Theor Appl Genet (1998) 97 : 473–478

R. Bernardo

A model for marker-assisted selection among single crosses with multiple genetic markers

Received: 22 June 1997 / Accepted: 25 February 1998

Abstract Trait means of marker genotypes are often inconsistent across experiments, thereby hindering the use of regression techniques in marker-assisted selection. Best linear unbiased prediction based on trait and marker data (TM-BLUP) does not require prior information on the mean effects associated with specific marker genotypes and, consequently, may be useful in applied breeding programs. The objective of this paper is to present a flanking-marker, TM-BLUP model that is applicable to interpopulation single crosses that characterize maize (*Zea mays* L.) breeding programs. The performance of a single cross is modeled as the sum of testcross additive and dominance effects at unmarked quantitative trait loci (QTL) and at marked QTL (MQTL). The TM-BLUP model requires information on the recombination frequencies between flanking markers and the MQTL and on MQTL variances. A tabular method is presented for calculating the conditional probability that MQTL alleles in two inbreds are identical by descent given the observed marker genotypes (G^k_{obs}) at the kth MQTL. Information on identity by descent of MQTL alleles can then be used to calculate the conditional covariance of MQTL effects between single crosses given G^k_{obs}. The inverse of the covariance matrix for dominance effects at unmarked QTL and MQTL can be written directly from the inverse of the covariance matrices of the corresponding testcross additive effects. In practice, the computations required in TM-BLUP may be prohibitive. The computational requirements may be reduced with simplified TM-BLUP models wherein dominance effects at MQTL are excluded, only the single crosses that have been tested are included, or information is pooled across several MQTL.

Key words Best linear unbiased prediction · Marker-assisted selection · Quantitative trait locus · Single cross

Communicated by H. C. Becker

R. Bernardo (✉)
Department of Agronomy, Purdue University,
1150 Lilly Hall of Life Sciences,
West Lafayette, IN 47907-1150, USA

Introduction

Molecular genetic markers enable the mapping of genes controlling quantitative traits and subsequent marker-assisted selection (Soller and Beckmann 1983; Lande and Thompson 1990). If flanking marker loci and a quantitative trait locus (QTL) are in linkage disequilibrium, the recombination frequencies between the flanking markers and the QTL as well as the mean effects of QTL alleles can be estimated (Haley and Knott 1992; Martinez and Curnow 1992; Zeng 1994). In maize (*Zea mays* L.), many studies have been conducted to map QTL for grain yield (Edwards et al. 1987; Stuber et al. 1992; Zehr et al. 1992; Beavis et al. 1994; Veldboom and Lee 1994; Ajmone-Marsan et al. 1995; Austin and Lee 1996; Eathington et al. 1997), disease resistance (Bubeck et al. 1993; Freymark et al. 1993), insect tolerance (Lee 1993; Schön et al. 1993), kernel chemical composition (Goldman et al. 1993; Schön et al. 1994), and morphological traits (Beavis et al. 1991; Koester et al. 1993; Schön et al. 1994; Veldboom et al. 1994).

Despite the detection of QTL for such traits, marker-assisted selection based on regression of trait means on marker genotypes has not been widely used in maize breeding programs (Smith and Beavis 1996). Genotype x environment interaction causes estimated QTL effects to vary across environments (Beavis and Keim 1995). The QTL allele linked to a given marker allele may vary among maize inbreds, thereby limiting the

estimates of marker-associated effects to the mapping population studied. The resulting inconsistency in means of marker genotypes among mapping populations and environments (Dudley 1993) causes difficulty in improving traits by selecting for desirable marker alleles in any given breeding population.

Best linear unbiased prediction based on trait and marker information (TM-BLUP) may be an alternative procedure that would make marker-assisted selection feasible in maize breeding programs. Neither linkage disequilibrium between marker loci and QTL nor information on the mean effect associated with a particular marker allele is needed in TM-BLUP (Wang et al. 1995). Rather, TM-BLUP requires information on recombination frequencies and QTL variances. Although estimates of marker-associated effects may be specific to mapping populations and target environments, the recombination frequencies between flanking markers and a QTL are assumed consistent across populations and environments. These recombination frequencies can be estimated from a large mapping population evaluated in a large number of environments (Beavis 1994). The QTL variances, as well as trait variances, can be estimated ad hoc from data sets routinely generated in breeding programs. The ability of TM-BLUP to accommodate general and unbalanced data structures makes it a potentially useful procedure in applied breeding programs (Goddard 1992).

Intrapopulation TM-BLUP has been proposed for single marker (Fernando and Grossman 1989), multiple non-flanking marker (van Arendonk et al. 1994), and multiple flanking marker (Goddard 1992) additive genetic models. However, a multiple flanking-marker, interpopulation TM-BLUP model that is applicable to hybrid species such as maize has not been developed. My objectives are to: (1) describe a method of calculating the covariance between single crosses for a QTL flanked by markers; (2) propose an interpopulation TM-BLUP model for markers that flank a QTL; (3) present a simple method for calculating the inverse of the dominance covariance matrices in TM-BLUP; and (4) discuss the applicability of TM-BLUP in hybrid breeding programs.

Theory

Single crosses are made between homozygous lines from two unrelated heterotic groups. Assume i and j are any two inbreds from Group I whereas i' and j' are any two inbreds from Group II. Linkage equilibrium among QTL is assumed in the Group I and Group II base populations. The genetic model includes testcross additive and dominance effects, but epistasis is assumed absent.

Covariance between single crosses at an unmarked QTL

Schnell (1965) and Stuber and Cockerham (1966) derived the covariance between single crosses at a single locus:

$$\mathrm{Cov}[(i \times i'), (j \times j')] = f_{ij} V^l_{A(1)} + f_{i'j'} V^l_{A(2)} + f_{ij} f_{i'j'} V^l_D$$

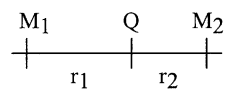
where: f_{ij} = coefficient of coancestry between i and j; $f_{i'j'}$ = coefficient of coancestry between i' and j'; $V^l_{A(1)}$ = additive variance of alleles from Group I inbreds, testcrossed to Group II inbreds, at the lth unmarked QTL; $V^l_{A(2)}$ = additive variance of alleles from Group II inbreds, testcrossed to Group I inbreds, at the lth unmarked QTL; and V^l_D = dominance variance of paired Group I and Group II alleles at the lth unmarked QTL. The covariance between $i \times i'$ and $j \times j'$ across all unmarked QTL is:

$$\mathrm{Cov}[(i \times i'), (j \times j')] = f_{ij} V_{A(1)} + f_{i'j'} V_{A(2)} + f_{ij} f_{i'j'} V_D$$

where: $V_{A(1)} = \sum V^l_{A(1)}$; $V_{A(2)} = \sum V^l_{A(2)}$; and $V_D = \sum V^l_D$.

Covariance between single crosses at a QTL flanked by markers

Assume that two marker loci (M_1 and M_2) flank a QTL (Q):



The recombination frequencies (Haldane 1919) are r_1 between M_1 and Q and r_2 between Q and M_2. The alleles that are homozygous in a given inbred are denoted in superscript; i.e., i has the alleles M^i_1, Q^i, and M^i_2.

The conditional covariance between single crosses at the kth marked QTL (MQTL), given the observed marker genotypes flanking the kth MQTL (G^k_{obs}), is:

$$\begin{aligned}\mathrm{Cov}^k[(i \times i'), (j \times j')] = {} & \Pr(Q^i \equiv Q^j | G^k_{obs}) V^k_{MA(1)} \\ & + \Pr(Q^{i'} \equiv Q^{j'} | G^k_{obs}) V^k_{MA(2)} \\ & + \Pr(Q^i \equiv Q^j | G^k_{obs}) \Pr(Q^{i'} \equiv Q^{j'} | G^k_{obs}) V^k_{MD}\end{aligned}$$

where: $\Pr(Q^i \equiv Q^j | G^k_{obs})$ = conditional probability that Q^i is identical by descent (denoted by $\equiv$) to Q^j given G^k_{obs}; $\Pr(Q^{i'} \equiv Q^{j'} | G^k_{obs})$ = conditional probability that $Q^{i'} \equiv Q^{j'}$ given G^k_{obs}; $V^k_{MA(1)}$ = testcross additive variance of alleles from Group I at the kth MQTL; $V^k_{MA(2)}$ = testcross additive variance of alleles from Group II at the kth MQTL; and V^k_{MD} = dominance variance of paired Group I and Group II alleles at the kth MQTL. The $\Pr(Q^i \equiv Q^j | G^k_{obs})$ and $\Pr(Q^{i'} \equiv Q^{j'} | G^k_{obs})$ terms for MQTL are analogous to the f_{ij} and $f_{i'j'}$ quantities for unmarked QTL. The calculation of $\Pr(Q^i \equiv Q^j | G^k_{obs})$ and $\Pr(Q^{i'} \equiv Q^{j'} | G^k_{obs})$ is outlined below.

Probability of descent of a QTL allele with flanking markers

Suppose a and b are the parental inbreds of i, and j is not a direct descendant of i. The $\Pr(Q^i \equiv Q^j | G^k_{obs})$ term can be expressed in terms of the conditional probability that $Q^j \equiv Q^a$ and $Q^j \equiv Q^b$ given G^k_{obs}:

$$\begin{aligned}\Pr(Q^i \equiv Q^j | G^k_{obs}) = {} & \Pr(Q^i \Leftarrow Q^a | G^k_{obs}) \Pr(Q^a \equiv Q^j | G^k_{obs}) \\ & + \Pr(Q^i \Leftarrow Q^b | G^k_{obs}) \Pr(Q^b \equiv Q^j | G^k_{obs})\end{aligned}$$

where $\Leftarrow$ indicates that the Q^i allele descended from Q^a or Q^b.

There are four possible ways by which $Q^i \Leftarrow Q^a$:

	M_1	M_2	Between M_1 and Q	Between Q and M_2
(1)	$M_1^i \Leftarrow M_1^a$	$M_2^i \Leftarrow M_2^a$	No recombination	No recombination
(2)	$M_1^i \Leftarrow M_1^a$	$M_2^i \Leftarrow M_2^b$	No recombination	Recombination
(3)	$M_1^i \Leftarrow M_1^b$	$M_2^i \Leftarrow M_2^a$	Recombination	No recombination
(4)	$M_1^i \Leftarrow M_1^b$	$M_2^i \Leftarrow M_2^b$	Recombination	Recombination

Assuming no interference, the frequency of non-recombinant marker genotypes among recombinant inbred lines [i.e., ($M_1^i \Leftarrow M_1^a$, $M_2^i \Leftarrow M_2^a$) and ($M_1^i \Leftarrow M_1^b$, $M_2^i \Leftarrow M_2^b$)] is $(1 - R) = [1 - (R_1 + R_2 - 2R_1R_2)]$, where R_1, the frequency of recombinants between M_1 and Q among recombinant inbred lines, is equal to $2r_1/(1 + 2r_1)$ (Haldane and Waddington 1931) and R_2 is equal to $2r_2/(1 + 2r_2)$. The frequency of recombinant marker genotypes [i.e., ($M_1^i \Leftarrow M_1^a$, $M_2^i \Leftarrow M_2^b$), ($M_1^i \Leftarrow M_1^b$, $M_2^i \Leftarrow M_2^a$)] is $R = R_1 + R_2 - 2R_1R_2$. The conditional probability that $Q^i \Leftarrow Q^a$, given G^k_{obs}, is:

$$\begin{aligned}
&\Pr(Q^i \Leftarrow Q^a|G^k_{obs}) \\
&= [\Pr(M_1^i \Leftarrow M_1^a)(1 - R_1)\ \Pr(M_2^i \Leftarrow M_2^a)(1 - R_2)]/(1 - R) \\
&\quad + [\Pr(M_1^i \Leftarrow M_1^a)(1 - R_1)\ \Pr(M_2^i \Leftarrow M_2^b)R_2]/R \\
&\quad + [\Pr(M_1^i \Leftarrow M_1^b)R_1\Pr(M_2^i \Leftarrow M_2^a)(1 - R_2)]/R \\
&\quad + [\Pr(M_1^i \Leftarrow M_1^b)R_1\Pr(M_2^i \Leftarrow M_2^b)R_2]/(1 - R)
\end{aligned}$$

The values of $\Pr(M_1^i \Leftarrow M_1^a)$, $\Pr(M_1^i \Leftarrow M_1^b)$, $\Pr(M_2^i \Leftarrow M_2^a)$, and $\Pr(M_2^i \Leftarrow M_2^b)$ can be easily determined from the marker genotypes, and are equal to 1 or 0 if the marker is polymorphic between *a* and *b*. Suppose *a* has the + + genotype, *b* has the − − genotype, and *i* has the + + genotype at the M_1 locus. In this example, $\Pr(M_1^i \Leftarrow M_1^a) = 1$ and $\Pr(M_1^i \Leftarrow M_1^b) = 0$. But if *a*, *b*, and *i* all have the + + genotype, the values of $\Pr(M_1^i \Leftarrow M_1^a)$ and $\Pr(M_1^i \Leftarrow M_1^b)$ need to be determined from the parental contributions to inbred progeny; i.e., the proportion of the genome derived directly by an inbred from each of its two parents (Bernardo et al. 1997). For example, if *i* is a BC_1-derived inbred, *a* is the recurrent parent whereas *b* is the donor parent, and all three inbreds have the + + genotype, then $\Pr(M_1^i \Leftarrow M_1^a) = 0.75$ and $\Pr(M_1^i \Leftarrow M_1^b) = 0.25$.

The $\Pr(Q^i \Leftarrow Q^b|G^k_{obs})$ term can be obtained in the same manner as $\Pr(Q^i \Leftarrow Q^a|G^k_{obs})$, leading to a solution to $\Pr(Q^i \equiv Q^j|G^k_{obs})$. Or, because Q^i must have descended from either Q^a or Q^b, $\Pr(Q^i \Leftarrow Q^b|G^k_{obs})$ can simply be calculated as $[1 - \Pr(Q^i \Leftarrow Q^a|G^k_{obs})]$. The use of a tabular method for obtaining $\Pr(Q^i \equiv Q^j|G^k_{obs})$ is illustrated in the following numerical example.

Numerical example

Suppose the map distances are $r_1 = 9$ centiMorgans (cM) between M_1 and Q and $r_2 = 6$ cM between Q and M_2. These map distances are equivalent to $R_1 = 0.141432$ and $R_2 = 0.101592$. The frequency of recombinant marker genotypes among recombinant inbred lines is $R = (R_1 + R_2 - 2R_1R_2) = 0.214287$. Consider five hypothetical inbreds with different genotypes at M_1 and M_2 (Table 1). Inbreds L1 and L2 are progenitor inbreds that are unrelated to each other. L3 and L4 are F_2-derived inbreds, whereas L5 is derived from the $[(L4 \times L2) \times L4]BC_1$ population.

Begin by sorting the inbreds so that each parent is listed before any of its progeny. Hence, the inbreds can be arranged as (L1, L2, L3, L4, L5) or (L2, L1, L3, L4, L5) but not in any other order. Set up a table, with a row and a column for each inbred, of $\Pr(Q^i \equiv Q^j|G^k_{obs})$ values by entering 1 in the diagonal elements of this table, i.e., the two alleles at the MQTL in a homozygous line are identical by descent with each other. For the progenitor inbreds (L1 and L2) that are unrelated with each other, enter a $\Pr(Q^i \equiv Q^j|G^k_{obs})$ value of zero in the corresponding off-diagonal element.

Table 1 Pedigrees and marker genotypes of hypothetical inbreds

Inbred	Parent 1	Parent 2	p^a	M_1 locus	M_2 locus
L1				+ +	− −
L2				− −	+ +
L3	L1	L2	0.50	+ +	− −
L4	L2	L3	0.50	+ +	+ +
L5	L4	L2	0.75	+ +	+ +

[a] Parental contribution of Parent 1 to the inbred. The *p* values ignore any relationship between Parent 1 and Parent 2 and have expected values of 0.50 for an F_2-derived inbred, 0.75 for a BC_1-derived inbred and its recurrent parent, 0.25 for a BC_1-derived inbred and its donar parent, 0.875 for a BC_2-derived inbred, etc

The $\Pr(Q^i \equiv Q^j|G^k_{obs})$ value for the first non-zero off-diagonal element in the first row (i.e., L1, L3) is calculated as (Eq. 1):

$$\begin{aligned}
\Pr(Q^{L3} \equiv Q^{L1}|G^k_{obs}) &= \Pr(Q^{L3} \Leftarrow Q^{L1}|G^k_{obs})\ \Pr(Q^{L1} \equiv Q^{L1}|G^k_{obs}) \\
&\quad + \Pr(Q^{L3} \Leftarrow Q^{L2}|G^k_{obs})\ \Pr(Q^{L2} \equiv Q^{L1}|G^k_{obs})
\end{aligned}$$

Now,

$$\begin{aligned}
&\Pr(Q^{L3} \Leftarrow Q^{L1}|G^k_{obs}) \\
&= [\Pr(M_1^{L3} \Leftarrow M_1^{L1})\ (1 - R_1)\ \Pr(M_2^{L3} \Leftarrow M_2^{L1})\ (1 - R_2)]/(1 - R) \\
&\quad + [\Pr(M_1^{L3} \Leftarrow M_1^{L1})\ (1 - R_1)\ \Pr(M_2^{L3} \Leftarrow M_2^{L2})\ R_2]/R \\
&\quad + [\Pr(M_1^{L3} \Leftarrow M_1^{L2})\ R_1\Pr(M_2^{L3} \Leftarrow M_2^{L1})\ (1 - R_2)]/R \\
&\quad + [\Pr(M_1^{L3} \Leftarrow M_1^{L2})\ R_1\Pr(M_2^{L3} \Leftarrow M_2^{L2})R_2]/(1 - R) \\
&= [(1)\ (0.858568)\ (1)\ (0.898408)]/0.785713 \\
&\quad + [(1)\ (0.858568)\ (0)\ (0.101592)]/0.214287 \\
&\quad + [(0)\ (0.141432)\ (1)\ (0.898408)]/0.214287 \\
&\quad + [(0)\ (0.141432)\ (0)\ (0.101592)]/0.785713 \\
&= 0.981713
\end{aligned}$$

Similarly, $\Pr(Q^{L3} \Leftarrow Q^{L2}|G^k_{obs}) = 0.018287$. Substituting the $\Pr(Q^{L3} \Leftarrow Q^{L1}|G^k_{obs})$ and $\Pr(Q^{L3} \Leftarrow Q^{L2}|G^k_{obs})$ values, which should sum to 1, in Eq. 1:

$$\Pr(Q^{L3} \equiv Q^{L1}|G^k_{obs}) = 0.981713(1) + 0.018287(0) = 0.981713$$

The rest of the $\Pr(Q^i \equiv Q^j|G^k_{obs}$ values in the first row [(L1, L4), (L1, L5)], followed by those in the second, third, fourth and fifth rows, can be calculated with the same procedure (Table 2). Note

Table 2 Conditional probability, given observed marker genotypes, of identity by descent of MQTL alleles among five hypothetical inbreds

	L1	L2	L3	L4	L5
L1	1	0	0.9817	0.3996	0.3349
L2		1	0.0183	0.6004	0.6651
L3			1	0.4179	0.3532
L4				1	0.9353
L5					1

that the parents of L5 are not polymorphic at M_2, and $\Pr(M_2^{L5} \Leftarrow M_2^{L4}) = 0.75$ and $\Pr(M_2^{L5} \Leftarrow M_2^{L2}) = 0.25$.

TM-BLUP for single crosses

Assume there are n_1 inbreds in Group I and n_2 inbreds in Group II. The population comprises the n_1n_2 possible crosses between Group I and Group II inbreds. In practice, many of the n_1n_2 potential single crosses may not have been evaluated in field trials, but the performance of such untested single crosses may be predicted from the performance of the tested single crosses (Bernardo 1996). Suppose there are q MQTL for the trait of interest, each MQTL being flanked by a pair of markers and each MQTL being independent from each other. The mixed linear model in TM-BLUP is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{Z}_2\mathbf{a}_2 + \mathbf{Z}\mathbf{d} + \sum_{k=1,q} \mathbf{W}_1^k\mathbf{v}_1^k + \sum_{k=1,q} \mathbf{W}_2^k\mathbf{v}_2^k + \sum_{k=1,q} \mathbf{W}^k\mathbf{d}_M^k + \mathbf{e}$$

where: $\mathbf{y}$ = vector of observed means of single crosses for a given trait; $\boldsymbol{\beta}$ = vector of fixed effects; $\mathbf{a}_1 = n_1 \times 1$ vector of testcross additive effects of Group I inbreds at unmarked QTL; $\mathbf{a}_2 = n_2 \times 1$ vector of testcross additive effects of Group II inbreds at unmarked QTL; $\mathbf{d} = n_1n_2 \times 1$ vector of dominance effects of tested and untested single crosses at unmarked QTL; $\mathbf{v}_1^k = n_1 \times 1$ vector of testcross additive effects of Group I inbreds at the kth MQTL; $\mathbf{v}_2^k = n_2 \times 1$ vector of testcross additive effects of Group II inbreds at the kth MQTL; $\mathbf{d}_M^k = n_1n_2 \times 1$ vector of dominance effects of tested and untested single crosses at the kth MQTL; $\mathbf{e}$ = vector of residual effects; and $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}$, $\mathbf{W}_1^k$, $\mathbf{W}_2^k$, and $\mathbf{W}^k$ are incidence matrices of 1's and 0's relating $\mathbf{y}$ to the respective fixed or random effects. The covariance matrix of random genetic effects with one MQTL is:

$$\mathrm{Var}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\\ \mathbf{d}\\ \mathbf{v}_1^k\\ \mathbf{v}_2^k\\ \mathbf{d}_M^k\end{bmatrix} =$$

$$\begin{bmatrix}
\mathbf{A}_1V_{A(1)} & & & & & \\
\mathbf{0} & \mathbf{A}_2V_{A(2)} & & & (\text{Symmetric}) & \\
\mathbf{0} & \mathbf{0} & \mathbf{D}V_D & & & \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}_1^kV_{MA}^k(1) & & \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}_2^kV_{MA}^k(2) & \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_M^kV_{MD}^k
\end{bmatrix}$$

The above covariance matrix can be expanded to include more than one MQTL. For the unmarked QTL, the elements of the genetic relationship matrices are f_{ij} in $\mathbf{A}_1$, $f_{i'j'}$ in $\mathbf{A}_2$, and $f_{ij}\,f_{i'j'}$ in $\mathbf{D}$. For the kth MQTL, the elements of the conditional genetic relationship matrices, given the observed marker genotypes, are $\Pr(Q^i \equiv Q^j|G_{obs}^k)$ in $\mathbf{M}_1^k$, $\Pr(Q^{i'} \equiv Q^{j'}|G_{obs}^k)$ in $\mathbf{M}_2^k$, and $\Pr(Q^i \equiv Q^j|G_{obs}^k)$ $\Pr(Q^{i'} \equiv Q^{j'}|G_{obs}^k)$ in $\mathbf{D}_M^k$. The $V_{MA(1)}^k$, $V_{MA(2)}^k$, and V_{MD}^k MQTL variances can be estimated by restricted maximum likelihood procedures outlined by Henderson (1985), i.e., in the same manner as the trait variances at unmarked QTL [$V_{A(1)}$, $V_{A(2)}$, and V_D] are estimated. Non-normal distributions are expected for MQTL effects. However, Goddard (1992) found that such departures from normality do not decrease the accuracy of BLUP.

Inverse of $\mathbf{D}$ and $\mathbf{D}_M^k$

Maize breeders typically evaluate up to a few hundred experimental inbreds in each heterotic group, but there are thousands of potential interpopulation single crosses (Bernardo 1996). Whereas obtaining the inverse of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{M}_1^k$ $\mathbf{M}_2^k$ will not be difficult, obtaining the inverse of $\mathbf{D}$ and $\mathbf{D}_M^k$ may be prohibitive. For example, if $n_1 = 200$ and $n_2 = 100$, then the dimensions of the conditional genetic relationship matrices for the MQTL are 200×200 for $\mathbf{M}_1^k$ 100×100 for $\mathbf{M}_2^k$, and $20,000 \times 20,000$ for $\mathbf{D}_M^k$. An efficient method to obtain the inverses of $\mathbf{D}$ and $\mathbf{D}_M^k$ would be useful.

For the sake of illustration, assume there are only three inbreds in Group I and two inbreds in Group II, and the Group I $\times$ Group II crosses are sorted as follows: $[1 \times 1, 1 \times 2, 2 \times 1, 2 \times 2, 3 \times 1, 3 \times 2]$. Because inbreds in Group I are unrelated to inbreds in Group II, $\mathbf{D}_M^k$ is equal to the Kronecker product (denoted by $\otimes$, Searle 1966 p 215) of $\mathbf{M}_1^k$ and $\mathbf{M}_2^k$:

$$\mathbf{D}_M^k = \mathbf{M}_1^k \otimes \mathbf{M}_2^k = \begin{bmatrix} m_1^{k(11)}\mathbf{M}_2^k & m_1^{k(12)}\mathbf{M}_2^k & m_1^{k(13)}\mathbf{M}_2^k \\ m_1^{k(21)}\mathbf{M}_2^k & m_1^{k(22)}\mathbf{M}_2^k & m_1^{k(23)}\mathbf{M}_2^k \\ m_1^{k(31)}\mathbf{M}_2^k & m_1^{k(32)}\mathbf{M}_2^k & m_1^{k(33)}\mathbf{M}_2^k \end{bmatrix}$$

Assume the elements of the inverse of $\mathbf{M}_1^k$ are designated $c_1^{k(ii')}$. The inverse of a Kronecker product is equal to the Kronecker product of the inverses of the matrices. Therefore, the inverse $\mathbf{D}_M^k$ is equal to the Kronecker product of the inverses of $\mathbf{M}_1^k$ and $\mathbf{M}_2^k$:

$$(\mathbf{D}_M^k)^{-1} = (\mathbf{M}_1^k)^{-1} \otimes (\mathbf{M}_2^k)^{-1} = \begin{bmatrix} c_1^{k(11)}(\mathbf{M}_2^k)^{-1} & c_1^{k(12)}(\mathbf{M}_2^k)^{-1} & c_1^{k(13)}(\mathbf{M}_2^k)^{-1} \\ c_1^{k(21)}(\mathbf{M}_2^k)^{-1} & c_1^{k(22)}(\mathbf{M}_2^k)^{-1} & c_1^{k(23)}(\mathbf{M}_2^k)^{-1} \\ c_1^{k(31)}(\mathbf{M}_2^k)^{-1} & c_1^{k(32)}(\mathbf{M}_2^k)^{-1} & c_1^{k(33)}(\mathbf{M}_2^k)^{-1} \end{bmatrix}$$

Likewise, for unmarked QTL, $\mathbf{D} = \mathbf{A}_1 \otimes \mathbf{A}_2$ and $\mathbf{D}^{-1} = (\mathbf{A}_1)^{-1} \otimes (\mathbf{A}_2)^{-1}$. Hence, if $n_1 = 200$ and $n_2 = 100$, the inverses of the $20,000 \times 20,000$ $\mathbf{D}$ and $\mathbf{D}_M^k$ matrices can each be obtained directly from the inverses of two much smaller (200×200 and 100×100 matrices.

Discussion

Recombination frequencies between flanking markers and MQTL are the crucial information needed in TM-BLUP. Despite the availability of molecular markers in maize for more than a decade, no comprehensive publicly available map of QTL is available for complex traits such as grain yield. Different methods for mapping QTL are discussed by Doerge et al. (1994). Perhaps a logical approach is to develop a single mapping population from a biparental cross that represents common heterotic groups, e.g., B73 $\times$ Mo17 in maize. Large numbers of progenies (> 500) need to be evaluated in a large number of environments to obtain reliable estimates of the location of MQTL (Beavis 1994).

The TM-BLUP approach requires marker information on inbreds and all their progenitors. This requirement should not be a limitation in current maize breeding programs, wherein (1) marker genotypes of elite inbreds are often, if not routinely, obtained for varietal protection purposes (Smith and Beavis 1996),

and (2) new inbreds are developed from crosses between elite inbreds, i.e., second-cycle breeding.

The TM-BLUP model may pose computational difficulties (i.e., inverting large matrices) given the numbers of inbreds and tested and untested single crosses available in maize breeding programs. The computational difficulties increase as the number of MQTL increases. Several approaches may reduce the computational requirements in TM-BLUP. First, a simplified model may be used wherein the dominance effects of MQTL alleles ($\mathbf{d}_M^k$) are excluded. Inclusion of $\mathbf{d}_M^k$ in TM-BLUP greatly increases the number of equations that need to be solved. If single crosses are made between n_1 inbreds in Group I and n_2 inbreds in Group 2, the number of random genetic effects in TM-BLUP is $(1 + q)$ $(n_1 + n_2 + n_1n_2)$, where q is the number of MQTL. Hence, if $n_1 = n_2 = 100$ and 10 MQTL are known, the size of the coefficient matrix for random genetic effects in TM-BLUP is $112{,}200 \times 112{,}200$. Bernardo (1996) found that, for data that did not include markers, the average ratio of V_D to $(V_{A(I)} + V_{A(II)} + V_D)$ for grain yield was only 0.26. If V_D is low, then estimates of V_D^k are also expected to be low and perhaps $\mathbf{d}_M^k$ may be excluded in the model to facilitate computation. In the preceding example, the size of the coefficient matrix for random effects is reduced to $12{,}200 \times 12{,}200$ if $\mathbf{d}_M^k$ is excluded.

A second approach for reducing the size of the coefficient matrix is to include only the tested single crosses in the model. Bernardo (1996) indicated that only about 10% of the potential n_1n_2 single crosses are actually tested in field trials and have prior data available. If the analysis is limited only to the tested single crosses, the number of random genetic effects in TM-BLUP is $(1 + q)$ $(n_1 + n_2 + tn_1n_2)$, where t is the proportion of potential single crosses that have been tested. If $n_1 = n_2 = 100$, $q = 10$, and $t = 0.10$, the size of the coefficient matrix for random effects is reduced to $13{,}200 \times 13{,}200$. The performance of the untested single crosses can subsequently be predicted from the performance of the tested single crosses (Bernardo 1996). But a disadvantage of limiting the TM-BLUP analysis to a subset of all potential n_1n_2 single crosses is that the inverse of $\mathbf{D}_M^k$ (or $\mathbf{D}$) can no longer be obtained directly as the Kronecker product of the inverses of $\mathbf{M}_1^k$ and $\mathbf{M}_2^k$ (or $\mathbf{A}_1$ and $\mathbf{A}_2$).

A third approach for reducing the size of the coefficient matrix is to pool information on conditional genetic covariances across the MQTL loci, as proposed by van Arendonk et al. (1994). If there are q independent MQTL, this procedure requires q separate TM-BLUP analyses, each involving one MQTL at a time. With $n_1 = n_2 = 100$ and $q = 10$, each of the 10 TM-BLUP analyses would involve $2(n_1 + n_2 + n_1n_2) = 20{,}400$ random effects. After $V_{MA(1)}^k$, $V_{MA(2)}^k$ and V_{MD}^k have been estimated for each of the q MQTL, matrices of conditional genetic covariances, pooled across the q MQTL, may be obtained (see van Arendonk et al. 1994 for details). In the final TM-BLUP analysis, which involves covariances at unmarked QTL and pooled covariances of MQTL effects, the size of the coefficient matrix for random effects is $20{,}400 \times 20{,}400$.

The three approaches for reducing the computational requirements in TM-BLUP may be used singly or in combination with each other. In the preceding example, the number of random effects is further reduced to $[(1 + q)$ $(n_1 + n_2) + tn_1n_2] = 3200$ if $\mathbf{d}_M^k$ is excluded and the analysis is limited to the single crosses that have been tested.

Bernardo (1996) found that best linear unbiased prediction based on trait data alone (T-BLUP) can effectively predict the performance of maize single crosses prior to field testing. The usefulness of TM-BLUP compared with T-BLUP needs to be studied. Computer simulation research is needed to determine the usefulness of TM-BLUP with varying trait heritability, number of MQTL, and proportion of the genetic variance explained by MQTL. Finally, TM-BLUP and T-BLUP need to be compared when applied to empirical data sets that are available to breeders.

Acknowledgements I thank Prof. W.E. Nyquist for critically reviewing the manuscript and offering helpful comments and suggestions.

References

Ajmone-Marsan P, Monfredini G, Ludwig WF, Melchinger AE, Franceschini P, Pagnotto G, Motto M (1995) In an elite single cross of maize quantitative trait locus controls one-fourth of the genetic variation for grain yield. Theor Appl Genet 90:415–424

Austin DF, Lee M (1996) Comparative mapping in $F_{2:3}$ and $F_{6:7}$ generations of quantitative trait loci for grain yield and yield components in maize. Theor Appl Genet 92:817–826

Beavis WD (1994) The power and deceit of QTL experiments. Proc Annu Corn Sorghum Res Conf 49:250–266

Beavis WD, Keim P (1995) Identification of quantitative trait loci that are affected by environment, In: Kang MS, Gauch HG (eds) Genotype-by-environment interaction. CRC Press, Boca Raton, pp 123–149

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative trait loci. Theor Appl Genet 83:141–145

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F_4 progeny from maize. Crop Sci 34:882–896

Bernardo R (1996) Best linear unbiased prediction of the performance of crosses between untested maize inbreds. Crop Sci 36:872–876

Bernardo R, Murigneux A, Maisonneuve JP, Johnson C, Karaman Z (1997) RFLP-based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor Appl Genet 94:652–656

Bubeck DM, Goodman MM, Beavis WD, Grant D (1993) Quantitative trait loci controlling resistance to gray leaf spot in maize. Crop Sci 33:838–847

Doerge RW, Zeng ZB, Weir BS (1994) Statistical issues in the search for genes affecting quantitative traits in populations. In: Analysis of molecular marker data (Proc Joint Plant Breed Symp Ser). Am Soc Hortic Sci (Alexandria, VA) and Crop Sci Soc Am, pp. 15–26

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

Eathington SR, Dudley JW, Rufener GK (1997) Marker effects estimated from testcrosses of early and late generations of inbreeding in maize. Crop Sci 37:1679–1685

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Envol 21:467–477

Freymark PJ, Lee M, Woodman WL, Martinson CA (1993) Quantitative and qualitative trait loci affecting host-plant response to *Exserohilum turcicum* in maize (*Zea mays* L.). Theor Appl Genet 87:537–544

Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878–886

Goldman IL, Rocheford TR, Dudley JW (1993) Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. Theor Appl Genet 87:217–224

Haldane JBS (1919) The combination of linkage values and the calculation of distance between loci of linked factors. J Genet 8:299–309

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J Anim Sci 60:111–117

Koester RP, Sisco PH, Stuber CW (1993) Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines of maize. Crop Sci 33:1209–1216

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lee M (1993) Genetic analysis of resistance to European corn borer and northern corn leaf blight in maize. Proc Annu Corn Sorghum Res Conf 48:213–223

Martinez O, Curnow RN (1992) Estimating the locations and sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

Schön CC, Lee M, Melchinger AE, Guthrie WD, Woodman WL (1993) Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Heredity 70:648–659

Schön CC, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG, Seitzer JF (1994) RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci 34:378–389

Schnell FW (1965) Die covarianz zwischen verwandten in einer gen-orthogonalen population. I. Allgemeine theorie. Biom 7:2–29

Searle SR (1996) Matrix algebra for the biological sciences (including applications in statistics). John Wiley & Sons, New York

Smith JSC, Beavis WD (1996) Molecular marker assisted breeding in a company environment. In: Sobral BWS (ed) The impact of plant molecular genetics. Bikhäuser, Boston, pp 259–272

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Stuber CW, Cockerham CC (1966) Gene effects and variances in hybrid populations. Genetics 54:1279–1286

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

van Arendonk JAM, Tier B, Kinghorn BP (1994) Use of multiple genetic markers in prediction of breeding values. Genetics 137:319–329

Veldboom LR, Lee M, (1994) Molecular marker-facilitated studies of morphological traits in maize. II. Determination of QTLs for grain yield and yield components. Theor Appl Genet 89:451–458

Veldboom LR, Lee M, Woodman WL (1994) Molecular marker-facilitated studies in an elite maize population. I. Linkage analysis and determination of QTL for morphological traits. Theor Appl Genet 88:7–16

Wang T, Fernando RL, van der Beek S, Grossman M, van Arendonk JAM (1995) Covariance between relatives for a marked quantitative trait locus. Genet Sel Evol 27:251–274

Zehr BE, Dudley JW, Chojecki J, Saghai-Maroof MA, Mowers RP (1992) Use of RFLP markers to search for alleles in a maize population for improvement of an elite hybrid. Theor Appl Genet 83:903–911

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468