Prediction of additive genetic effects for the QTL-cluster on the basis of data on surrounding markers in outbred populations

Hirokazu MATSUDA¹, Hiroaki IWAISAKI¹²

¹ Course of Environmental Management Science, Graduate School of Science and Technology, ² Department of Animal Science, Faculty of Agriculture, Niigata University, Niigata, Japan

Abstract. This paper described a method for predicting additive effects of a cluster of tightly linked QTLs for outbred populations of animals in the situation where the QTLs are located on a chromosome segment surrounded by multiple linked DNA markers. We present a mixed model method for best linear unbiased prediction (conditional to the marker data) of the additive effects of the QTL-cluster and of the remaining QTLs unlinked to the marker linkage group. This method takes into consideration the identity-by-descent proportion (IBDP) for the particular chromosomal segment, in contrast to some other methods which use IBD probabilities at one specific location. In this method, fully informative data on different flanking markers is used to calculate the values of the expectations of the IBDPs (EIBDPs) between gametes for animals to be evaluated. Then the expected values are used as the elements of the gametic relationship matrix required in the best linear unbiased prediction. Giving a small numerical example, we illustrate how the present method can be used for the prediction of the QTL-cluster effects and for genetic evaluation of animals in outbred populations. A computational strategy is discussed on the basis of the calculation of the EIBDPs and the inverted gametic relationship matrix in complex pedigrees.

Key words: additive genetic effect, BLUP, DNA marker linkage group, gametic relationships, identity-by-descent proportion, outbred population, QTL-cluster.

List of abbreviations: MAS = marker-assisted selection, QTLs = quantitative trait loci, BLUE = best linear unbiased estimation (or estimator), BLUP = best linear unbiased prediction (or predictor), IBD = identity-by-descent, IBDP = identity-by-descent proportion EIBDP = expectation of identity-by-descent proportion.

Received: September 11, 2001. Accepted: January 31, 2002.
Correspondence: H. IWASAKI, Department of Animal Science, Faculty of Agriculture, Niigata University, Igarashi nino-cho 8050, Niigata-shi 950-2181, Japan, e-mail: iwsk@agr.niigata-u.ac.jp
In recent years, a variety of studies have been undertaken in relation to manipulat-
ing loci that control quantitative traits with the use of genetic marker data. Artifi-
cial selection for quantitative traits utilizing marker information, that is
the so-called marker-assisted selection (MAS), could be effective in certain cases.
MAS is expected to increase the efficiency of selection, in particular, for lowly
heritable traits, sex-limited traits and traits for which selection is usually carried
out before the phenotypic information is provided (e.g., LANDE, THOMPSON
1990, GIMELFARB, LANDE 1994, RUANE, COLLEAU 1995, RUANE,

In a crossing design starting from inbred parental strains in which individuals
in each strain have identical genotypes, there are only two alleles segregating at
a locus. For the population of this type, there exists linkage disequilibrium maxi-
mized throughout the population between markers and quantitative trait loci
(QTLs), and all individuals are informative. Thus, the information about
the marker genotypes provides information on the QTL genotypes and conse-
quently on breeding values of individuals (e.g., LANDE, THOMPSON 1990,
GIMELFARB, LANDE 1994).

In contrast, in outbred populations where any number of alleles can be segre-
gating, marker-QTL linkage phases can differ among individuals and no associa-
tions between markers and quantitative traits are found at the population level,
so that marker-trait associations are necessary to examine for each parent
of a given individual. In this case, the information on the allele transmission
within the family for markers closely linked to the QTL(s) can rather be used to
predict breeding values (OLLIVIER 1998). For outbred populations of animals,
showing how a single marker and trait information can be utilized, a mixed model
method for the best linear unbiased prediction (BLUP) of individuals’ additive ge-
netic merits (HENDERSON 1973, 1975, 1984) was initially proposed by
FERNANDO and GROSSMAN (1989). The BLUP methods of this kind usually as-
sume the situation where a QTL is marked by a single marker or by flanking mark-
kers (e.g., CANTET, SMITH 1991, GODDARD 1992, HOESCHELE 1993, VAN
AREndonk et al. 1994, SAITO, IWAIASKI 1996, SAITO, IWAIASKI 1997a,b,
SAITO et al. 1998).

On the other hand, it is likely that a cluster of multiple linked QTLs is located
in a certain chromosome region (e.g., KHAVKIN, COE 1997, 1998). In such a set-
ing, since it is suggested that a detected major QTL may reflect the expression of
genes with a similar function for multiple linked QTLs positioned in the specific
region, at least one of those loci seems to be segregating and polymorphic. Ac-
cordingly, for outbred populations, a mixed model approach for prediction of ge-
netic effects using information on markers just flanking a cluster of QTLs was
first discussed by MATSUDA and IWAIASKI (1998). In outbred populations,
however, markers may be partially informative, so that the use of only the flank-
ing markers may not extract the maximum amount of information on the segregation of the QTLs marked. Additional use of information from multiple linked upstream markers outside the interval containing the QTLs is expected to increase the accuracy of the prediction.

The purpose of this paper is to present a mixed model procedure for predicting additive genetic effects of a marked cluster of linked QTLs that is contained in a chromosome segment and the effects of the remaining QTLs, using data on different flanking markers that are fully informative of multiple linked markers.

**Theory and method**

**The mixed linear model**

We consider a situation where a cluster of multiple linked QTLs is located in a particular region of a chromosome bracketed with two flanking marker loci, that is further followed by multiple linked upstream markers. For animal \( i \) in an outbred population, let the additive effects of paternal and maternal origin due to the marked cluster of QTLs be \( v_i^0 \) and \( v_i^1 \), respectively, and the additive effect for remaining polygenes be \( u_i \). Then the total breeding value of the \( i \)-th individual \((a_i)\) is expressed as

\[
a_i = u_i + v_i^0 + v_i^1. \tag{i}
\]

Then the mixed linear model to describe observations is written in matrix notation as

\[
y = Xb + Zu + Wv + e, \tag{ii}
\]

where \( y \) is an \( N \times 1 \) vector of observations, \( b \) is an \( f \times 1 \) vector of fixed effects including discrete variables and/or covariates, \( u \) is an \( h \times 1 \) random vector of the additive genetic effects due to the remaining polygenes, \( v \) is a \( 2h \times 1 \) random vector of the additive genetic effects due to the marked cluster of QTLs, \( e \) is an \( N \times 1 \) random vector of residuals, and \( X, Z \) and \( W \) are the \( N \times f, N \times h \) and \( N \times 2h \) incidence matrices pertaining to \( b, u \) and \( v \), respectively.

Since there is linkage equilibrium between markers and QTLs in the outbred population, marker genotypes do not provide information on the expectation of breeding value, but contribute to information about the variance. Accordingly, it is assumed here that

\[
E[[u \ v \ e]] = [[0 \ 0 \ 0]] \quad \text{and} \quad \text{Var}[[u \ v \ e]] = \begin{bmatrix} A_u \sigma_u^2 & 0 & 0 \\ 0 & G_v \sigma_v^2 & 0 \\ 0 & 0 & I \sigma_e^2 \end{bmatrix},
\]

where

- \( A_u \) is the additive relationship matrix
- \( G_v \) is the genetic relationship matrix
- \( I \) is the identity matrix
- \( \sigma_u, \sigma_v, \sigma_e \) are the additive genetic, dominance genetic, and residual variance components.
where \( A_u \) is the additive relationship matrix among animals, \( G_v \) is the so-called gametic relationship matrix for the marked cluster of QTLs, \( I \) is an identity matrix, \( \sigma_u^2 \) and \( 2\sigma_v^2 \) represent the components of variance due to the remaining polygenes and the cluster of QTLs, respectively, and \( \sigma_e^2 \) is the variance component for residuals.

**Calculation of the gametic relationship matrix**

As the elements of the gametic relationship matrix \( G_v \) in the model (ii), we use the EIBDP or the expectation of the proportion of genetic materials identical by descent (IBD) shared by a group of relatives in the specified chromosomal region in which the cluster of QTLs is contained (GUO 1994a,b, 1995). Note that for a given pedigree structure, while the single locus IBD probability is a constant, the proportion of IBD (IBDP) mentioned herein is a random variable with the range of 0 to 1.

We here assume the use of information on multiple linked markers surrounding the chromosome segment in which the cluster of QTLs is positioned. Let \( m+1 \) linked markers on the chromosome be \( M_0, M_1, ..., M_q, M_{q+1}, ..., M_m \). The relative positions (Morgans, \( M \)) of the markers are assumed to be \( 0 < l_1 < ... < l_q < l_{q+1} < ... < l_m \), and we denote the chromosome region containing the cluster of QTLs by \( l_q \leq t \leq l_{q+1} \), assuming that the marker linkage phases are known. The length of the region from the origin at which the \( k \)th crossover occurred is denoted by \( S_k \). Then, following GUO (1995), the paternal or maternal chromosome of the offspring can be represented by a time-continuous, two-state Markov chain \( g(t) \), as follows:

\[
g(t) = \begin{cases} C & S_{2k} \leq t < S_{2k+1} \leq l_m \\ 1-C & S_{2k+1} \leq t < S_{2k+2} \leq l_m \end{cases}
\]

where \( C \) is a random variable and takes values 0 and 1 with equal probability for the father’s or the mother’s paternal chromosome and for the father’s or the mother’s maternal chromosome, respectively, at any point \( t \) along the chromosome region, and the time parameter is the map distance along the chromosome. This was called the gametogenesis process by GUO (1995). All the gametogenesis processes in a pedigree are independent and stochastically identical. Then, using HALDANE’s (1919) mapping function, we have the following transition probability matrix for \( g(t) \):

\[
P(t) = (p_{g(t)}) = \begin{pmatrix} 1 + e^{-2t} & 1 - e^{-2t} \\ 1 - e^{-2t} & 1 + e^{-2t} \end{pmatrix}
\]
Further following GUO (1995), let the \( n \) relevant gametogenesis processes be

\[
\psi(t) = \left[ g_1(t), g_2(t), \ldots, g_n(t) \right].
\]

Then this joint gametogenesis process, which is a random vector, constitutes a random walk on an \( n \)-dimensional hypercube denoted as \( Z^n = \{ (\eta_1, \eta_2, \ldots, \eta_n) : \eta_i = 0 \text{ or } 1 \} \). Thus, defining the set of IBD states by \( D \) as the collection of vertices on \( Z^n \), for a given pedigree structure and \( D \), the IBDP is given as

\[
R(l_{q+1} - l_q) = \frac{1}{l_{q+1} - l_q} \int_0^{l_{q+1} - l_q} \delta[\psi(t) \in D] dt,
\]

where \( \delta(E) = 1 \) if \( E \) is true and \( \delta(E) = 0 \) if not.

Now, for the current situation with the \( m + 1 \) markers available, denote the information on parental origin by \( \psi(0) = \psi_0 \), \( \psi(l_1) = \psi_1 \), \( \psi(l_2) = \psi_2 \), \( \psi(l_{q+1}) = \psi_{q+1} \), \( \ldots, \psi(l_m) = \psi_m \). Then, given the marker information (denoted as \( M \)) that is 0 and 1 for the paternal and the maternal chromosome, respectively, the EIBDP between the gametes of given relatives can be expressed as

\[
E(R(l_{q+1} - l_q)|M) = \frac{1}{(l_{q+1} - l_q)^\frac{d}{2}} \prod_{r=0}^{m-q+1} \left[ P_{00}(l_q - l_r) \right]^{a_0 - H(v, \nu_r)} \left[ P_{01}(l_q - l_r) \right]^{H(v, \nu_r)}
\]

\[
\times \sum_{v, w \in D} \int_0^{l_{q+1} - l_q} \left[ P(\psi(t) = v) ight] \left[ P(\psi_0 = v, \psi(l_1) = \psi_1, \psi(l_2) = \psi_2, \ldots, \psi(l_{q+1}) = \psi_{q+1}, \ldots, \psi(l_m) = \psi_m | \psi(t) = v) \right] dt
\]

\[
= \frac{1}{(l_{q+1} - l_q)^\frac{d}{2}} \prod_{r=0}^{m-q+1} \left[ P_{00}(l_q - l_r) \right]^{a_0 - H(v, \nu_r)} \left[ P_{01}(l_q - l_r) \right]^{H(v, \nu_r)}
\]

\[
\times \sum_{v, w \in D} \int_0^{l_{q+1} - l_q} \left[ \int_0^{l_{q+1} - l_q} \left[ P_{00}(l_q - l_r + t) \right]^{a_0 - H(v, \nu_r)} \left[ P_{01}(l_q - l_r + t) \right]^{H(v, \nu_r)} dt \right]
\]

\[
\times \left[ P_{00}(l_q - l_r - t) \right]^{a_0 - H(v, \nu_r)} \left[ P_{01}(l_q - l_r - t) \right]^{H(v, \nu_r)} dt,
\]

(vii)
where $a_{rs}$ with $\sum a_{rs} = n$ represents the number of the gametogenesis processes for which the informative and possibly close two markers bracketing the QTLs-region are $r$ and $s$ among $n$ processes of the joint gametogenesis process, and with denoting two joint gametogenesis processes by

$$H(\eta_j, \eta_{j'}) = \sum (\eta_{j'} - \eta_j)$$

that is the sum in the cases where $M_q$ and $M_{q+1}$ is informative, when $\eta_j$ and $\eta_{j'}$ are $v_q$ or $v_{q+1}$, respectively, and the sum in the cases where $M_q$ and $M_{q+1}$ is not informative, when $\eta_j$ and $\eta_{j'}$ are $v_0, v_1, \ldots, v_{q-1}$ or $v_{q+2}, \ldots, v_m$, respectively. For instance, of $a_{rs}$, the numbers of the processes in which recombination occurred and did not occur between the two loci, given $r$ and $s$, are represented by $H(v_s, v_r)$ and $a_{rs} - H(v_s, v_r)$, respectively.

### Mixed model equations to be solved

Construction of the inverse of additive relationship matrix $A_u^{-1}$ can be performed by the usual procedure (HENDERSON 1976, QUAAS 1976). Then the predictands are predicted by a linear function of observations $L'y$, such that the predictor is unbiased and prediction error variance is minimized. The mixed model equations to be solved actually are given as

$$
\begin{bmatrix}
X'X & XZ & XW \\
Z'X & ZZ + A_u^{-1} \alpha_u & ZW \\
W'X & WZ & WW + G_v^{-1} \alpha_v
\end{bmatrix}
\begin{bmatrix}
\hat{b} \\
\hat{u} \\
\hat{v}
\end{bmatrix}
=
\begin{bmatrix}
X'y \\
Z'y \\
W'y
\end{bmatrix},
$$

where $\alpha_u$ and $\alpha_v$ are $\sigma_u^2 / \sigma_v^2$ and $\sigma_e^2 / \sigma_v^2$, respectively, $\hat{b}$ is the best linear unbiased estimators (BLUEs) of estimable functions of $b$, and $\hat{u}$ and $\hat{v}$ are BLUPs of $u$ and $v$, respectively.

### Numerical illustration

In this section, using a small data-sample as shown in Table 1, we illustrate the current prediction procedure. An illustrative representation of a marker linkage group assumed is given in Figure 1. Pedigree structure and pathways of marker allele transmission for the example are shown in Figure 2. It is assumed here that for the relative locations of the markers A-E, $l_A = 0$, $l_B = 0.1$, $l_C = 0.3$, $l_D = 0.35$ and $l_E = 0.4$M, and that a cluster of linked QTLs is located on the interval just bracketed with markers B and C. The variance components assumed are $\sigma^2_u = 120$, $\sigma^2_v = 20$ and $\sigma^2_e = 240$.

For the given data, we note that marker loci C and B are not informative for animals 1 and 2, respectively. Then, for instance, for the EIBDP between the mater-
nal gamete for animal 2 and that for animal 4, the current method uses information on the two informative markers A and C, as follows:

\[
\frac{1}{(0.3-0.1)p_{01}(0.3-0)} \int_{0}^{0.3-0.1} [p_{01}(0.2-0.1+t)p_{00}(0.3-0.1-t)]dt = 0.6656, \\
\]

where \( p_{01}(t) \) and \( p_{00}(t) \) denote the transition probabilities at the map distance \( t \), as shown in equation (iv). It should be noted that when data on the flanking markers B and C are used, with which the QTL-cluster is just bracketed, then marker locus B is uninformative, and consequently the corresponding value is regarded as a higher one:

\[
\frac{1}{(0.3-0.1)} \int_{0}^{0.3-0.1} [p_{01}(0.2-0.1+t)]dt = 0.9121. \\
\]

Furthermore, for marker allele transmission from animal 4 to animal 5, we note that animal 4 is homozygous at markers C and D. Thus, the information on markers B and E is used. Then, with the current method, the value of the EIBDP:

\[
\]

---

**Table 1. Trait data including genetic marker information**

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Marker haplotype(^1)</th>
<th>Level(^2)</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–</td>
<td>–</td>
<td>A1B1C2D1E1 / A2B2C2D2E2</td>
<td>1</td>
<td>143</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
<td>–</td>
<td>A3B3C3D3E3 / A4B3C2D1E4</td>
<td>2</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>A2B2C2D2E2 / A4B3C2D1E4</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>A1B1C2D1E1 / A3B3C2D1E4</td>
<td>2</td>
<td>130</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>A2B2C2D2E2 / A3B3C2D1E4</td>
<td>2</td>
<td>96</td>
</tr>
</tbody>
</table>

\(^1\) A – E stand for the five marker loci, with alleles represented by different figures. 
\(^2\) Level of the fixed factor.
Figure 2. Pedigree structure and pathways of marker allele transmission. The upper and lower boxes for each animal represent two chromosomes of paternal and maternal origin, respectively. The symbols 0 and 1, and * denote alleles of paternal and maternal origin in one parent, respectively, and the uninformative case. The five positions from left to right in a box correspond to markers A to E.
Prediction of effects for the QTL-cluster

\[
\frac{1}{(0.3-0.1)} p_{00}(0.3-0) p_{01}(0.3-0) p_{00}(0.4-0.1) \\
\times \int_{0}^{0.3-0.1} [p_{00}(0.1+0+t) p_{01}(0.1-0+t) p_{01}(0.3-0-1-t) p_{00}(0.4-01-t)]dt
\]

is utilized between the maternal gamete for animal 3 and that for animal 5, in which the gametogenesis process from animal 4 to animal 5 is taken into account. We again note that the use of only flanking markers B and C results in a higher corresponding value:

\[
\frac{1}{(0.3-0.1)} \int_{0}^{0.3-0.1} [p_{00}^2(0.3-0-1-t) p_{00}(t) + p_{00}^2(0.3-0-1-t) p_{00}(t)]dt = 0.6507
\]

Table 2. Elements of gametic relationship matrices for the presented method (above the diagonal) and for the method using only the flanking markers B and C (below the diagonal)

<table>
<thead>
<tr>
<th>Animal no.</th>
<th>Animal 1</th>
<th>Animal 2</th>
<th>Animal 3</th>
<th>Animal 4</th>
<th>Animal 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>0.9886</td>
<td>0</td>
<td>0.0114</td>
<td>0.9886</td>
<td>0.0112</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.3344</td>
<td>0</td>
<td>0.3301</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9839</td>
<td>0.6656</td>
<td>0</td>
<td>0</td>
<td>0.6538</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.0879</td>
<td>0.9121</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.0879</td>
<td>0.9121</td>
<td>0.9121</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.9121</td>
<td>0.0879</td>
<td>0.9121</td>
<td>0.8442</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8442</td>
<td>0.0879</td>
<td>0.9121</td>
<td>0.8442</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.0879</td>
<td>0.9121</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.0879</td>
<td>0.9121</td>
<td>0.8442</td>
<td>0.1558</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.1558</td>
<td>0.8442</td>
<td>0.9121</td>
<td>0.1558</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0174</td>
</tr>
<tr>
<td></td>
<td>0.0174</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.0879</td>
<td>0.9121</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.0879</td>
<td>0.9121</td>
<td>0.8442</td>
<td>0.1558</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.1558</td>
<td>0.8442</td>
<td>0.9121</td>
<td>0.1558</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

All the elements of \( G \) used in the current method are listed in Table 2, and compared with those for the approach using only the flanking markers B and C.

The incidence matrix \( X \) is given as

\[
X = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1
\end{bmatrix}
\]
and we have $Z = I_5$ and $W = I_5 J_2$, where $I$ denotes the direct product. The additive relationship matrix among the five animals is of the form:

$$A_u = \begin{bmatrix}
1 & 0 & 0.5 & 0.5 & 0.75 \\
1 & 0.5 & 0.5 & 0.25 \\
1 & 0.5 & 0.5 & \\
\text{Sym.} & 1 & 0.75 \\
1.25
\end{bmatrix}$$

Then, by solving the mixed model equations (vii), we obtain the solutions 118.5087 and 99.0206 for $b_1$ and $b_2$, respectively, and 7.1733, −7.1733, −3.6670, 5.0937, and 4.5461 for $u_i$ for $i = 1, 2, ..., 5$, respectively. The BLUPs of the QTL-cluster effects are yielded as 3.1365 and −0.4869, −0.9714 and −1.6757, −0.4733 and −1.7001, 3.1369 and −0.8012, and −0.4577 and −0.7591 for $\hat{v}_i^0$ and $\hat{v}_i^1$ for $i = 1, 2, ..., 5$, respectively. In contrast to these, the approach using only the flanking markers B and C does not always utilize the informative data and does not extract as much information as possible on the segregation of the marked QTLs gives, for example, −1.4143 and −1.2144 for $v_2^0$ and $v_2^1$, and −0.0646 and −0.7491 for $v_5^0$ and $v_5^1$, respectively.

**Discussion**

Some studies suggest that some QTLs affecting certain quantitative traits may be a cluster of QTLs positioned in a particular chromosome region (e.g., KHAVKIN, COE 1997, 1998). For such situations, instead of one locus or a limited number of loci, consideration of a chromosomal segment itself would be a reasonable approach. The chromosomal segment is not a collection of independently segregating genes, and the proportion of genome shared IBD provides an ultimate measure of genetic resemblance at the chromosome level among relatives (DONELLY 1983, GOLDGAR 1990, GUO 1994a, b, 1995). Applying the ideas of GUO (1995) to the case with general pedigrees in outbred populations and using the mixed model methodology of HENDERSON (1973, 1975, 1984), MATSUDA and IWAIKSAK (1998) initially discussed BLUE of fixed effects pertaining to macroenvironmental factors and BLUP of random effects of the QTL-cluster and of the remaining QTLs, assuming the use of marker data. Their procedure is also an approach to predicting the effects of the chromosome segment marked, regardless of the number of QTLs contained in the segment. For situations in which QTLs more than two loci are located on a chromosome segment of 20 cM in length bracketed with two markers, MATSUDA and IWAIKSAKI (2000) further applied a deterministic sampling technique to evaluate the expected log-likelihood surfaces of a model, as used by MACKINNON et al. (1996), and showed that
the model fitting the QTL-cluster effects is expected to provide unbiased prediction in the cases where the single-QTL model leads to biased prediction.

The procedure presented herein is a generalization of the method of MATSUDA and IWAISAKI (1998), that takes into consideration the EIBDPs for the marked chromosome region in contrast to IBD probabilities at a specific location, to deal with the situation where data on multiple linked markers surrounding a cluster of QTLs are available. For the usual single-QTL model, it is shown that computation of IBD probabilities using information on multiple markers can essentially be similar to that using information on only the two flanking markers (e.g., VAN ARENDONK et al. 1998, PONG-WONG et al. 2001). However, when dealing with a particular chromosome segment, given multiple marker information, one cannot use the equations shown by MATSUDA and IWAISAKI (1998) straightforwardly. Then, in this study, taking into consideration the length of the segment of concern and the relative positions of multiple markers, we make it possible to calculate the EIBDPs.

When the IBD patterns at the flanking marker loci are unknown, the gain in information on the EIBDPs could be more substantial if more markers are typed (GUO 1994b). As demonstrated in the above example, the possibly reasonable values of the EIBDPs and consequently those of the elements of the gametic relationship matrix are used with the current method through incorporating the information of two informative markers. Hence, the use of the current method is theoretically expected to improve the accuracy of the prediction.

Application of the current method requires some knowledge of the position of the QTL-cluster and the related components of variances. The model (ii) assumed here can be used to estimate these parameters, for instance, by restricted maximum likelihood (PATTERSON, THOMPSON 1971). Some approaches for interval mapping assuming a single putative QTL via restricted maximum likelihood have already been developed for outbred populations (e.g., GRIGNOLA et al. 1996, GRIGNOLA et al. 1997, van ARENDONK et al. 1998, SAIT0, IWAISAKI 2000). With these procedures, if some QTLs of relatively large effects are not tightly linked to each other (20 cM apart) and each of the QTLs is always separated by two markers, then it would be possible to individually detect each QTL, conducting the intersection-union test (GRIGNOLA et al. 1997). In the situation where a cluster of tightly linked multiple QTLs is located in a marked chromosome region, however, the power of such interval mapping procedures would be inferior, and a wrong inference would possibly be made. Contrary to this, the current approach makes no assumptions on how many QTLs are contained in a tested chromosomal segment, and therefore it could be robust not only if there are many linked QTLs, but even if only a single QTL is located in the marked region. Moreover, it is known that relative to the usual interval mapping, multipoint mapping using information on multiple markers can increase the power of QTL detection (e.g., FULKER et al. 1995, ALMASY, BLANGER0 1998). It is also known that the use of all existing relationships increases the power of QTL detec-
tion and the accuracy of the estimates (ALMASY, BLANGERO 1998). Hence, the application of the current approach using data on informative markers of multiple linked ones and on all animals with general pedigrees would be of considerable value for detection of the segment containing QTL(s).

The inverse of the gametic relationship matrix is definitely necessitated in the current method. Because of this, the application of the direct inversion might be limited to populations of a relatively small size, even if it is possible to utilize sparse matrix techniques. For the case allowing a single QTL, some algorithms to efficiently calculate the gametic relationship matrix and its inverse have already been proposed (FERNANDO, GROSSMAN 1989, VAN ARENDONK et al. 1994, WANG et al. 1995). However, these algorithms cannot be applied directly to the current procedure because of the difference in nature between the IBD probability in the single-QTL case and the EIBDP in the current case, which are the elements of the corresponding gametic relationship matrices. On this point, we comment that for the prediction procedures of the present type we have presented a computing algorithm to set up the gametic relationship and the inverted matrices recursively by recursively constructing the IBD sets in equation (vii) (MATSUDA, IWAIISAKI 2001). According to the procedure, the IBD set \(D(i^h, j^{h'})\) between gamete \(h\) for individual \(i\) originated from its one parent \(s\) and gamete \(h'\) for individual \(j\) can be constructed recursively, as follows:

\[
D(i^h, j^{h'}) = [D(s^0, j^{h'}) \times D(s^0, i^h)] \cup [D(s^1, j^{h'}) \times D(s^1, i^h)],
\]

using the IBD sets between \(h\) and the paternal and the maternal gametes of \(s\), \(D(s^0, i^h)\) and \(D(s^1, i^h)\), and the IBD sets between \(h'\) and the paternal and the maternal gametes of \(j\), \(D(s^0, j^{h'})\) and \(D(s^1, j^{h'})\), where the notation ‘×’ indicates that the whole set \(D\) can be composed of lower dimensional, mutually exclusive, disjointed subsets. Note that since \(D(s^0, i^h)\) and \(D(s^1, i^h)\) are the IBD sets between the progeny and its one parent, these are the smallest units, \(\{0\}\) or \(\{1\}\). Then, using the information on the gametic relationship matrix \(G_{v,k}\) for gametes 1 to \(k - 1\) that is based on the recursive procedure above, it is possible to calculate the inverse, as follows:

\[
G_{v,k}^{-1} = \begin{bmatrix}
G_{v,k-1}^{-1} & 0 \\
0 & 0
\end{bmatrix} + (1 - s'_{k} G_{v,k-1} s_{k})^{-1} \begin{bmatrix}
s'_{k} s_{k} \\
-s'_{k} & 1
\end{bmatrix},
\]

where \(s_{k}\) is the vector of contribution by ancestral gametes and is obtained as \(G_{v,k-1}^{-1} G_{v,k}\). Hence, applying the computing procedure of MATSUDA and IWAIISAKI (2001), the current method can be used in outbred populations with complex pedigrees, as some other methods using the single-QTL model can be done.

The current method allows the situation where the parental linkage phases of markers are known. It is further necessary to adapt it to the situations with
the linkage phases between unknown markers. Application of Markov chain Monte Carlo (MCMC) approximation (e.g., GRIGNOLA et al. 1996, BINK, VAN ARENDONK 1999, PEREZ-ENCISO et al. 2000) would be useful for this purpose. The MCMC approach requires the sampling of each set of possible linkage phases for markers. Therefore, it seems that the use of the proposed procedure in each linkage phase sampled could be useful in efficiently making up the (inverted) gametic relationship matrix of concern. Moreover, PONG-WONG et al. (2001) recently proposed an interesting algorithm to effectively approximate the gametic relationship matrix for one marked QTL, which provides a smaller computational load compared to the MCMC approach. To calculate the IBD probabilities, this algorithm uses the information on the selected markers where the phases will be known with absolute certainty. It would be possible to apply a similar algorithm to the current method. At that time, in order to compute the EIBDPs among sibs produced by base parents whose linkage phases are not known, application of the approach by XU and GESSLER (1998) could be useful.

REFERENCES


