ be the dominance relationship matrix at a given QTL locus, then the additive by additive relationship matrix is $\mathbf{R}^m \# \mathbf{R}^n$, the additive by dominance relationship matrix $\mathbf{R}^m \# \mathbf{U}^n$, the dominance by additive relationship matrix $\mathbf{U}^m \# \mathbf{R}^n$, and the dominance by dominance relationship matrix $\mathbf{U}^m \# \mathbf{U}^n$, where the symbol $\#$ is the Hadamard product between two matrices of the same size.

4. NUMERICAL EXAMPLE

A pedigree data set is given (Tab. III) to demonstrate the method described. We assumed that marker locus 1 is linked to QTL 1 with a recombination rate of 0.1. In another linkage group, marker loci 2 and 3 bracket QTL 2. The recombination rate between marker locus 2 and QTL 2 is taken as 0.1 and between QTL 2 and marker locus 3 as 0.2. The parents of individuals 1 and 2 are unknown. These two individuals were assumed to be unrelated and non-inbred. Therefore, the uppermost-left 4×4 submatrix of matrix **G** for these two individuals is an identity matrix for both loci. The matrices of the conditional marker transmission probabilities from parents to offspring are listed in Table IV for both single marker and flanking marker cases. For individual 5, the linkage phase between markers 2 and 3 was uncertain. Given the genotype of individual 5, the probabilities of two possible linkage phases

Table III. Example pedigree and marker data.

Indiv.	Father	Mother	Marker Genotypes		
			Marker 1	Marker 2	Marker 3
1	–	–	12	11	13
2	–	–	34	22	24
3	1	2	13	12	34
4	1	2	23	12	14
5	3	4	33	12	34
6	5	4	32	22	34

Table IV. Conditional transmission probabilities of marker haplotypes ($\mathbf{S}_i$).

Ind.	Marker 1				Markers 2 and 3							
	M_s^1	M_s^2	M_d^1	M_d^2	$M_s^1 N_s^1$	$M_s^1 N_s^2$	$M_s^2 N_s^1$	$M_s^2 N_s^2$	$M_d^1 N_d^1$	$M_d^1 N_d^2$	$M_d^2 N_d^1$	$M_d^2 N_d^2$
$\mathbf{S}_3'$	1	0	0	0	0	0.26	0	0.74	0	0	0	0
	0	0	1	0	0	0	0	0	0	0.26	0	0.74
$\mathbf{S}_4'$	0	1	0	0	0.74	0	0.26	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	0.26	0	0.74
$\mathbf{S}_5'$	0	1	0	0	0.954	0	0	0	0	0.046	0	0
	0	0	0	1	0	0	0.046	0	0	0	0	0.954
$\mathbf{S}_6'$	0.5	0.5	0	0	0	0	0.954	0.046	0	0	0	0
	0	0	1	0	0	0	0	0	0	0	0	1

Table V. Conditional transmission probabilities of QTL alleles ($\mathbf{T}_i$).

Individual	QTL locus 1				QTL locus 2			
	Q_s^{11}	Q_s^{12}	Q_d^{11}	Q_d^{12}	Q_s^{21}	Q_s^{22}	Q_d^{21}	Q_d^{22}
$\mathbf{T}_3'$	0.9	0.1	0	0	0.2	0.8	0	0
	0	0	0.9	0.1	0	0	0.2	0.8
$\mathbf{T}_4'$	0.1	0.9	0	0	0.8	0.2	0	0
	0		0.9	0.1	0	0	0.2	0.8
$\mathbf{T}_5'$	0.1	0.9	0	0	0.928	0.026	0.032	0.014
	0	0	0.1	0.9	0.014	0.032	0.026	0.928
$\mathbf{T}_6'$	0.5	0.5	0	0	0.295	0.705	0	0
	0	0	0.9	0.1	0	0	0.027	0.973

$\frac{13}{24}$ and $\frac{14}{23}$ are 0.954 and 0.046, respectively. Both $\mathbf{S}_5$ and $\mathbf{S}_6$, therefore, were replaced by their expected values. Formula (10) was used to obtain the conditional QTL transmission probabilities for individuals 3, 4, 5 and 6 (Tab. V).

Once the conditional QTL allelic transmission probabilities and QTL allelic identity probabilities of founders were computed, the construction of $\mathbf{G}$ proceeded by computing the submatrices $\mathbf{P}_{ij}$ and $\mathbf{P}_{ii}$. The pedigree data were sorted in ascending order by birth dates and the animal ID was coded sequentially. We calculated the submatrix $\mathbf{P}_{ij}$ for each possible pair of individuals and submatrix $\mathbf{P}_{ii}$ for each individual. The pedigree data, submatrices $\mathbf{P}_{ij}$ and $\mathbf{P}_{ii}$, and elements of $\mathbf{G}$ are all indexed by the successive identification number of the individuals, such that $\mathbf{G}$ can be constructed step by step, recursively. The resulting matrix $\mathbf{G}$ is listed in Table VI.

Table VI. Conditional gametic relationship matrix (**G**) for QTL 1 (upper diagonal) and QTL 2 (lower diagonal).

	Q_1^1	Q_1^2	Q_2^1	Q_2^2	Q_3^1	Q_3^2	Q_4^1	Q_4^2	Q_5^1	Q_5^2	Q_6^1	Q_6^2
Q_1^1	1	0	0	0	0.9	0	0.1	0	0.09	0.01	0.05	0.09
Q_1^2	0	1	0	0	0.1	0	0.9	0	0.01	0.09	0.05	0.81
Q_2^1	0	0	1	0	0	0.9	0	0.9	0.81	0.81	0.81	0.09
Q_2^2	0	0	0	1	0	0.1	0	0.1	0.09	0.09	0.09	0.01
Q_3^1	0.2	0.8	0	0	1	0	0.18	0	0.1	0.018	0.059	0.162
Q_3^2	0	0	0.2	0.8	0	1	0	0.82	0.9	0.738	0.819	0.082
Q_4^1	0.8	0.2	0	0	0.32	0	1	0	0.018	0.1	0.059	0.9
Q_4^2	0	0	0.2	0.8	0	0.68	0	1	0.738	0.9	0.819	0.1
Q_5^1	0.211	0.749	0.008	0.032	0.938	0.035	0.329	0.317	1		0.666	0.833	0.09
Q_5^2	0.024	0.017	0.192	0.768	0.022	0.663	0.030	0.95	0.035	1	0.833	0.18
Q_6^1	0.079	0.232	0.138	0.551	0.292	0.478	0.118	0.679	0.319	0.716	1	0.135
Q_6^2	0.022	0.005	0.195	0.778	0.009	0.662	0.027	0.973	0.04	0.925	0.664	1

Table VII. The identity measures of additive genetic effects for QTL 1 (upper off-diagonal) and QTL 2 (lower off-diagonal).

	Diagonal		Off-diagonal					
	QTL 1	QTL 2	a_1	a_2	a_3	a_4	a_5	a_6
a_1	1	1		0	0.5	0.5	0.1	0.5
a_2	1	1	0		0.5	0.5	0.9	0.5
a_3	1	1	0.5	0.5		0.5	0.878	0.561
a_4	1	1	0.5	0.5	0.5		0.878	0.939
a_5	1.666	1.035	0.5	0.5	0.83	0.670		0.968
a_6	1.135	1.664	0.169	0.831	0.720	0.899	0.999	

The conditional additive relationship matrix (Tab. VII) and dominance relationship matrix (Tab. VIII) are calculated from **G**. Subsequently, the epistatic relationship matrices $\mathbf{R}^1 \# \mathbf{R}^2$, $\mathbf{R}^1 \# \mathbf{U}^2$, $\mathbf{R}^2 \# \mathbf{U}^1$ and $\mathbf{U}^1 \# \mathbf{U}^2$ can be computed.

The **G** matrix can also be calculated recursively by extending it by two rows and two columns for each new individual using matrix $\mathbf{T}_i$, in a manner similar to the construction of numerator relationship matrix [1, 19, 35]. The principle and computing formulae of this study apply to this method. The only difference is that matrix $\mathbf{T}_i$ needs to be filled by place-holding zeros to adjust the positions of the non-zero elements to match the positions of parental QTL identity probabilities in the preceding upper left submatrix of matrix **G**. Taking the first QTL of individual 5 as an example, matrix $\mathbf{T}_5$ needs to be filled with

Table VIII. The identity measures of dominance effects for QTL 1 (upper off-diagonal) and QTL 2 (lower off-diagonal).

	Diagonal		Off-diagonal					
	QTL 1	QTL 2	d_1	d_2	d_3	d_4	d_5	d_6
d_1	1	1		0	0	0	0.008	0.045
d_2	1	1	0		0	0	0.146	0.016
d_3	1	1	0	0		0.148	0.09	0.138
d_4	1	1	0	0	0.218		0.09	0.743
d_5	1.444	1.001	0.021	0.012	0.623	0.313		0.225
d_6	1.018	1.441	0.006	0.215	0.198	0.134	0.324	

zeros as:

$$\mathbf{T}_5^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.9 \end{pmatrix}'.$$

Both 2×2 original submatrices in the extended matrix $\mathbf{T}_i^*$ must correspond to the father's and mother's positions in the upper left submatrix of matrix $\mathbf{G}$. Beginning with the upper-left 4×4 submatrix, the remainder of matrix $\mathbf{G}$ can be obtained by recursive extension.

5. DISCUSSION

5.1. Differences between classical and marker-based methods

Molecular genotyping techniques have a profound impact on the traditional theories of quantitative genetics. They make it possible to analyze a quantitative trait on the basis of individual loci. As a consequence, some theories of quantitative genetics have to be modified to account for marker information. In this study, the classical formulae for the covariance between relatives [7, 21,22] were reformulated to allow for locus-specific information about the resemblance between relatives provided by genetic markers. In classical theory, identity by descent is assumed to be equal among loci. Therefore, σ_A^2 and σ_D^2 are summed over all loci affecting a quantitative trait and the epistatic variances are also summed over all possible combinations of the same kind of interactions. In the new formulation, the identities by descent are generally not equal among loci.

The coancestry, which is half of the additive relationship coefficient r_{ij} in the case of no inbreeding [9], needs to be modified as well to incorporate marker information. The coancestry f_{ij} between two individuals, i and j, depends on the relationship of i's parents s and d with j's parents s' and d' [10]:

$$f_{ij} = \frac{1}{4}\left(f_{ss'} + f_{sd'} + f_{ds'} + f_{dd'}\right).$$

This equation assumes that the transmission probability of parental alleles to descendants is 0.5 as in the absence of marker information. When marker data are available to track QTL transmission, the formula is no longer optimal. Conditional on markers, the coancestry between individuals i and j needs to be defined as:

$$f_{ij} = \frac{1}{4} \left[Pr(Q_i^1 \equiv Q_j^1 | \mathcal{M}) + Pr(Q_i^1 \equiv Q_j^2 | \mathcal{M}) \right.$$
$$\left. + Pr(Q_i^2 \equiv Q_j^1 | \mathcal{M}) + Pr(Q_i^2 \equiv Q_j^2 | \mathcal{M}) \right].$$

Classically, the identity by descent between relatives depends only on pedigree and reflects the average relatedness between individuals in a population. In this study, the identity measures depend not only on pedigree data, but also on linked marker genotypes and genetic distances between markers and QTLs. They would vary from trait to trait and reflect the actual genetic resemblance between relatives. The conditional identity by descent provides a more accurate measure of genetic resemblance regarding a specific trait, and results in a more accurate analysis of quantitative traits, such as the estimation of genetic parameters.

Furthermore, the classical formula for the covariance between relatives was derived based on the assumption that no linkage exists between quantitative trait loci. When the loci are linked, the joint probability of identical genes among loci is not a product of marginal probabilities. Therefore, the classical method results in a biased estimation if linkage between loci exists and epistatic effects are not negligible [8, 37]. On the contrary to the classical method, the assumption of no linkage between QTLs is not required for the marker-based method as long as the markers are available for tracking QTLs. Consider an individual i with genotype $\frac{Q_i^{11} \; M_i^1 \; Q_i^{21}}{Q_i^{12} \; M_i^2 \; Q_i^{22}}$ and an individual j with genotype $\frac{Q_j^{11} \; M_j^1 \; Q_j^{21}}{Q_j^{12} \; M_j^2 \; Q_j^{22}}$, where two QTLs and a marker locus are linked. If marker M is ignored as in the classical method, $P(Q_i^{21} \equiv Q_j^{21} | Q_i^{11} \equiv Q_j^{11}) \neq P(Q_i^{21} \equiv Q_j^{21})$ because of the linkage. On the contrary, conditional on the marker, $P(Q_i^{21} \equiv Q_j^{21} | Q_i^{11} \equiv Q_j^{11}, \; \mathcal{M}) = P(Q_i^{21} \equiv Q_j^{21}) | \mathcal{M})$ because the marker information $\mathcal{M}$ removes the influence of Q_1. That is, with marker information, the joint probability of identical QTL alleles among loci can be expressed as the product of marginal probabilities.

5.2. Incomplete marker information

In data analysis, incomplete marker information, such as missing marker genotypes and unknown linkage phases, presents a tedious problem which might increase computation considerably. A solution to this problem suggested in the literature is to replace the gametic relationship matrix **G** by its expectation

conditional on observed markers M_{obs} [15, 35]:

$$E(\mathbf{G}|M_{\text{obs}}) = \sum_{\omega \in \Omega} \mathbf{G}_\omega \, Pr(\omega|M_{\text{obs}})$$

where $\mathbf{G}_\omega$ is a gametic relationship matrix conditional on a single phase-known marker configuration ω for the pedigree from a set of all possible marker configurations (Ω), and $Pr(\omega|M_{\text{obs}})$ is the probability of the complete marker configuration ω conditional on observed markers. MCMC algorithms are often used in exploring the possible configurations and their probabilities conditional on observed data [33, 34]. George *et al.* [15] gave a detailed review in this regard.

Calculating the expectation of matrix $\mathbf{G}$ for a pedigree with substantial missing marker data usually increases computing time considerably due to the potentially large number of configurations [15]. This computing burden is from the repeated calculation of $\mathbf{G}_\omega$ for each complete marker configuration because the expectation of $\mathbf{G}$ is taken at the last stage of the procedure, after $\mathbf{G}_\omega$'s are obtained. Therefore, we previously suggested taking the expectation of marker transmission probability ($\mathbf{S}_i$) for the issue of unknown linkage phase. Since the expectation of $\mathbf{S}_i$ rather than $\mathbf{G}$ is taken in the early stages of tracing marker origins, the repeated calculations of matrix $\mathbf{G}_\omega$ can be avoided. A similar idea can also be used to tackle the problem of missing markers. For example, the marker genotype of an individual i's mother, d, is missing. Based on the marker genotypes of d's parents, mates and descendants, it is possible to calculate the probability distribution of d's marker genotype: $P(M_d^1 M_d^2 \equiv k) = p^k$, where $\sum p^k = 1$. An $\mathbf{S}_i^k$ for each possible genotype k can be calculated to get the expectation of $\mathbf{S}_i$,

$$E(\mathbf{S}_i) = \sum_k \mathbf{S}_i^k \, p^k.$$

Then $\mathbf{S}_i$ can be replaced with $E(\mathbf{S}_i)$ in calculating matrix $\mathbf{G}$. Since the expectation is taken at the stage of tracing markers, the $\mathbf{G}$ matrix is calculated only once.

Recently, Perez-Enciso *et al.* [30] developed a MCMC method for identity-by-descent probabilities of a chromosome region conditional on DNA markers. This procedure appears promising for the situations of incomplete marker information and could be used for computing the gametic relationship matrix at a QTL by some minor modifications.

5.3. Possible benefits of the present study

There have been several studies on the conditional covariance between relatives for marker assisted genetic evaluation within the framework of mixed model equations [11, 16, 35]. In comparison with these studies, the present study differs as follows:

Y. Liu *et al.*

(1) This study presents a general framework for the conditional covariance between relatives including additive effects, dominance effects and epistatic interactions whereas the previous studies generally focused on the covariance components due to allelic effects. This study extends the classical covariance between relatives to cover the situation where both marker information and pedigree are available.

(2) In real data analysis, a solution for unknown parental origins of the offspring's marker alleles is unavoidable. Among the previous studies on gametic relationship, only Wang *et al.* [35] considered the problem of unknown parental origins and quantified the transmission probabilities by tracing back to both paternal and maternal origins for each allele of an offspring. Their algorithm was developed only for a single marker situation. This study developed a clear, stepwise procedure based on Bayes theorem for calculating marker transmission probabilities, which is applicable to both single and flanking marker situations with known or unknown parental origins. Probability tree analysis [20] shows that the quantification of the marker transmission probability requires assessing the reverse probability of a transmission path for a marker haplotype given the descendant's observed marker haplotype.

(3) This study provides a systematic procedure for constructing the additive and nonadditive relationship matrices conditional on marker information besides the allelic relationship, and makes it possible to model additive and nonadditive effects in the framework of random models. Therefore, it provides the opportunity to improve the models for marker assisted genetic evaluation and QTL mapping by including other QTL effects aside from QTL allelic effects. In marker assisted selection, modeling QTL additive effects, instead of QTL allelic effects [11], will reduce the number of equations and decrease the problem of overparameterization. Also, including nonadditive effects will refine marker assisted genetic evaluation by separating the nonadditive QTL effects from additive effects, thereby improving the estimates of additive effects. Modeling the nonadditive effects may be especially meaningful for refining QTL mapping of livestock using random models. There have been attempts to extend Fernando and Grossman's [11] model to QTL mapping [4, 17] which were generally based on modeling QTL alleles. The feasibility of modeling the additive and nonadditive effects in QTL mapping will certainly improve the accuracy of QTL mapping. This is especially true for QTL mapping based on full-sib designs.

REFERENCES

[1] Abdel-Aziz G., Freeman A.E., A rapid method for computing the inverse of the gametic covariance relationship matrix between relatives for a marked quantitative trait locus, Genet. Sel. Evol. 33 (2001) 153–173.

[2] Almasy L., Blangero J., Multipoint quantitative-trait linkage analysis in general pedigrees, Am. J. Hum. Genet. 62 (1998) 1198–1211.

[3] Van Arendonk J.A.M., Tier B., Kinghorn B.P., Use of multiple genetic markers in prediction of breeding values, Genetics 137 (1994) 319-329.

[4] Bink M.C.A.M., Janss L.L.G., Quaas R.L., Markov Chain Monte Carlo for mapping a quantitative trait locus in outbred populations, Genet. Res. 75 (2000) 231–241.

[5] Brunk H.D., An introduction to mathematical statistics, Third edn., Xerox College Publishing, Lexington, MA, USA, 1975.

[6] Cockerham C.C., Genetic covariation among characteristics of swine, Ph.D. thesis, Iowa State College, Ames, IA, USA, 1952.

[7] Cockerham C.C., An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present, Genetics 39 (1954) 859–882.

[8] Cockerham C.C., Effects of linkage on the covariances between relatives, Genetics 41 (1956) 138–141.

[9] Crow J.F., Kimura M., An introduction to population genetics theory, Happer & Row, New York, 1970.

[10] Falconer D.S., Mackay T.F., Introduction to quantitative genetics, Fourth edn., Longman, UK, 1996.

[11] Fernando R.L., Grossman M., Marker assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 21 (1989) 467–477.

[12] Fisher R.A., The correlation between relatives on the supposition of Mendelian inheritance, Trans Royal Soc. Edinburgh 52 (1918) 399–433.

[13] Fulker D.W., Cardon L.R., A sib-pair approach of interval mapping of quantitative trait loci, Am. J. Hum. Genet. 54 (1994) 1092–1103.

[14] Fulker D.W., Cherny S.S., Cardon L.R., Multipoint interval mapping of quantitative trait loci, using sib pairs, Am. J. Hum. Genet. 56 (1995) 1224–1233.

[15] George A.W., Visscher R.M., Haley C.S., Mapping quantitative trait loci in complex pedigree: A two-step variance component approach, Genetics 156 (2000) 2081–2092.

[16] Goddard M.E., A mixed model for analyses of data on multiple genetic markers, Theor. Appl. Genet. 83 (1992) 878–886.

[17] Grignola F.E., Hoeschele I., Tier B., Mapping quantitative loci in outcross populations *via* residual maximum likelihhod. I. Methodology, Genet. Sel. Evol. 28 (1996) 479–490.

[18] Haseman J.K., Elston R.C., The investigation of linkage between a quantitative trait and a marker locus, Behav. Genet. 2 (1972) 3–19.

[19] Henderson C.R., A general method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, Biometrics 32 (1976) 69–83.

[20] Keller G., Warrack B., Bartel H., Statistics for management and economics, Wadsworth Publishing Company, Belmont, CA, USA, 1988.

[21] Kempthore O., The correlation between relatives in a random mating populations, Proc. Royal Soc. London B 143 (1954) 103–113.

[22] Kempthore O., The theoretical values of correlations between relatives in random mating populations, Genetics 40 (1955) 153–167.

[23] Kempthore O., The correlations between relatives in random mating populations, Cold Spring Harbor Symposia on Quantitative Biology, Vol. XX, 1955, pp. 60–78.

[24] Kempthore O., The correlations between relatives in inbred populations, Genetics 40 (1955) 681–691.

[25] Kempthore O., The correlations between relatives in a simple autotetroploid population, Genetics 40 (1955) 168–174.

[26] Kruglyak L., Lander E.S., Complete multipoint sib-pair analysis of qualitative and quantitative traits, Am. J. Hum. Genet. 57 (1995) 439–545.

[27] Malecot G., Les mathématiques de l'hérédité, Masson et Cie, Paris, 1948.

[28] Meuwissen T.H.E., Goddard M.E., Prediction of identity by descent probabilities from marker-haplotypes, Genet. Sel. Evol. 33 (2001) 605–634.

[29] Pearson K., On a generalized theory of alternative inheritance with special reference to Mendel's laws, Phil. Trans. Royal. Soc. A 203 (1904) 53–86.

[30] Perez-Enciso M., Varona L., Rothschild M.F., Computation of identity by descent probabilities conditional on DNA markers *via* a Monte Carlo Markov Chain method, Genet. Sel. Evol. 32 (2000) 467–482.

[31] Plum M., Computation of inbreeding and relationship coefficients, J. Hered. 45 (1954) 92–94.

[32] Pong-Wong R., George A.W., Woolliams J.A., Haley C.S., A simple and rapid method for calculating identity-by-descent matrices using multiple markers, Genet. Sel. Evol. 33 (2001) 453-471.

[33] Thompson E.A., Monte Carlo estimation of multilocus autozygosity probabilities, Proceeding of the 1994 Interface Conference, Fairfax Station, VA, USA, 1994, pp. 498–506.

[34] Thompson E.A., Heath C., Estimation of conditional multilocus gene identity among relatives, in: Seillier-Moseiwitch F. (Ed.), Statistics in Molecular Biology and Genetics, Verlager-Verlage IMS lecture note series, New York, 1999, pp. 95–113.

[35] Wang T., Fernando R.L., Van Der Beek S., Grossman M., Van Arendonk J.A.M., Covariance between relatives for a marked quantitative trait locus, Genet. Sel. Evol. 27 (1995) 251–272.

[36] Weinberg W., Uber Vererbungsgesetze beim Menschen, Zeit. Ind. Abst. Ver I (1909) 377–392, 440–460; II (1909) 276–330.

[37] Weir B.S., Cockerham C.C., Reynolds J., The effects of linkage and linkage disequilibrium on the covariances of noninbred relatives, Heredity 45 (1980) 351–359.

[38] Xu S., Gessler D.D.G., Multipoint genetic mapping of quantitative trait loci using variable number of sibs per family, Genet. Res. 71 (1998) 73–83.