

Covariance between relatives for a marked quantitative trait locus*

T Wang¹, RL Fernando^{1**}, S van der Beek²,
M Grossman¹, JAM van Arendonk²

¹ University of Illinois, Department of Animal Sciences, 1207, W Gregory Drive, Urbana,
IL 61801, USA;

² Wageningen Agricultural University, Department of Animal Breeding, PO Box 338,
6700 AH Wageningen, The Netherlands

(Received 6 April 1994; accepted 1st February 1995)

Summary – Best linear unbiased prediction (BLUP) can be applied to marker-assisted selection. This application requires computation of the inverse of the conditional covariance matrix (\mathbf{G}_v) of additive effects for the quantitative trait locus (QTL) linked to the marker locus (ML), given marker genotypes. This paper presents theory and algorithms to construct \mathbf{G}_v and to obtain its inverse efficiently. These algorithms are sufficiently general to accommodate situations (1) where paternal or maternal origin of marker alleles cannot be determined and (2) where the marker genotypes of some individuals in the pedigree are unknown.

genetic marker / marker-assisted selection / best linear unbiased prediction / covariance between relatives / gametic relationship

Résumé – Covariance entre apparentés pour un locus de caractère quantitatif marqué. La meilleure prédiction linéaire sans biais (BLUP) s'applique à la sélection assistée par marqueur. Cela demande d'inverser la matrice (\mathbf{G}_v) des covariances entre apparentés des effets génétiques additifs du locus quantitatif lié au locus marqueur, covariances conditionnelles aux génotypes du marqueur. Cet article présente la théorie et les algorithmes pour établir \mathbf{G}_v et pour obtenir son inverse d'une manière efficace. Ces algorithmes sont assez généraux pour prendre en compte des situations i) où l'origine paternelle ou maternelle des allèles marqueurs ne peut pas être déterminée, ii) où le génotype marqueur de certains individus dans le pedigree n'est pas connu.

marqueur génétique / sélection assistée par marqueur / meilleure prédiction linéaire sans biais / covariance entre apparentés / parenté gamétique

* Supported in part by the Illinois Experiment Station, Hatch Projects 35-0345 (RLF) and 35-0367 (MG).

** Correspondence and reprints.

INTRODUCTION

Theory for covariance between relatives provides the basis for use of data from relatives in genetic evaluation. At present, genetic evaluations in animal populations are primarily obtained by best linear unbiased prediction (BLUP; Henderson 1973) using trait phenotypes (T-BLUP). Due to advances in molecular biology, genetic markers are becoming increasingly available for use in genetic evaluation. Several approaches for use in genetic evaluation using marker genotypes and trait phenotypes have been discussed (Geldermann, 1975; Soller, 1978; Soller and Beckmann, 1982; Smith and Simpson, 1986; Kashi *et al*, 1990). In addition, Fernando and Grossman (1989) described how BLUP can be used for genetic evaluation using marker genotypes and trait phenotypes (TM-BLUP). Some strategies have been proposed to make TM-BLUP computationally efficient (Cantet and Smith, 1991; Hoeschele, 1993; van Arendonk *et al*, 1994). TM-BLUP has also been extended to accommodate multiple markers (Goddard, 1992; van Arendonk *et al* 1994).

TM-BLUP requires computation of the inverse of the conditional covariance matrix (\mathbf{G}_v) of additive effects for the quantitative trait locus linked to the marker locus, given marker genotypes. To compute this inverse, Fernando and Grossman (1989) provided an algorithm that required information on the parental (paternal or maternal) origin of marker alleles, in addition to information on marker genotypes. The parental origin of marker alleles in an individual, however, is not always known. For example, if 2 parents and their offspring each has genotype A_1A_2 at the same marker locus, marker allele A_1 in the offspring could have descended from either of the parents, thus the parental origin of A_1 in the offspring is unknown.

The objective of this paper is to present theory and algorithms to compute the conditional covariance matrix and its inverse when parental origin of the marker alleles may not be known. Theory and algorithms are developed for pedigrees where the marker genotype of each individual is known (complete marker data). Application of this theory is given for pedigrees where the marker genotype of some individuals is unknown (incomplete marker data).

Wang *et al* (1991) presented, without proof, a recursive equation to construct \mathbf{G}_v and an efficient algorithm to compute its inverse. This recursive equation has been used by van Arendonk *et al* (1994) and Hoeschele (1993). In the present paper, we prove that the recursive equation holds when marker data are complete, but does not hold generally when marker data are incomplete.

Chevalet *et al* (1984) have described a method to compute \mathbf{G}_v given marker phenotypes. This method does not require knowing the parental origin of marker alleles and can accommodate missing marker phenotypes. The method, however, is not computationally feasible for the large pedigrees typically encountered in animal breeding. Computation of the conditional covariance matrix and its inverse become feasible by conditioning on marker genotypes instead of marker phenotypes.

NOTATION AND ASSUMPTIONS

Consider a single polymorphic marker locus (ML) closely linked to a quantitative trait locus (QTL), which will be referred to as the marked QTL (MQTL). Assume linkage equilibrium between the ML and MQTL. For individual i , let M_i^1 and M_i^2

denote 2 alleles at the ML, and let Q_i^1 and Q_i^2 denote MQTL alleles linked to M_i^1 and M_i^2 as shown below

$$\begin{array}{c} M_i^1 \quad Q_i^1 \\ \hline M_i^2 \quad Q_i^2 \end{array}$$

If the 2 marker alleles for individual i are known, then they will be arbitrarily labelled as M_i^1 and M_i^2 . For example, suppose individual i has marker alleles A_3 and A_1 , then A_3 can be labelled as M_i^1 and A_1 as M_i^2 , or A_1 can be labelled as M_i^1 and A_3 as M_i^2 . If the 2 marker alleles for individual i , however, are unknown, M_i^1 can be any of the marker alleles segregating in the population, and M_i^2 can also be any of the marker alleles. For example, suppose there are 3 marker alleles (A_1, A_2 , and A_3) segregating in the population, then M_i^1 can be A_1, A_2 , or A_3 , and M_i^2 can also be A_1, A_2 , or A_3 .

Further, let v_j^1 and v_i^2 be the additive effects of Q_i^1 and Q_i^2 , and let $\sigma_v^2 = \text{Var}(v_i^1) = \text{Var}(v_i^2)$ be their variance, for $i = 1, \dots, n$. Observed marker genotypes are denoted by \mathbf{G}_{obs} .

COVARIANCE OF MQTL EFFECTS GIVEN COMPLETE MARKER DATA

The conditional covariances of additive effects of MQTL alleles will be derived separately for alleles between individuals and for alleles within an individual.

Covariance between individuals

Suppose s and d are parents of i , and j is not a direct descendant of i (fig 1). The conditional covariance of the additive effects of MQTL alleles $Q_i^{k_i}$ and $Q_j^{k_j}$ in individuals i and j , given the observed marker genotypes (\mathbf{G}_{obs}), is

$$\text{Cov}(v_i^{k_i}, v_j^{k_j} | \mathbf{G}_{obs}) = \Pr(Q_i^{k_i} \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \sigma_v^2 \tag{1}$$

where k_i and k_j can be 1 or 2, and $\Pr(Q_i^{k_i} \equiv Q_j^{k_j} | \mathbf{G}_{obs})$ is the conditional probability that $Q_i^{k_i}$ is identical by descent to $Q_j^{k_j}$ given \mathbf{G}_{obs} (eg, Fernando and Grossman, 1989).

Because individuals s and d are parents of i , $Q_i^{k_i}$ can be identical by descent to $Q_j^{k_j}$ in 1 of 4 ways:

1. $Q_i^{k_i}$ descended from Q_s^1 and Q_s^1 was identical by descent to $Q_j^{k_j}$, denoted by $(Q_i^{k_i} \Leftarrow Q_s^1, Q_s^1 \equiv Q_j^{k_j})$
2. $Q_i^{k_i}$ descended from Q_s^2 and Q_s^2 was identical by descent to $Q_j^{k_j}$, denoted by $(Q_i^{k_i} \Leftarrow Q_s^2, Q_s^2 \equiv Q_j^{k_j})$
3. $Q_i^{k_i}$ descended from Q_d^1 and Q_d^1 was identical by descent to $Q_j^{k_j}$, denoted by $(Q_i^{k_i} \Leftarrow Q_d^1, Q_d^1 \equiv Q_j^{k_j})$

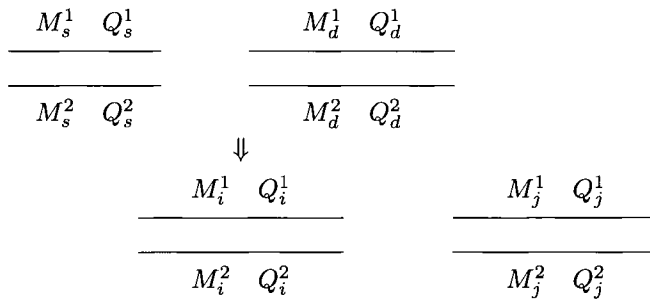


Fig 1. Chromosome fragments containing the ML and the MQTL for individuals *s*, *d*, *i* and *j*.

4. $Q_i^{k_i}$ descended from Q_d^2 and Q_d^2 was identical by descent to $Q_j^{k_j}$, denoted by $(Q_i^{k_i} \leftarrow Q_d^2, Q_d^2 \equiv Q_j^{k_j})$

Therefore, the probability in [1] can be written as

$$\begin{aligned}
 & \Pr(Q_i^{k_i} \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \\
 &= \Pr(Q_i^{k_i} \leftarrow Q_s^1, Q_s^1 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) + \Pr(Q_i^{k_i} \leftarrow Q_s^2, Q_s^2 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \\
 &+ \Pr(Q_i^{k_i} \leftarrow Q_d^1, Q_d^1 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) + \Pr(Q_i^{k_i} \leftarrow Q_d^2, Q_d^2 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \quad [2]
 \end{aligned}$$

Because individual *j* is not a direct descendant of individual *i*, and marker genotypes of *s* and *d* are known, the conditional sampling of $Q_i^{k_i}$ from *s* or *d* is independent of alleles in *j* being identical by descent to alleles in *s* or *d* (fig 1), given \mathbf{G}_{obs} . Thus, the probability in [1] can be computed recursively as

$$\begin{aligned}
 & \Pr(Q_i^{k_i} \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \\
 &= \Pr(Q_i^{k_i} \leftarrow Q_s^1 | \mathbf{G}_{obs}) \Pr(Q_s^1 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) + \Pr(Q_i^{k_i} \leftarrow Q_s^2 | \mathbf{G}_{obs}) \Pr(Q_s^2 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \\
 &+ \Pr(Q_i^{k_i} \leftarrow Q_d^1 | \mathbf{G}_{obs}) \Pr(Q_d^1 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) + \Pr(Q_i^{k_i} \leftarrow Q_d^2 | \mathbf{G}_{obs}) \Pr(Q_d^2 \equiv Q_j^{k_j} | \mathbf{G}_{obs}) \quad [3]
 \end{aligned}$$

Equation [3] was first given by Wang *et al* (1991). It will be shown later that [3] does not hold generally when marker data are incomplete.

Generalizing in [3], $\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs})$ is the conditional probability that allele $Q_i^{k_i}$ in offspring *i* descended from allele $Q_p^{k_p}$ in parent $p = s$ or d for $k_i, k_p = 1$ or 2 . This conditional probability will be referred to as the probability of descent for a QTL allele (PDQ). There are 8 PDQs for each individual, as shown in *Appendix B*, and each PDQ can be expressed as

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs}) = (1 - \rho) \Pr(M_i^{k_i} \leftarrow M_p^1 | \mathbf{G}_{obs}) + \rho \Pr(M_i^{k_i} \leftarrow M_p^2 | \mathbf{G}_{obs}) \quad [4]$$

for $k_i = 1$ or 2 and $p = s$ or d , where $\rho = r$ when $k_p = 1$ and $\rho = 1 - r$ when $k_p = 2$, and where r is the recombination rate between the ML and MQTL. Further, $\Pr(M_i^{k_i} \leftarrow M_p^{k_p} | \mathbf{G}_{obs})$ is the conditional probability that marker allele $M_i^{k_i}$ in offspring i descended from marker allele $M_p^{k_p}$ in parent p , given the pedigree and marker genotypes. This conditional probability will be referred to as the probability of descent for a marker allele (PDM). There are 8 PDMs for each individual, and their computations are explained in *Appendix A*. Note that the PDMs and PDQs associated with the unknown parent(s) are undefined.

Equation [4] explicitly shows the relationship between PDQs and PDMs in scalar notation. For convenience, it is rewritten in matrix notation as

$$\mathbf{B}_i = \mathbf{S}_i \mathbf{R} \quad [5]$$

where

$$\mathbf{B}_i = \begin{bmatrix} \Pr(Q_i^1 \leftarrow Q_s^1 | \mathbf{G}_{obs}) & \Pr(Q_i^1 \leftarrow Q_s^2 | \mathbf{G}_{obs}) & \Pr(Q_i^1 \leftarrow Q_d^1 | \mathbf{G}_{obs}) & \Pr(Q_i^1 \leftarrow Q_d^2 | \mathbf{G}_{obs}) \\ \Pr(Q_i^2 \leftarrow Q_s^1 | \mathbf{G}_{obs}) & \Pr(Q_i^2 \leftarrow Q_s^2 | \mathbf{G}_{obs}) & \Pr(Q_i^2 \leftarrow Q_d^1 | \mathbf{G}_{obs}) & \Pr(Q_i^2 \leftarrow Q_d^2 | \mathbf{G}_{obs}) \end{bmatrix} \quad [6]$$

$$\mathbf{S}_i = \begin{bmatrix} \Pr(M_i^1 \leftarrow M_s^1 | \mathbf{G}_{obs}) & \Pr(M_i^1 \leftarrow M_s^2 | \mathbf{G}_{obs}) & \Pr(M_i^1 \leftarrow M_d^1 | \mathbf{G}_{obs}) & \Pr(M_i^1 \leftarrow M_d^2 | \mathbf{G}_{obs}) \\ \Pr(M_i^2 \leftarrow M_s^1 | \mathbf{G}_{obs}) & \Pr(M_i^2 \leftarrow M_s^2 | \mathbf{G}_{obs}) & \Pr(M_i^2 \leftarrow M_d^1 | \mathbf{G}_{obs}) & \Pr(M_i^2 \leftarrow M_d^2 | \mathbf{G}_{obs}) \end{bmatrix} \quad [7]$$

$$\mathbf{R} = \begin{bmatrix} 1 - r & r & 0 & 0 \\ r & 1 - r & 0 & 0 \\ 0 & 0 & 1 - r & r \\ 0 & 0 & r & 1 - r \end{bmatrix} \quad [8]$$

Covariance within an individual

The conditional covariance between additive effects v_i^1 and v_i^2 of MQTL alleles Q_i^1 and Q_i^2 in individual i with parents s and d , given \mathbf{G}_{obs} , can be written from [1] as

$$\text{Cov}(v_i^1, v_i^2 | \mathbf{G}_{obs}) = \Pr(Q_i^1 \equiv Q_i^2 | \mathbf{G}_{obs}) \sigma_v^2 = f_i \sigma_v^2 \quad [9]$$

where $f_i = \Pr(Q_i^1 \equiv Q_i^2 | \mathbf{G}_{obs})$ is the conditional probability that 2 homologous alleles at the MQTL in individual i are identical by descent, given \mathbf{G}_{obs} . Thus, f_i is the conditional inbreeding coefficient of individual i for the MQTL, given \mathbf{G}_{obs} . This is different from Wright's inbreeding coefficient, which is the conditional probability that 2 homologous alleles at any locus in individual i are identical by descent, given only the pedigree.

The pair of 2 homologous alleles at the MQTL, Q_i^1 and Q_i^2 , in individual i descended from 1 of the following parental pairs: (Q_s^1, Q_d^1) , (Q_s^1, Q_d^2) , (Q_s^2, Q_d^1) or (Q_s^2, Q_d^2) . Let $T_{k_s k_d}$ denote the event that the pair of alleles in i descended from the parental pair $(Q_s^{k_s}, Q_d^{k_d})$ for $k_s, k_d = 1$ or 2 . Now, f_i can be written as

$$f_i = \sum_{k_s=1}^2 \sum_{k_d=1}^2 \Pr(Q_i^1 \equiv Q_i^2 | T_{k_s k_d}, \mathbf{G}_{obs}) \Pr(T_{k_s k_d} | \mathbf{G}_{obs}) \quad [10]$$

Because $(Q_i^1 \equiv Q_i^2 | T_{k_s k_d}, \mathbf{G}_{obs})$ implies $(Q_s^{k_s} \equiv Q_d^{k_d} | \mathbf{G}_{obs})$, [10] becomes

$$f_i = \sum_{k_s=1}^2 \sum_{k_d=1}^2 \Pr(Q_s^{k_s} \equiv Q_d^{k_d} | \mathbf{G}_{obs}) \Pr(T_{k_s k_d} | \mathbf{G}_{obs}) \quad [11]$$

The $\Pr(T_{k_s k_d} | \mathbf{G}_{obs})$ can be expressed in terms of PDQs (see *Appendix C*) as

$$\Pr(T_{k_s k_d} | \mathbf{G}_{obs}) = \frac{\Pr(Q_i^1 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs}) \Pr(Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow Q_s^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_s^2 | \mathbf{G}_{obs})} + \frac{\Pr(Q_i^1 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs}) \Pr(Q_i^2 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow Q_d^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_d^2 | \mathbf{G}_{obs})} \quad [12]$$

For example,

$$\Pr(T_{11} | \mathbf{G}_{obs}) = \frac{B_i(1, 1)B_i(2, 3)}{B_i(1, 1) + B_i(1, 2)} + \frac{B_i(1, 3)B_i(2, 1)}{B_i(1, 3) + B_i(1, 4)}$$

where $B_i(l, k)$ are elements of \mathbf{B}_i in [5]. If 1 of the denominators in [12] is zero, then the entire corresponding term is set to zero.

Tabular method to construct covariance matrix \mathbf{G}_v

The conditional covariance matrix (\mathbf{G}_v) between additive effects of MQTL alleles can be written, from [1] and [9], as

$$\mathbf{G}_v = \mathbf{\Lambda} \sigma_v^2 \quad [13]$$

where $\mathbf{\Lambda}$ is the matrix of conditional probabilities that the 2 homologous alleles at MQTL are identical by descent, given \mathbf{G}_{obs} . The matrix $\mathbf{\Lambda}$ includes a row and column for each of the 2 MQTL alleles in each individual. Thus the order of $\mathbf{\Lambda}$ is $2n$, where n is the number of individuals in the pedigree. This matrix is the conditional gametic relationship matrix (Smith and Allaire, 1985), given \mathbf{G}_{obs} . It follows that each diagonal element of this matrix is unity. The tabular method to construct $\mathbf{\Lambda}$ is explained below.

Following Henderson (1976), individuals are ordered such that parents precede their progeny, and individuals 1 through b are considered to be unrelated and non-inbred. Thus, the upper left submatrix of $\mathbf{\Lambda}$ is an identity matrix of order $2b$, which is expanded sequentially by the 2 rows and 2 columns corresponding to individual i , for $i = b + 1, \dots, n$, as follows:

Let $\delta_i^1 = 2(i - 1) + 1$ and $\delta_i^2 = 2(i - 1) + 2$ be the row indices of $\mathbf{\Lambda}$ corresponding to the 2 MQTL alleles Q_i^1 and Q_i^2 of individual i . From [3], the elements of the 2 rows δ_i^1 and δ_i^2 , corresponding to the 2 MQTL alleles of individual i with parents s and d , are computed as

$$\begin{aligned}\lambda_{\delta_s^1, j} &= B_i(1, 1)\lambda_{\delta_s^1, j} + B_i(1, 2)\lambda_{\delta_s^2, j} + B_i(1, 3)\lambda_{\delta_d^1, j} + B_i(1, 4)\lambda_{\delta_d^2, j} \\ \lambda_{\delta_s^2, j} &= B_i(2, 1)\lambda_{\delta_s^1, j} + B_i(2, 2)\lambda_{\delta_s^2, j} + B_i(2, 3)\lambda_{\delta_d^1, j} + B_i(2, 4)\lambda_{\delta_d^2, j}\end{aligned}\quad [14]$$

for $j = 1, \dots, \delta_i^1 - 1$, where $B_i(l, k)$ were defined in [6]. Element $\lambda_{\delta_s^2, \delta_i^1} = f_i$, where f_i is given in [11]. Elements of columns δ_i^1 and δ_i^2 are obtained by symmetry. If 1 parent is unknown, terms involving the unknown parent are dropped from [14].

For convenience, the tabular algorithm described above can be written in matrix notation. Let $\mathbf{\Lambda}_{i-1}$ be the upper left submatrix of $\mathbf{\Lambda}$ expanded up to $i - 1$. For individual i , with parents s and d , $\mathbf{\Lambda}_{i-1}$ is expanded to $\mathbf{\Lambda}_i$ as

$$\mathbf{\Lambda}_i = \begin{bmatrix} \mathbf{\Lambda}_{i-1} & \mathbf{\Lambda}_{i-1}\mathbf{q}_i \\ \mathbf{q}_i'\mathbf{\Lambda}_{i-1} & \mathbf{C}_i \end{bmatrix}\quad [15]$$

where

$$\mathbf{C}_i = \begin{bmatrix} 1 & f_i \\ f_i & 1 \end{bmatrix}\quad [16]$$

and

$$\mathbf{q}_i' = \begin{bmatrix} 0 & \dots & 0 & B_i(1, 1) & B_i(1, 2) & 0 & \dots & 0 & B_i(1, 3) & B_i(1, 4) & 0 & \dots & 0 \\ 0 & \dots & 0 & B_i(2, 1) & B_i(2, 2) & 0 & \dots & 0 & B_i(2, 3) & B_i(2, 4) & 0 & \dots & 0 \end{bmatrix}\quad [17]$$

In [17], \mathbf{q}_i' is a $2 \times 2(i - 1)$ matrix with at most 8 non-zero elements, which are from \mathbf{B}_i and are located in columns δ_s^1 , δ_s^2 , δ_d^1 and δ_d^2 .

The above tabular algorithm to construct $\mathbf{\Lambda}$ is similar to that used to construct the numerator relationship matrix (Emik and Terrill, 1949; Henderson, 1976). Further, $\mathbf{\Lambda}$ plays the same role in prediction of MQTL effects as the numerator relationship matrix, \mathbf{A} , does in prediction of breeding values.

ALGORITHM TO INVERT COVARIANCE MATRIX OF MQTL ALLELE EFFECTS

Theory

Tier and Sölkner (personal communication, 1994) and van Arendonk *et al* (1994) used partitioned matrix theory to develop rules to invert the numerator relationship matrix efficiently for populations with unusual relationships. A similar approach is used here to invert $\mathbf{\Lambda}$ efficiently.

From [13], $\mathbf{G}_v^{-1} = \mathbf{\Lambda}^{-1}/\sigma_v^2$. In general, the inverse of $\mathbf{\Lambda}_i$, partitioned as in [15], can be obtained as

$$\mathbf{\Lambda}_i^{-1} = \begin{bmatrix} \mathbf{\Lambda}_{i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{q}_i\mathbf{D}_i^{-1}\mathbf{q}_i' & -\mathbf{q}_i\mathbf{D}_i^{-1} \\ -\mathbf{D}_i^{-1}\mathbf{q}_i' & \mathbf{D}_i^{-1} \end{bmatrix}\quad [18]$$

where $\mathbf{D}_i = \mathbf{C}_i - \mathbf{q}_i'\mathbf{\Lambda}_{i-1}\mathbf{q}_i$ is 2×2 matrix (Searle, 1982). From [18], the contribution of individual i to $\mathbf{\Lambda}_i^{-1}$ is given by the second term on the right-hand side of this equation, for which, as shown below, there are at most 36 non-zero elements.

Because of the sparse structure of \mathbf{q}_i as shown in [17], $\mathbf{q}'_i \boldsymbol{\Lambda}_{i-1} \mathbf{q}_i$ can be written as $\mathbf{B}_i \mathbf{C}_{s,d} \mathbf{B}'_i$, where $\mathbf{C}_{s,d}$ is the 4×4 conditional gametic relationship matrix for parents of i , s and d , the elements of which are in $\boldsymbol{\Lambda}_{i-1}$, and \mathbf{B}_i is the matrix of PDQs defined in [6]. Thus

$$\mathbf{D}_i = \mathbf{C}_i - \mathbf{B}_i \mathbf{C}_{s,d} \mathbf{B}'_i \tag{19}$$

If f_i , f_s and f_d are nulle, then

$$\mathbf{D}_i = \mathbf{I}_2 - \mathbf{B}_i \mathbf{B}'_i \tag{20}$$

where \mathbf{I}_2 is a 2×2 identity matrix.

The submatrix $\mathbf{q}_i \mathbf{D}_i^{-1} \mathbf{q}'_i$ in [18] is a square matrix of order $2(i - 1)$ that contains only 16 non-zero elements, which are given by $\mathbf{B}_i \mathbf{D}_i^{-1} \mathbf{B}'_i$. The submatrix $\mathbf{D}_i^{-1} \mathbf{q}'_i$ is a matrix of order $2 \times 2(i - 1)$ that contains only 8 non-zero elements, which are given by $\mathbf{D}_i^{-1} \mathbf{B}'_i$. Thus, there is a total of 36 non-zero elements contributing to $\boldsymbol{\Lambda}_i^{-1}$ from individual i . For convenience, these 36 non-zero elements are collected into a 6×6 matrix:

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{B}_i \mathbf{D}_i^{-1} \mathbf{B}'_i & -\mathbf{B}_i \mathbf{D}_i^{-1} \\ -\mathbf{D}_i^{-1} \mathbf{B}'_i & \mathbf{D}_i^{-1} \end{bmatrix} = \begin{bmatrix} -\mathbf{B}_i \\ \mathbf{I}_2 \end{bmatrix} \mathbf{D}_i^{-1} \begin{bmatrix} -\mathbf{B}'_i & \mathbf{I}_2 \end{bmatrix} \tag{21}$$

Because \mathbf{W}_i contains all contributions to $\boldsymbol{\Lambda}_i^{-1}$ from individual i , we refer to it as the ‘contribution matrix’. The position of contribution element $\mathbf{W}_i(l, k)$ is given by element $\boldsymbol{\Pi}_i(l, k)$, so we define the corresponding ‘position matrix’ for \mathbf{W}_i as

$$\boldsymbol{\Pi}_i = \begin{bmatrix} (\delta_s^1, \delta_s^1) & (\delta_s^1, \delta_s^2) & (\delta_s^1, \delta_d^1) & (\delta_s^1, \delta_d^2) & (\delta_s^1, \delta_i^1) & (\delta_s^1, \delta_i^2) \\ (\delta_s^2, \delta_s^1) & (\delta_s^2, \delta_s^2) & (\delta_s^2, \delta_d^1) & (\delta_s^2, \delta_d^2) & (\delta_s^2, \delta_i^1) & (\delta_s^2, \delta_i^2) \\ (\delta_d^1, \delta_s^1) & (\delta_d^1, \delta_s^2) & (\delta_d^1, \delta_d^1) & (\delta_d^1, \delta_d^2) & (\delta_d^1, \delta_i^1) & (\delta_d^1, \delta_i^2) \\ (\delta_d^2, \delta_s^1) & (\delta_d^2, \delta_s^2) & (\delta_d^2, \delta_d^1) & (\delta_d^2, \delta_d^2) & (\delta_d^2, \delta_i^1) & (\delta_d^2, \delta_i^2) \\ (\delta_i^1, \delta_s^1) & (\delta_i^1, \delta_s^2) & (\delta_i^1, \delta_d^1) & (\delta_i^1, \delta_d^2) & (\delta_i^1, \delta_i^1) & (\delta_i^1, \delta_i^2) \\ (\delta_i^2, \delta_s^1) & (\delta_i^2, \delta_s^2) & (\delta_i^2, \delta_d^1) & (\delta_i^2, \delta_d^2) & (\delta_i^2, \delta_i^1) & (\delta_i^2, \delta_i^2) \end{bmatrix}$$

where $\delta_a^b = 2(a - 1) + b$ for $a = s, d$, or i and $b = 1$ or 2 . If both parents of individual i are known, then all elements in $\boldsymbol{\Pi}_i$ are defined. If at least 1 parent is unknown, then elements in $\boldsymbol{\Pi}_i$ associated with the unknown parent(s) are not defined.

Because \mathbf{q}_i has at most 8 non-zero elements, and the positions of these elements are simple functions of s and d , [18] leads to an efficient algorithm to invert $\boldsymbol{\Lambda}$, where the number of arithmetic operations for inverting is proportional to $2n$, the size of $\boldsymbol{\Lambda}$. It is noteworthy that any symmetric positive definite matrix can be inverted using [18]. Unless \mathbf{q}_i is sparse and the positions of the non-zero elements can be determined easily, this approach will not be efficient. Note that [19] requires $\mathbf{C}_{s,d}$, which is from $\boldsymbol{\Lambda}_{i-1}$. Thus for an inbred pedigree, $\mathbf{C}_{s,d}$ needs first to be computed, similar to the situation where inbreeding coefficients need first to be computed when Henderson’s rapid algorithm (Henderson, 1976) is used to invert a numerator relationship matrix.

Algorithm

1. Set Λ^{-1} equal to the null matrix.
2. For individual i , $i = 1, \dots, n$:
 - (a) if both parents are unknown, then add 1s to $\Lambda_{\delta_i^1, \delta_i^1}^{-1}$ and $\Lambda_{\delta_i^2, \delta_i^2}^{-1}$
 - (b) if at least 1 parent is known, then:
 - i) compute \mathbf{B}_i according to [5]
 - ii) compute \mathbf{D}_i according to [19] for inbreeding or [20] for non-inbreeding
 - iii) compute \mathbf{W}_i according to [21]
 - iv) for each 'defined' element in $\mathbf{\Pi}_i$, add element $\mathbf{W}_i(l, k)$ to Λ^{-1} at the position given by $\mathbf{\Pi}_i(l, k)$

NUMERICAL EXAMPLE WITH COMPLETE MARKER DATA

Consider the pedigree of 5 individuals in table I. These 5 individuals are numbered sequentially so that parents precede their offspring, and are assumed to be from a population with marker allele frequencies of $p(A_1) = 0.7$, $p(A_2) = 0.1$, and $p(A_3) = 0.2$. For convenience, we assumed that $\sigma_v^2 = 1.0$ and $r = 0.1$. For this example, genotype A_2A_2 is assigned to individual 2, so that marker data are complete.

Table I. Pedigree of 5 individuals with marker genotypes.

<i>Individual</i>	<i>Genotype</i>	<i>Parent s</i>	<i>Parent d</i>
1	A_1A_1	0	0
2	Unknown	0	0
3	A_1A_2	1	2
4	A_1A_2	0	2
5	A_1A_2	3	4

Zeros indicate unknown parent.

Computing PDMs

The PDMs are undefined for individuals 1 and 2, because their parents are unknown. Individual 3 has parents 1 and 2. Thus, as shown in *Appendix A*, the 8 PDMs for individual 3 can be computed as

$$\Pr(M_3^{k_3} \leftarrow M_p^{k_p} | \mathbf{G}_{obs}) = \Pr(M_3^{k_3} \leftarrow M_p^{k_p} | G_1, G_2, G_3) \quad [23]$$

for $k_3, k_p, p = 1$ or 2 , where G_1 , G_2 , and G_3 represent marker genotypes of individuals 1, 2, and 3. The right-hand side of [23] can be computed from Mendelian principles (see example after equation [A.1] in *Appendix A*), and the resulting PDMs are stored in matrix \mathbf{S}_3 , defined in [7], as

$$\mathbf{S}_3 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \quad [24]$$

For individual 4, the paternal parent is unknown. Thus, PDMs for individual 4 can be computed as

$$\Pr(M_4^{k_4} \leftarrow M_2^{k_2} | \mathbf{G}_{obs}) = \sum_{i=1}^3 \sum_{j=1}^3 \Pr(M_4^{k_4} \leftarrow M_2^{k_2} | G_u = A_i A_j, G_2, G_4) \Pr(G_u = A_i A_j | G_2, G_4) \quad [25]$$

for $k_4, k_2 = 1$ or 2 where $G_u = A_i A_j$ is the ordered marker genotype for the unknown paternal parent. The upper limit of the summation is the number of marker alleles segregating in the population. The resulting PDMs are

$$\mathbf{S}_4 = \begin{bmatrix} - & - & 0 & 0 \\ - & - & 1/2 & 1/2 \end{bmatrix}$$

The first 2 columns in \mathbf{S}_4 are undefined because the paternal parent is unknown.

For individual 5, both parents are known. Thus, computation of PDMs for individual 5 is similar to that for individual 3, and the resulting PDMs are

$$\mathbf{S}_5 = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

Constructing $\mathbf{\Lambda}$

Individuals 1 and 2 are unrelated and non-inbred (table I), thus the upper left submatrix of the conditional gametic relationship matrix $\mathbf{\Lambda}$ is an identity matrix of order 4. This submatrix will be expanded by the tabular method for individuals 3, 4, and 5, as shown below.

The matrix \mathbf{B}_3 of PDQs for individual 3 with parents 1 and 2 is computed using \mathbf{S}_3 according to [5]:

$$\mathbf{B}_3 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Now, from [14], elements $\lambda_{5,j}$ and $\lambda_{6,j}$, for $j = 1, \dots, 4$, which correspond to individual 3, are computed as linear functions of elements in the first 4 rows, which correspond to the parents 1 and 2:

$$\begin{bmatrix} \lambda_{5,j} \\ \lambda_{6,j} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} \lambda_{1,j} \\ \lambda_{2,j} \\ \lambda_{3,j} \\ \lambda_{4,j} \end{bmatrix}$$

Diagonal elements $\lambda_{5,5}$ and $\lambda_{6,6}$ for individual 3 are unity. Off-diagonal element $\lambda_{6,5}$, which is defined as the conditional inbreeding coefficient in [10], is null because the parents of individual 3 are unrelated. For individual 3, therefore, numerical values

of elements $\lambda_{5,j}$ for $j = 1, \dots, 5$ and $\lambda_{6,j}$ for $j = 1, \dots, 6$ are

j	1	2	3	4	5	6
$\lambda_{5,j}$	1/2	1/2	0	0	1	
$\lambda_{6,j}$	0	0	1/2	1/2	0	1

The corresponding column elements are obtained by symmetry.

The PDQs for individual 4 are computed using [5]:

$$\mathbf{B}_4 = \begin{bmatrix} - & - & 0 & 0 \\ - & - & 1/2 & 1/2 \end{bmatrix}$$

For individual 4, numerical values of elements $\lambda_{7,j}$ for $j = 1, \dots, 7$ and $\lambda_{8,j}$ for $j = 1, \dots, 8$ are

j	1	2	3	4	5	6	7	8
$\lambda_{7,j}$	0	0	0	0	0	0	1	
$\lambda_{8,j}$	0	0	1/2	1/2	0	1/2	0	1

The PDQs for individual 5 are computed using [5]:

$$\mathbf{B}_5 = \begin{bmatrix} 9/20 & 1/20 & 9/20 & 1/20 \\ 1/20 & 9/20 & 1/2 & 9/20 \end{bmatrix}$$

To compute f_5 defined in [10], we need $\Pr(Q_3^{k_3} \equiv Q_4^{k_4} | \mathbf{G}_{obs})$ and $\Pr(T_{k_3 k_4} | \mathbf{G}_{obs})$ for $k_3, k_4 = 1$ or 2. Probabilities, $\Pr(Q_3^{k_3} \equiv Q_4^{k_4} | \mathbf{G}_{obs})$, have already been computed as

$$\begin{aligned} \Pr(Q_3^1 \equiv Q_4^1 | \mathbf{G}_{obs}) &= \lambda_{5,7} = 0 \\ \Pr(Q_3^1 \equiv Q_4^2 | \mathbf{G}_{obs}) &= \lambda_{5,8} = 0 \\ \Pr(Q_3^2 \equiv Q_4^1 | \mathbf{G}_{obs}) &= \lambda_{6,7} = 0 \\ \Pr(Q_3^2 \equiv Q_4^2 | \mathbf{G}_{obs}) &= \lambda_{6,8} = 1/2 \end{aligned}$$

Probabilities, $\Pr(T_{k_3 k_4} | \mathbf{G}_{obs})$, can be obtained according to [12] as

$$\Pr(T_{11} | \mathbf{G}_{obs}) = \frac{B_5(1, 1) B_5(2, 3)}{B_5(1, 1) + B_5(1, 2)} + \frac{B_5(1, 3) B_5(2, 1)}{B_5(1, 3) + B_5(1, 4)} = 9/100$$

Similarly, $\Pr(T_{12} | \mathbf{G}_{obs}) = 41/100$, $\Pr(T_{21} | \mathbf{G}_{obs}) = 41/100$, and $\Pr(T_{22} | \mathbf{G}_{obs}) = 9/100$. Therefore,

$$f_5 = \sum_{k_3=1}^2 \sum_{k_4=1}^2 \Pr(Q_3^{k_3} \equiv Q_4^{k_4} | \mathbf{G}_{obs}) \Pr(T_{k_3 k_4} | \mathbf{G}_{obs}) = 9/200$$

For individual 5, numerical values of elements $\lambda_{9,j}$ for $j = 1, \dots, 9$ and $\lambda_{10,j}$ for $j = 1, \dots, 10$ are

j	1	2	3	4	5	6	7	8	9	10
$\lambda_{9,j}$	0.225	0.225	0.050	0.050	0.450	0.075	0.450	0.075	1.000	
$\lambda_{10,j}$	0.025	0.025	0.450	0.450	0.050	0.675	0.050	0.675	0.045	1.000

The conditional gametic relationship matrix ($\mathbf{\Lambda}$) is

	v_1^1	v_1^2	v_2^1	v_2^2	v_3^1	v_3^2	v_4^1	v_4^2	v_5^1	v_5^2
v_1^1	1.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_1^2	0.000	1.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_2^1	0.000	0.000	1.000	0.000	0.000	0.500	0.000	0.500	0.050	0.450
v_2^2	0.000	0.000	0.000	1.000	0.000	0.500	0.000	0.500	0.050	0.450
v_3^1	0.500	0.500	0.000	0.000	1.000	0.000	0.000	0.000	0.450	0.050
v_3^2	0.000	0.000	0.500	0.500	0.000	1.000	0.000	0.500	0.075	0.675
v_4^1	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.450	0.050
v_4^2	0.000	0.000	0.500	0.500	0.000	0.500	0.000	1.000	0.075	0.675
v_5^1	0.225	0.225	0.050	0.050	0.450	0.075	0.450	0.075	1.000	0.045
v_5^2	0.025	0.025	0.450	0.450	0.050	0.675	0.050	0.675	0.045	1.000

Inverting $\mathbf{\Lambda}$

Set $\mathbf{\Lambda}^{-1}$ to the null matrix. For each of the 5 individuals, the contribution matrix \mathbf{W}_i and corresponding position matrix $\mathbf{\Pi}_i$ are computed as described below. The inverse of $\mathbf{\Lambda}$ is obtained by adding elements $\mathbf{W}_i(l, k)$ to $\mathbf{\Lambda}^{-1}$ at positions indicated by elements $\mathbf{\Pi}_i(l, k)$.

For the first 2 individuals, the parents are unknown. Thus, add 1s to $\mathbf{\Lambda}_{1,1}^{-1}$, $\mathbf{\Lambda}_{2,2}^{-1}$, $\mathbf{\Lambda}_{3,3}^{-1}$ and $\mathbf{\Lambda}_{4,4}^{-1}$. For individual 3, PDQs (\mathbf{B}_3) can be obtained as shown earlier. Because individual 3 is not inbred, $\mathbf{D}_3 = \mathbf{I}_2 - \mathbf{B}_3\mathbf{B}'_3$, from [20]. Matrix \mathbf{W}_3 is in table II and $\mathbf{\Pi}_3$ is in table III.

Similarly, for individual 4, matrices \mathbf{W}_4 and $\mathbf{\Pi}_4$ are in tables II and III. Note that 1 parent of individual 4 is unknown. Those elements in \mathbf{W}_4 and $\mathbf{\Pi}_4$ associated with the unknown parent are undefined.

From the previous section, individual 5 is inbred ($f_5 = 0.045$). Thus, [19] is used to obtain $\mathbf{D}_5 = \mathbf{C}_5 - \mathbf{B}_5\mathbf{C}_{3,4}\mathbf{B}'_5$, where \mathbf{C}_5 and $\mathbf{C}_{3,4}$ were computed in the previous section:

$$\mathbf{C}_5 = \begin{bmatrix} 1 & 9/200 \\ 9/200 & 1 \end{bmatrix}, \quad \mathbf{C}_{3,4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1 \end{bmatrix}$$

Matrices \mathbf{W}_5 and $\mathbf{\Pi}_5$ are given in tables II and III.

Table II. Contribution matrices \mathbf{W}_i for $i = 3, 4, 5$.

$\mathbf{W}_3 =$	$\begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 & -1.000 & 0.000 \\ 0.500 & 0.500 & 0.000 & 0.000 & -1.000 & 0.000 \\ 0.000 & 0.000 & 0.500 & 0.500 & 0.000 & -1.000 \\ 0.000 & 0.000 & 0.500 & 0.500 & 0.000 & -1.000 \\ -1.000 & -1.000 & 0.000 & 0.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 \end{bmatrix}$
$\mathbf{W}_4 =$	$\begin{bmatrix} - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & 0.500 & 0.500 & 0.000 & 1.000 \\ - & - & 0.500 & 0.500 & 0.000 & 1.000 \\ - & - & 0.000 & 0.000 & 1.000 & 0.000 \\ - & - & 1.000 & 1.000 & 0.000 & 2.000 \end{bmatrix}$
$\mathbf{W}_5 =$	$\begin{bmatrix} 0.372 & 0.160 & 0.372 & 0.160 & -0.797 & -0.268 \\ 0.160 & 0.551 & 0.160 & 0.551 & -0.223 & -1.200 \\ 0.372 & 0.160 & 0.372 & 0.160 & -0.797 & -0.268 \\ 0.160 & 0.551 & 0.160 & 0.551 & -0.223 & -1.200 \\ -0.797 & -0.223 & -0.797 & -0.223 & 1.737 & 0.303 \\ -0.268 & -1.200 & -0.268 & -1.200 & 0.303 & 2.633 \end{bmatrix}$

Table III. Position matrices $\mathbf{\Pi}_i$ for $i = 3, 4, 5$.

$\mathbf{\Pi}_3 =$	$\begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix}$
$\mathbf{\Pi}_4 =$	$\begin{bmatrix} - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & (3,3) & (3,4) & (3,7) & (3,8) \\ - & - & (4,3) & (4,4) & (4,7) & (4,8) \\ - & - & (7,3) & (7,4) & (7,7) & (7,8) \\ - & - & (8,3) & (8,4) & (8,7) & (8,8) \end{bmatrix}$
$\mathbf{\Pi}_5 =$	$\begin{bmatrix} (5,5) & (5,6) & (5,7) & (5,8) & (5,9) & (5,10) \\ (6,5) & (6,6) & (6,7) & (6,8) & (6,9) & (6,10) \\ (7,5) & (7,6) & (7,7) & (7,8) & (7,9) & (7,10) \\ (8,5) & (8,6) & (8,7) & (8,8) & (8,9) & (8,10) \\ (9,5) & (9,6) & (9,7) & (9,8) & (9,9) & (9,10) \\ (10,5) & (10,6) & (10,7) & (10,8) & (10,9) & (10,10) \end{bmatrix}$

The \mathbf{A}^{-1} matrix is

	v_1^1	v_1^2	v_2^1	v_2^2	v_3^1	v_3^2	v_4^1	v_4^2	v_5^1	v_5^2
v_1^1	1.500	0.500	0.000	0.000	-1.000	0.000	0.000	0.000	0.000	0.000
v_1^2	0.500	1.500	0.000	0.000	-1.000	0.000	0.000	0.000	0.000	0.000
v_2^1	0.000	0.000	2.000	1.000	0.000	-1.000	0.000	-1.000	0.000	0.000
v_2^2	0.000	0.000	1.000	2.000	0.000	-1.000	0.000	-1.000	0.000	0.000
v_3^1	-1.000	-1.000	0.000	0.000	2.372	0.160	0.372	0.160	-0.797	-0.268
v_3^2	0.000	0.000	-1.000	-1.000	0.160	2.551	0.160	0.551	-0.223	-1.200
v_4^1	0.000	0.000	0.000	0.000	0.372	0.160	1.372	0.160	-0.797	-0.268
v_4^2	0.000	0.000	-1.000	-1.000	0.160	0.551	0.160	2.551	-0.223	-1.200
v_5^1	0.000	0.000	0.000	0.000	-0.797	-0.223	-0.797	-0.223	1.737	0.303
v_5^2	0.000	0.000	0.000	0.000	-0.268	-1.200	-0.268	-1.200	0.303	2.633

COVARIANCE OF MQTL EFFECTS GIVEN INCOMPLETE MARKER DATA

Algorithms to construct and invert the conditional gametic relationship matrix (\mathbf{A}), given complete marker data, are based on the recursive equation [3]. In deriving [3] from [2], it was assumed, given complete marker data, that events $Q_i^{k_i} \leftarrow Q_s^1$ and $Q_s^1 \equiv Q_j^{k_j}$, for example, are independent. They may not always be independent, however, when marker genotypes of the parents are unknown. Thus, although [2] holds for complete and incomplete marker data, [3] may not hold for incomplete marker data. Therefore, algorithms developed for complete marker data cannot be directly applied, in general, to pedigrees with incomplete marker data. In this section, we first demonstrate that [3] may not hold when marker genotypes of parents are unknown. A strategy to accommodate pedigrees with incomplete marker data is then presented.

The pedigree in table I is used to demonstrate that [3] may not hold when marker genotypes of the parents are unknown. In this pedigree, marker genotype of individual 2, the maternal parent of individuals 3 and 4, is unknown. Thus, as shown below, $\Pr(Q_4^1 \equiv Q_3^2)$ cannot be computed using [3].

From [2],

$$\Pr(Q_4^1 \equiv Q_3^2) = \Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) + \Pr(Q_4^1 \leftarrow Q_2^2, Q_2^2 \equiv Q_3^2 | \mathbf{G}_{obs}) + 0 + 0 \tag{26}$$

The last 2 terms in [26] are null because the QTL alleles in the unknown parent of individual 4 cannot be identical by descent to QTL alleles in individual 3. In deriving [3] from [2], it was assumed, given \mathbf{G}_{obs} , that $Q_4^1 \leftarrow Q_2^1$ and $Q_2^1 \equiv Q_3^2$, for example, are independent, *ie*

$$\Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) = \Pr(Q_4^1 \leftarrow Q_2^1 | \mathbf{G}_{obs}) \Pr(Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs})$$

Because the marker genotype for the maternal parent of individual 4 is unknown, however, the above equality does not hold. This is illustrated numerically.

Given the parents' genotypes, the genotypes of offspring are independent. Therefore, $\Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs})$ can be computed by conditioning on the genotype of individual 2 (parent of individuals 3 and 4) as

$$\begin{aligned} & \Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) \\ &= \sum_i^3 \sum_j^3 \Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2, G_2 = A_i A_j | \mathbf{G}_{obs}) \\ &= \sum_i^3 \sum_j^3 \Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | G_2 = A_i A_j, \mathbf{G}_{obs}) \Pr(G_2 = A_i A_j | \mathbf{G}_{obs}) \\ &= \sum_i^3 \sum_j^3 \Pr(Q_4^1 \leftarrow Q_2^1 | G_2 = A_i A_j, \mathbf{G}_{obs}) \Pr(Q_2^1 \equiv Q_3^2 | G_2 = A_i A_j, \mathbf{G}_{obs}) \\ &\quad \times \Pr(G_2 = A_i A_j | \mathbf{G}_{obs}) \\ &= 3/400 \end{aligned}$$

The probabilities required in the above computation are

$G_2 = A_i A_j$	$\Pr(Q_4^1 \leftarrow Q_2^1 G_2 = A_i A_j, \mathbf{G}_{obs})$	$\Pr(Q_2^1 \equiv Q_3^2 G_2 = A_i A_j, \mathbf{G}_{obs})$	$\Pr(G_2 = A_i A_j \mathbf{G}_{obs})$
$A_1 A_1$	0	0	0
$A_1 A_2$	9/80	1/10	1/3
$A_1 A_3$	0	0	0
$A_2 A_1$	1/80	9/10	1/3
$A_2 A_2$	0	1/2	1/6
$A_2 A_3$	0	9/10	1/12
$A_3 A_1$	0	0	0
$A_3 A_2$	0	1/10	1/12
$A_3 A_3$	0	0	0

From the above table, $\Pr(Q_4^1 \leftarrow Q_2^1 | \mathbf{G}_{obs})$ and $\Pr(Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs})$ can also be computed as

$$\begin{aligned} \Pr(Q_4^1 \leftarrow Q_2^1 | \mathbf{G}_{obs}) &= \sum_i^3 \sum_j^3 \Pr(Q_4^1 \leftarrow Q_2^1 | G_2 = A_i A_j, \mathbf{G}_{obs}) \Pr(G_2 = A_i A_j | \mathbf{G}_{obs}) \\ &= 1/24 \\ \Pr(Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) &= \sum_i^3 \sum_j^3 \Pr(Q_2^1 \equiv Q_3^2 | G_2 = A_i A_j, \mathbf{G}_{obs}) \Pr(G_2 = A_i A_j | \mathbf{G}_{obs}) \\ &= 1/2 \end{aligned}$$

The values of $\Pr(Q_4^1 \leftarrow Q_2^1 | \mathbf{G}_{obs}) = 1/24$, $\Pr(Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) = 1/2$, and $\Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) = 3/400$ illustrate that

$$\Pr(Q_4^1 \leftarrow Q_2^1, Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs}) \neq \Pr(Q_4^1 \leftarrow Q_2^1 | \mathbf{G}_{obs}) \Pr(Q_2^1 \equiv Q_3^2 | \mathbf{G}_{obs})$$

Because [3] may not hold when marker genotypes of parents s and d are unknown, the tabular algorithm for complete marker data cannot be applied directly to construct $\mathbf{\Lambda}$, given incomplete marker data. The tabular algorithm can be used, however, to construct $\mathbf{\Lambda}$ given incomplete marker data, as described below.

Let Ω be the set of all possible marker genotype configurations for individuals with unknown genotypes, and let \mathbf{G}_{obs} be the observed marker genotypes for individuals with known genotypes. The conditional gametic relationship matrix given incomplete marker data, $\mathbf{\Lambda}_{|\mathbf{G}_{obs}}$, can then be computed as

$$\mathbf{\Lambda}_{|\mathbf{G}_{obs}} = \sum_{\omega \in \Omega} \mathbf{\Lambda}_{|\omega, \mathbf{G}_{obs}} \Pr(\omega | \mathbf{G}_{obs}) \tag{27}$$

where $\mathbf{\Lambda}_{|\omega, \mathbf{G}_{obs}}$ is the conditional gametic relationship matrix given marker genotypes ω for individuals with unknown genotypes and \mathbf{G}_{obs} for individuals with known genotypes, and $\Pr(\omega | \mathbf{G}_{obs})$ is the conditional probability of individuals with unknown genotypes having marker genotypes ω , given marker genotypes \mathbf{G}_{obs} for individuals with known genotypes. The matrix $\mathbf{\Lambda}_{|\omega, \mathbf{G}_{obs}}$ can be constructed using the tabular method given complete marker data, and the probability $\Pr(\omega | \mathbf{G}_{obs})$ can be computed as

$$\Pr(\omega | \mathbf{G}_{obs}) = \frac{\Pr(\omega, \mathbf{G}_{obs})}{\sum_{\omega} \Pr(\omega, \mathbf{G}_{obs})}$$

where $\Pr(\omega, \mathbf{G}_{obs})$ can be computed efficiently (Elston and Stewart, 1971; Bonney, 1984).

The conditional gametic relationship matrix ($\mathbf{\Lambda}$) for the pedigree in table I, computed using [27], is

	v_1^1	v_1^2	v_2^1	v_2^2	v_3^1	v_3^2	v_4^1	v_4^2	v_5^1	v_5^2
v_1^1	1.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_1^2	0.000	1.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_2^1	0.000	0.000	1.000	0.000	0.000	0.500	0.042	0.458	0.067	0.433
v_2^2	0.000	0.000	0.000	1.000	0.000	0.500	0.042	0.458	0.067	0.433
v_3^1	0.500	0.500	0.000	0.000	1.000	0.000	0.000	0.000	0.450	0.050
v_3^2	0.000	0.000	0.500	0.500	0.000	1.000	0.015	0.698	0.092	0.765
v_4^1	0.000	0.000	0.042	0.042	0.000	0.015	1.000	0.000	0.451	0.057
v_4^2	0.000	0.000	0.458	0.458	0.000	0.698	0.000	1.000	0.085	0.764
v_5^1	0.225	0.225	0.067	0.067	0.450	0.092	0.451	0.085	1.000	0.069
v_5^2	0.025	0.025	0.433	0.433	0.050	0.765	0.057	0.764	0.069	1.000

Computing $\mathbf{\Lambda}$ using [27] is not efficient when a large number of individuals have unknown genotypes because the summation in [27] is over all combinations of the

unknown genotypes. Further, an efficient algorithm to invert $\Lambda_{|\mathbf{G}_{obs}}$ has not been found. Therefore, 2 approximate methods to compute $\Lambda_{|\mathbf{G}_{obs}}$ and its inverse are presented:

1) We have already shown that [3] may not hold for incomplete marker data because, given $\mathbf{G}_{obs}, Q_i^{k_i} \Leftarrow Q_s^1$ and $Q_s^1 \equiv Q_j^{k_j}$ in [2], for example, may not be independent. If we ignore this dependency, then [15] and [18], which are based on [3], can be used to approximate Λ and its inverse. This approximation will require PDMs for individuals with incomplete marker data. For individual i , with unknown marker genotypes for parents s and d , PDMs can be computed as

$$\Pr(M_i^1 \Leftarrow M_s^1 | \mathbf{G}_{obs}) = \sum_{G_s} \sum_{G_d} \sum_{G_i} \Pr(M_i^1 \Leftarrow M_s^1 | G_s, G_d, G_i) \Pr(G_s, G_d, G_i | \mathbf{G}_{obs}) \quad [28]$$

where each summation is over all possible genotypes at the ML. If G_s , G_d , or G_i is not missing, then the corresponding summation should be dropped from [28]. The computation of $\Pr(G_s, G_d, G_i | \mathbf{G}_{obs})$ can be very time-consuming when a large number of individuals have unknown marker genotypes. An approximation for $\Pr(G_s, G_d, G_i | \mathbf{G}_{obs})$ can be obtained, however, by conditioning only on marker information of 'close' relatives of i , s and d , where, for example, a set of 'close' relatives for an individual could be its parents, sibs and offspring. The conditional gametic relationship matrix (Λ), for the pedigree in table I, using this approximation is

	v_1^1	v_1^2	v_2^1	v_2^2	v_3^1	v_3^2	v_4^1	v_4^2	v_5^1	v_5^2
v_1^1	1.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_1^2	0.000	1.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_2^1	0.000	0.000	1.000	0.000	0.000	0.500	0.042	0.458	0.067	0.433
v_2^2	0.000	0.000	0.000	1.000	0.000	0.500	0.042	0.458	0.067	0.433
v_3^1	0.500	0.500	0.000	0.000	1.000	0.000	0.000	0.000	0.450	0.050
v_3^2	0.000	0.000	0.500	0.500	0.000	1.000	0.042	0.458	0.092	0.658
v_4^1	0.000	0.000	0.042	0.042	0.000	0.042	1.000	0.000	0.452	0.069
v_4^2	0.000	0.000	0.458	0.458	0.000	0.458	0.000	1.000	0.073	0.656
v_5^1	0.225	0.225	0.067	0.067	0.450	0.092	0.452	0.073	1.000	0.058
v_5^2	0.025	0.025	0.433	0.433	0.050	0.658	0.069	0.656	0.058	1.000

The consequence of this approximation is that the summation in [27] has been brought into inside of Λ and performed on \mathbf{B}_i (or \mathbf{S}_i , see [5]).

2) Let ω_{max} be the genotype configuration in Ω with the largest probability. Given ω_{max} and \mathbf{G}_{obs} , [15] and [18] can be used to approximate Λ and its inverse. Sheehan *et al* (1993) proposed a sampling scheme to compute the probability of genotype configurations. For the pedigree in table I, given \mathbf{G}_{obs} , $G_2 = A_1A_2$ has the largest conditional probability (2/3) among all possible genotypes for G_2 , *ie* $\omega_{max} = (G_2 = A_1A_2)$. Thus, [15] can be used to construct Λ with $G_2 = A_1A_2$. The conditional gametic relationship matrix (Λ) using this approximation is:

	v_1^1	v_1^2	v_2^1	v_2^2	v_3^1	v_3^2	v_4^1	v_4^2	v_5^1	v_5^2
v_1^1	1.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_1^2	0.000	1.000	0.000	0.000	0.500	0.000	0.000	0.000	0.225	0.025
v_2^1	0.000	0.000	1.000	0.000	0.000	0.100	0.113	0.088	0.060	0.090
v_2^2	0.000	0.000	0.000	1.000	0.000	0.900	0.013	0.788	0.090	0.760
v_3^1	0.500	0.500	0.000	0.000	1.000	0.000	0.000	0.000	0.450	0.050
v_3^2	0.000	0.000	0.100	0.900	0.000	1.000	0.023	0.718	0.096	0.774
v_4^1	0.000	0.000	0.113	0.013	0.000	0.023	1.000	0.000	0.451	0.060
v_4^2	0.000	0.000	0.088	0.788	0.000	0.718	0.000	1.000	0.086	0.773
v_5^1	0.225	0.225	0.060	0.090	0.450	0.096	0.451	0.086	1.000	0.074
v_5^2	0.025	0.025	0.090	0.760	0.050	0.774	0.060	0.773	0.074	1.000

The consequence of this approximation is that the resulting $\mathbf{\Lambda}$ is conditional on ω_{\max} .

A measure of how well an approximation compares to the exact method is the correlations coefficient, $r_{\text{exact,approx}}$, between upper off-diagonal elements of $\mathbf{\Lambda}_{|\mathbf{G}_{obs}}$, computed exactly by [27], and corresponding elements computed by approximate methods. For the pedigree in table I, $r_{\text{exact,approx1}} = 0.9877$ for approximation 1 and $r_{\text{exact,approx2}} = 0.8735$ for approximation 2.

To further examine these approximations $r_{\text{exact,approx1}}$ and $r_{\text{exact,approx2}}$ were computed for a pedigree of 99 individuals with 3 generations. The first generation consisted of 3 grandsires, each mated with 12 granddams. The second generation consisted of 2 sires and 10 dams from each grandsire for a total of 6 sires and 30 dams. Each sire was randomly mated with 4 dams, avoiding full-sib and half-sib matings. The third generation consisted of 2 grandsons and 2 granddaughters from each sire for a total of 12 grandsons and 12 granddaughters. Marker genotypes were assumed missing for the 30 maternal granddams. Thus covariances were only computed for the remaining 69 individuals in the pedigree. Marker genotypes for these 69 individuals were generated randomly. Granddaughters and dams without progeny were assigned missing marker genotypes with probability 0.6.

Exact and approximate covariances were computed for 20 randomly generated marker genotype configurations. The average for $r_{\text{exact,approx1}}$ was 0.8923 and for $r_{\text{exact,approx2}}$ was 0.8939. The effect of these approximations on marker-assisted genetic evaluation needs to be studied.

DISCUSSION

Theory and algorithms are presented here to construct the conditional covariance matrix between relatives for a marked quantitative trait locus ($\mathbf{G}_v = \mathbf{\Lambda}\sigma_v^2$) and to obtain its inverse efficiently. These algorithms extend those of Fernando and Grossman (1989) to accommodate situations (1) where paternal or maternal origin of marker alleles cannot be determined and (2) where marker genotypes of some individuals in the pedigree are unknown. The exact procedure presented here to

construct $\Lambda_{|\mathbf{G}_{obs}}$ for incomplete marker data may not be efficient for large pedigree. Therefore, we presented 2 alternative strategies to approximate $\Lambda_{|\mathbf{G}_{obs}}$ and its inverse. Simulation results indicate that the 2 approximations are similar because they have similar correlations with the exact method ($r_{\text{exact,approx1}} = 0.8923$, $r_{\text{exact,approx2}} = 0.8939$). Approximation (1) is preferred, however, because it may be difficult to search for ω_{max} when a large number of individuals have unknown marker genotypes.

We also presented an algorithm to compute the conditional inbreeding coefficient (f_i) for a QTL given \mathbf{G}_{obs} , which is different from Wright's inbreeding coefficient. This conditional inbreeding coefficient is the probability that the 2 homologous alleles at the MQTL in an individual are identical by descent given the pedigree and marker information, whereas Wright's inbreeding coefficient is the conditional probability that the 2 homologous alleles at any locus in an individual are identical by descent given only the pedigree. A numerical example is used to show that equation [3], which is the basis of tabular method to construct \mathbf{G}_v , does not hold generally when marker data are incomplete.

In most practical situations, marker information will not be available on distant ancestors. Thus, TM-BLUP cannot be computed. One of the 2 approximations presented in this paper, however, can be employed to compute $\Lambda_{|\mathbf{G}_{obs}}$. Thus available marker information can be used to obtain improved genetic evaluations by approximate TM-BLUP. Further, in general, information on distant ancestors has little impact on genetic evaluations.

If the ML and MQTL are in linkage disequilibrium, marker data provide information on the first moment of MQTL effects. In this situation, regression techniques can be used for genetic evaluation using marker and trait information (Lande and Thompson, 1990; Zhang and Smith, 1992). If the ML and MQTL are in linkage equilibrium, marker data do not provide information on the first moment of the MQTL effects. Even with equilibrium, however, marker data do provide information on covariances of MQTL effects. In this situation, TM-BLUP can be used for genetic evaluation by fitting MQTL effects as random effects within animal (Fernando and Grossman, 1989; Cantet and Smith, 1991; Goddard, 1992; Hoeschele, 1993).

Genetic evaluation by TM-BLUP requires knowledge of genetic parameters, such as r and σ_v^2 . This is also true for T-BLUP, which requires knowledge of genetic variances and covariances. In practice, true values of genetic parameters are unknown and estimates are used in their places. Both restricted maximum likelihood and maximum likelihood approaches can be used to estimate parameters required for TM-BLUP (Weller and Fernando, 1991).

Ideally, marker-assisted selection will be based on multiple marker loci. When the linkage phase between flanking marker loci is known in addition to the parental origin of marker alleles, the method presented by Goddard (1992) for multiple markers can be used for TM-BLUP. Further research is needed for TM-BLUP using multiple markers when both the linkage phase between flanking marker loci and the parental origin of marker alleles are unknown.

ACKNOWLEDGMENT

The authors would like to thank an anonymous reviewer for pointing out an error in the manuscript.

REFERENCES

- Bonney GE (1984) On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am J Med Genet* 18, 731-749
- Cantet RJC, Smith C (1991) Reduced animal model for marker-assisted selection using best linear unbiased prediction. *Genet Sel Evol* 23, 221-233
- Chevalet C, Gillois M, Khang JVT (1984) Conditional probabilities of identity of genes at a locus linked to a marker. *Genet Sel Evol* 16, 431-444
- Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21, 523-542
- Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. *J Hered* 40, 51-55
- Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. *Genet Sel Evol* 21, 467-477
- Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor Appl Genet* 46, 319-330
- Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. *Theor Appl Genet* 83, 878-886
- Henderson CR (1973) Sire evaluation and genetic trend. In: *Anim Breed Genet Symp in Honor of Dr J L Lush*, Am Soc Anim Sci and Am Dairy Sci Assoc, Champaign, IL, USA, 10-41
- Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69-83
- Hoeschele I (1993) Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. *J Dairy Sci* 76, 1693-1713
- Kashi Y, Hallerman EM, Soller M (1990) Marker-assisted selection of candidates sires for progeny testing programs. *Anim Prod* 51, 63-74
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Searle SR (1982) *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York, USA
- Sheehan N, Thomas A (1993) On the irreducibility of a Markov chain defined on a space of genotype configurations by a sample scheme. *Biometrics* 49, 163-175
- Smith C, Simpson SP (1986) The use of genetic polymorphisms in livestock improvement. *J Anim Breed Genet* 103, 205-217
- Smith SP, Allaire FR (1985) Efficient selection rules to increase non-linear merit: application in mate selection. *Genet Sel Evol* 17, 387-406
- Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. *Anim Prod* 27, 133-139
- Soller M, Beckmann JS (1982) Restriction fragment length polymorphisms and genetic improvement. In: *2nd World Congress Genet Appl Livest Prod*, Madrid, Editorial Garsi, Madrid, Spain, vol 6, 396-404
- van Arendonk JAM, Tier B, Kinghorn BP (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319-329
- Wang T, van der Beek S, Fernando RL, Grossman M (1991) Covariance between effects of marked QTL alleles. *J Anim Sci* 69 (Suppl 1), 202 (Abstr)

Weller JI, Fernando RL (1991) Strategies for the improvement of animal production using marker-assisted selection. *In: Gene Mapping: Strategies, Techniques and Applications* (LB Schook, HA Lewin, DG McLaren, eds), Marcel Dekker, New York, USA, 305-328
 Zhang W, Smith C (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor Appl Genet* 83, 813-820

APPENDIX A

Theory for computation of PDMs

Let $G_s = M_s^1 M_s^2$, $G_d = M_d^1 M_d^2$ and $G_i = M_i^1 M_i^2$ be the marker genotypes of 2 parents s and d and their offspring i . Given G_s , G_d and G_i , the probability that $M_i^{k_i}$ descended from $M_p^{k_p}$ does not depend on other information in the pedigree. Thus, $\Pr(M_i^{k_i} \leftarrow M_p^{k_p} | \mathbf{G}_{obs}) = \Pr(M_i^{k_i} \leftarrow M_p^{k_p} | G_s, G_d, G_i)$, which can be obtained as

$$\Pr(M_i^{k_i} \leftarrow M_p^{k_p} | G_s, G_d, G_i) = \frac{\Pr(M_i^{k_i} \leftarrow M_p^{k_p}, G_i | G_s, G_d)}{\Pr(G_i | G_s, G_d)} \quad [A1]$$

The numerator and denominator of [A1] are easily computed from Mendelian principles. For example, if 2 parents and their offspring each has marker genotype $A_1 A_2$, ie $G_s = M_s^1 M_s^2 = A_1 A_2$, $G_d = M_d^1 M_d^2 = A_1 A_2$ and $G_i = M_i^1 M_i^2 = A_1 A_2$, then

$$\Pr(G_i | G_s, G_d) = 1/2, \quad \text{and} \quad \Pr(M_i^1 \leftarrow M_s^1, G_i | G_s, G_d) = 1/4$$

Thus, $\Pr(M_i^1 \leftarrow M_s^1 | G_s, G_d, G_i) = 1/2$.

Other examples are listed below. Eight PDMs for each individual i are collected into matrix \mathbf{S}_i , which is defined in [7].

s	d	i	$S_i(1,1)$	$S_i(1,2)$	$S_i(1,3)$	$S_i(1,4)$	$S_i(2,1)$	$S_i(2,2)$	$S_i(2,3)$	$S_i(2,4)$
$A_1 A_1$	$A_1 A_1$	$A_1 A_1$	1/4	1/4	1/4	1/4	1/4	1/4	1/4	1/4
$A_1 A_1$	$A_1 A_2$	$A_1 A_2$	1/2	1/2	0	0	0	0	0	1
$A_1 A_2$	$A_1 A_2$	$A_1 A_2$	1/2	0	1/2	0	0	1/2	0	1/2
$A_2 A_1$	$A_1 A_2$	$A_1 A_2$	0	1/2	1/2	0	1/2	0	0	1/2

APPENDIX B

Theory for computation of PDQs

The conditional probability that allele $Q_i^{k_i}$ of individual i descended from allele $Q_p^{k_p}$ of parent p (fig 1), given \mathbf{G}_{obs} , will be denoted by $\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs})$, which is called PDQ. This conditional probability can be expressed as

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs}) = \sum_{p' \in (s,d)} \sum_{k_{p'}=1}^2 \Pr(M_i^{k_i} \leftarrow M_{p'}^{k_{p'}}, Q_i^{k_i} \leftarrow Q_{p'}^{k_{p'}} | \mathbf{G}_{obs}) \quad [B1]$$

Because $Q_i^{k_i}$ and $M_i^{k_i}$ are on the same chromosome of individual i , each must have descended from the same parent. Thus, $\Pr(M_i^{k_i} \leftarrow M_{p'}^{k_p'}, Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs})$ is null. Now,

$$\begin{aligned} \Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs}) &= \sum_{k'_p=1}^2 \Pr(M_i^{k_i} \leftarrow M_{p'}^{k'_p}, Q_i^{k_i} \leftarrow Q_p^{k_p} | \mathbf{G}_{obs}) \\ &= \sum_{k'_p=1}^2 \Pr(M_i^{k_i} \leftarrow M_{p'}^{k'_p} | \mathbf{G}_{obs}) \Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p}, \mathbf{G}_{obs}) \quad [\text{B2}] \end{aligned}$$

There are 2 probabilities on the right-hand side in [B2]. The first probability

$$\Pr(M_i^{k_i} \leftarrow M_{p'}^{k'_p} | \mathbf{G}_{obs})$$

is a PDM for individual i (see [A1] for its computation). The second probability

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p}, \mathbf{G}_{obs})$$

can be expressed in terms of PDMs and of the recombination rate r between the ML and the MQTL as explained below.

Given $M_i^{k_i} \leftarrow M_{p'}^{k'_p}$, the probability that $Q_i^{k_i}$ descended from $Q_p^{k_p}$ does not depend on other information in the pedigree. Thus,

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p}, \mathbf{G}_{obs}) = \Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p}) \quad [\text{B3}]$$

If $k'_p = k_p$, then recombination has not taken place, so that

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p=k_p}) = 1 - r$$

If $k'_p \neq k_p$, then recombination has taken place, so that

$$\Pr(Q_i^{k_i} \leftarrow Q_p^{k_p} | M_i^{k_i} \leftarrow M_{p'}^{k'_p \neq k_p}) = r$$

For each combination of k_i , k_p , $k'_p = 1, 2$, we have

$$\begin{aligned} \Pr(Q_i^1 \leftarrow Q_s^1 | M_i^1 \leftarrow M_s^1) &= 1 - r, & \Pr(Q_i^2 \leftarrow Q_s^1 | M_i^2 \leftarrow M_s^1) &= 1 - r \\ \Pr(Q_i^1 \leftarrow Q_s^2 | M_i^1 \leftarrow M_s^1) &= r, & \Pr(Q_i^2 \leftarrow Q_s^2 | M_i^2 \leftarrow M_s^1) &= r \\ \Pr(Q_i^1 \leftarrow Q_s^1 | M_i^1 \leftarrow M_s^2) &= r, & \Pr(Q_i^2 \leftarrow Q_s^1 | M_i^2 \leftarrow M_s^2) &= r \\ \Pr(Q_i^1 \leftarrow Q_s^2 | M_i^1 \leftarrow M_s^2) &= 1 - r, & \Pr(Q_i^2 \leftarrow Q_s^2 | M_i^2 \leftarrow M_s^2) &= 1 - r \end{aligned}$$

The PDQs, $\Pr(Q_i^{k_i} \Leftarrow Q_p^{k_p} | \mathbf{G}_{obs})$, for $k_i, k_p = 1, 2$, can be obtained by using the above in [B2]:

$$\begin{aligned} \Pr(Q_i^1 \Leftarrow Q_p^1 | \mathbf{G}_{obs}) &= (1-r) \Pr(M_i^1 \Leftarrow M_p^1 | \mathbf{G}_{obs}) + r \Pr(M_i^1 \Leftarrow M_p^2 | \mathbf{G}_{obs}) \\ \Pr(Q_i^1 \Leftarrow Q_p^2 | \mathbf{G}_{obs}) &= r \Pr(M_i^1 \Leftarrow M_p^1 | \mathbf{G}_{obs}) + (1-r) \Pr(M_i^1 \Leftarrow M_p^2 | \mathbf{G}_{obs}) \\ \Pr(Q_i^2 \Leftarrow Q_p^1 | \mathbf{G}_{obs}) &= (1-r) \Pr(M_i^2 \Leftarrow M_p^1 | \mathbf{G}_{obs}) + r \Pr(M_i^2 \Leftarrow M_p^2 | \mathbf{G}_{obs}) \\ \Pr(Q_i^2 \Leftarrow Q_p^2 | \mathbf{G}_{obs}) &= r \Pr(M_i^2 \Leftarrow M_p^1 | \mathbf{G}_{obs}) + (1-r) \Pr(M_i^2 \Leftarrow M_p^2 | \mathbf{G}_{obs}) \end{aligned}$$

where $p = s$ or d .

In summary,

$$\Pr(Q_i^{k_i} \Leftarrow Q_p^{k_p} | \mathbf{G}_{obs}) = (1-\rho) \Pr(M_i^{k_i} \Leftarrow M_p^1 | \mathbf{G}_{obs}) + \rho \Pr(M_i^{k_i} \Leftarrow M_p^2 | \mathbf{G}_{obs}) \quad [\text{B4}]$$

for $k_i = 1$ or 2 and $p = s$ or d , where $\rho = r$ when $k_p = 1$ and $\rho = 1-r$ when $k_p = 2$. Note that PDQs are now expressed in terms of PDMs and r .

APPENDIX C

Theory for computation of $\Pr(T_{k_s k_d} | \mathbf{G}_{obs})$

The event that the pair of alleles (Q_i^1, Q_i^2) in individual i descended from parental pair ($Q_s^{k_s}, Q_d^{k_d}$) (fig 1), is denoted by $T_{k_s k_d}$ for $k_s, k_d = 1$ or 2 . This event can occur in 1 of 2 ways:

1. Q_i^1 descended from $Q_s^{k_s}$ and Q_i^2 from $Q_d^{k_d}$, denoted ($Q_i^1 \Leftarrow Q_s^{k_s}, Q_i^2 \Leftarrow Q_d^{k_d}$)
2. Q_i^1 descended from $Q_d^{k_d}$ and Q_i^2 from $Q_s^{k_s}$, denoted ($Q_i^1 \Leftarrow Q_d^{k_d}, Q_i^2 \Leftarrow Q_s^{k_s}$)

Given the pedigree and marker genotypes, the probability of $T_{k_s k_d}$, which is denoted by $\Pr(T_{k_s k_d} | \mathbf{G}_{obs})$, can be written as

$$\begin{aligned} \Pr(T_{k_s k_d} | \mathbf{G}_{obs}) &= \Pr(Q_i^1 \Leftarrow Q_s^{k_s}, Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs}) \\ &\quad + \Pr(Q_i^1 \Leftarrow Q_d^{k_d}, Q_i^2 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs}) \quad [\text{C1}] \end{aligned}$$

Consider the first probability on the right-hand side in [C1], which can be expressed as

$$\begin{aligned} \Pr(Q_i^1 \Leftarrow Q_s^{k_s}, Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs}) \\ = \Pr(Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs}) \Pr(Q_i^1 \Leftarrow Q_s^{k_s} | Q_i^2 \Leftarrow Q_d^{k_d}, \mathbf{G}_{obs}) \quad [\text{C2}] \end{aligned}$$

where $\Pr(Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs})$ is a PDQ and $\Pr(Q_i^1 \Leftarrow Q_s^{k_s} | Q_i^2 \Leftarrow Q_d^{k_d}, \mathbf{G}_{obs})$ can be expressed in terms of PDQs for individual i , as explained below.

Note that if Q_i^2 descended from $Q_d^{k_d}$ of parent d , Q_i^1 must have descended from the other parent s ; ie $Q_i^2 \Leftarrow Q_d^{k_d}$ is equivalent to $Q_i^1 \Leftarrow s$. Therefore,

$$\begin{aligned} \Pr(Q_i^1 \Leftarrow Q_s^{k_s} | Q_i^2 \Leftarrow Q_d^{k_d}, \mathbf{G}_{obs}) &= \Pr(Q_i^1 \Leftarrow Q_s^{k_s} | Q_i^1 \Leftarrow s, \mathbf{G}_{obs}) \\ &= \frac{\Pr(Q_i^1 \Leftarrow Q_s^{k_s}, Q_i^1 \Leftarrow s | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow s | \mathbf{G}_{obs})} \quad [\text{C3}] \end{aligned}$$

Observe that event $Q_i^1 \Leftarrow s$ is implied by $Q_i^1 \Leftarrow Q_s^{k_s}$; therefore,

$$\Pr(Q_i^1 \Leftarrow Q_s^{k_s}, Q_i^1 \Leftarrow s | \mathbf{G}_{obs}) = \Pr(Q_i^1 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs})$$

Further,

$$\Pr(Q_i^1 \Leftarrow s | \mathbf{G}_{obs}) = \Pr(Q_i^1 \Leftarrow Q_s^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_s^2 | \mathbf{G}_{obs})$$

Thus, [C3] can be rewritten in terms of PDQs as

$$\Pr(Q_i^1 \Leftarrow Q_s^{k_s} | Q_i^2 \Leftarrow Q_d^{k_d}, \mathbf{G}_{obs}) = \frac{\Pr(Q_i^1 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow Q_s^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_s^2 | \mathbf{G}_{obs})} \quad [C4]$$

After substituting [C4] in [C2], the first probability on the right-hand side in [C1] can be written in terms of PDQs. The same approach is applied to the second probability in [C1]. Then, $\Pr(T_{k_s k_d} | \mathbf{G}_{obs})$ can be expressed in terms of PDQs as

$$\begin{aligned} \Pr(T_{k_s k_d} | \mathbf{G}_{obs}) &= \frac{\Pr(Q_i^1 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs}) \Pr(Q_i^2 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow Q_s^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_s^2 | \mathbf{G}_{obs})} \\ &\quad + \frac{\Pr(Q_i^1 \Leftarrow Q_d^{k_d} | \mathbf{G}_{obs}) \Pr(Q_i^2 \Leftarrow Q_s^{k_s} | \mathbf{G}_{obs})}{\Pr(Q_i^1 \Leftarrow Q_d^1 | \mathbf{G}_{obs}) + \Pr(Q_i^1 \Leftarrow Q_d^2 | \mathbf{G}_{obs})} \end{aligned} \quad [C5]$$

If 1 of the denominators in [C5] is zero (indicating the event in 1 of the terms in [C1] is impossible), then the corresponding term in [C5] is set to zero.