

Identification of gametes and treatment of linear dependencies in the gametic QTL-relationship matrix and its inverse

Armin TUCHSCHERER, Manfred MAYER, Norbert REINSCH*

Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich Genetik und Biometrie, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

(Received 29 December 2003; accepted 14 June 2004)

Abstract – The estimation of gametic effects *via* marker-assisted BLUP requires the inverse of the conditional gametic relationship matrix \mathbf{G} . Both gametes of each animal can either be identified (distinguished) by markers or by parental origin. By example, it was shown that the conditional gametic relationship matrix is not unique but depends on the mode of gamete identification. The sum of both gametic effects of each animal – and therefore its estimated breeding value – remains however unaffected. A previously known algorithm for setting up the inverse of \mathbf{G} was generalized in order to eliminate the dependencies between columns and rows of \mathbf{G} . In the presence of dependencies the rank of \mathbf{G} also depends on the mode of gamete identification. A unique transformation of estimates of QTL genotypic effects into QTL gametic effects was proven to be impossible. The properties of both modes of gamete identification in the fields of application are discussed.

marker assisted selection / best linear unbiased prediction / linkage analysis / gametic relationship matrix

1. INTRODUCTION

Fernando and Grossman [2] described how to incorporate genetic markers linked to quantitative trait loci (QTL) into best linear unbiased prediction (BLUP) for genetic evaluation. For this, the inverse of the conditional gametic relationship matrix \mathbf{G} is needed. This matrix mirrors the (co-)variances between QTL allele effects of all animals for a marked QTL (MQTL).

For offspring of so-called informative matings the paternal or maternal origin of gametes can be identified by one or several markers in the surroundings of the QTL. The QTL-allele on the paternal (maternal) gamete can then be

* Corresponding author: reinsch@fhn-dummerstorf.de

taken as the first (second) MQTL-allele effect of such an individual. Below this is termed “gamete identification by parental origin”.

An alternative mode of gamete identification has been employed by Wang *et al.* [21] and Abdel-Azim and Freeman [1]: for an individual with a heterozygous (1, 2) marker genotype, the gamete with the first (1, in alphanumerical order) marker allele is taken to carry the first and the gamete with the other (2) allele, the second MQTL allele effect. This is denoted as “gamete identification by markers”.

Both modes of gamete identification have been used before in publications dealing with the computation of \mathbf{G} and its inverse from pedigrees and marker data. Until now – to the authors’ knowledge – the consequences of changing the mode of gamete identification in a marker assisted BLUP (MA-BLUP) model have, however, not been investigated.

Abdel-Azim and Freeman [1] – based on the results of [2] and [21] – developed a numerically efficient algorithm for the computation of \mathbf{G} and its inverse. This algorithm has been tailored for situations where \mathbf{G} has full row and column rank and the number of MQTL effects is twice the number of animals in the pedigree. However, under certain circumstances, linear dependencies may occur between gametic MQTL effects and \mathbf{G} may therefore be rank-deficient. This could *e.g.* arise from a microsatellite located within an intron (zero recombination rate) of that gene, which is responsible for the QTL or if double recombinants are ignored for a QTL between two flanking markers [10].

This article first demonstrates by example that \mathbf{G} is not unique but depends on the mode of gamete identification, and as do the MA-BLUP estimates of gametic MQTL effects. Then a generalization of the Abdel-Azim and Freeman algorithm [1] is developed, allowing for the elimination of linear dependencies in \mathbf{G} and its inverse.

2. MODEL, NOTATION, DEFINITIONS, ASSUMPTIONS

Let us consider the following mixed linear model (gametic effects model)

$$\mathbf{y} = \mathbf{Xf} + \mathbf{Zu} + \mathbf{ZTv} + \mathbf{e}, \quad (1)$$

where $\mathbf{y}_{(m \times 1)}$ denotes the vector of m phenotypic records for n animals, $\mathbf{f}_{(n_f \times 1)}$ is the vector of fixed effects, $\mathbf{u}_{(n \times 1)}$ is the vector of random polygenic effects and $\mathbf{v}_{(2n \times 1)}$ is the vector of the random gametic effects $(v_1^1, v_1^2, \dots, v_i^1, v_i^2, \dots, v_n^1, v_n^2)'$ of a marked quantitative trait locus (MQTL) that is linked to a single polymorphic marker locus (ML). Linkage equilibrium between ML and MQTL is assumed. Observed marker genotypes are denoted

by \mathbf{M} . $\mathbf{X}_{(m \times n_f)}$, $\mathbf{Z}_{(m \times n)}$ are known incidence matrices and $\mathbf{T}_{(n \times 2n)} = \mathbf{I}_n \otimes [1 \ 1]$, where \otimes stands for the Kronecker product. Subscripts in parentheses of the vectors and matrices denote their dimensions. Expectations of \mathbf{u} , \mathbf{v} and \mathbf{e} and covariances between them are assumed to be 0. Furthermore, let $\text{Cov}(\mathbf{u}) = \sigma_u^2 \mathbf{V}$, $\text{Cov}(\mathbf{v}) = \sigma_v^2 \mathbf{G}$, $\text{Cov}(\mathbf{e}) = \sigma_e^2 \mathbf{R}$, with the $(n \times n)$ -dimensional numerator relationship matrix \mathbf{V} , the $(m \times m)$ -dimensional residual covariance matrix \mathbf{R} and the $(2n \times 2n)$ -dimensional conditional gametic relationship matrix \mathbf{G} and the variance components σ_u^2 , σ_v^2 and σ_e^2 of the polygenic effects, the effects of the MQTL and the residual effects.

Let α_i^1, α_i^2 , $i = 1, \dots, n$ denote the two MQTL alleles of individual i having the additive effects $\mathbf{v}_i = (v_i^1, v_i^2)'$, and $P(\alpha_i^k \leftarrow \alpha_j^t | \mathbf{M})$ defines the probability that the k th allele, $k = 1, 2$, of individual i descends from the t th allele α_j^t , $t = 1, 2$, of parent j given the observed marker genotypes \mathbf{M} , and, r is the recombination rate between the maker locus and the MQTL. In the following paragraphs let us assume that individuals are ordered such that parents precede their progeny (ordered pedigree).

3. COMPUTING G AND ITS INVERSE

Abdel-Azim's and Freeman's example [1] is used to demonstrate that \mathbf{G} and its inverse are not unique but depend on the mode of gamete identification. With the assumptions made above and a recombination rate $r > 0$, gamete identification by markers is considered first.

3.1. Gametes are identified by markers

Let s and d denote paternal and maternal parents of animal i . The eight probabilities that the MQTL alleles (α_i^1, α_i^2) of animal i descended from any of the parents' four MQTL alleles, paternal (α_s^1, α_s^2) and maternal (α_d^1, α_d^2) , for given observed marker genotypes \mathbf{M} , can be written as a matrix \mathbf{Q}_i as defined by Wang *et al.* [21]:

$$\mathbf{Q}_i = \begin{bmatrix} P(\alpha_i^1 \leftarrow \alpha_s^1 | \mathbf{M}) P(\alpha_i^1 \leftarrow \alpha_s^2 | \mathbf{M}) P(\alpha_i^1 \leftarrow \alpha_d^1 | \mathbf{M}) P(\alpha_i^1 \leftarrow \alpha_d^2 | \mathbf{M}) \\ P(\alpha_i^2 \leftarrow \alpha_s^1 | \mathbf{M}) P(\alpha_i^2 \leftarrow \alpha_s^2 | \mathbf{M}) P(\alpha_i^2 \leftarrow \alpha_d^1 | \mathbf{M}) P(\alpha_i^2 \leftarrow \alpha_d^2 | \mathbf{M}) \end{bmatrix}. \quad (2a)$$

It must be defined what is the first and what is the second MQTL allele in (2a): in heterozygotes (1,2 at the marker) the first MQTL allele is on the gamete with the first marker allele (1) and the second MQTL allele is on the gamete with the second marker allele (2), as already described in the introduction.

In homozygotes, the MQTL alleles can not be distinguished. The \mathbf{Q}_i for the base animals, *i.e.* animals having no parents in the pedigree, are not defined. Non-base animals have \mathbf{Q}_i s with first and the second row sums equal to one as well as the sum of the elements of the sire block (first two columns of \mathbf{Q}_i) and the sum of the elements of the dam block (last two columns of \mathbf{Q}_i).

The \mathbf{Q}_i matrices are of key importance, because once these \mathbf{Q}_i s have been computed for all individuals in an ordered pedigree, the tabular method [21] can be applied for the construction of \mathbf{G} and \mathbf{G}^{-1} – no matter what method has been used for the computation of \mathbf{Q}_i s before:

$$\mathbf{G}_1 = \mathbf{C}_{11} = \mathbf{I}_2 \quad \text{and} \quad \mathbf{G}_i = \left[\begin{array}{c|c} \mathbf{G}_{i-1} & \mathbf{G}_{i-1}\mathbf{A}'_i \\ \hline \mathbf{A}_i\mathbf{G}_{i-1} & \mathbf{C}_{ii} \end{array} \right],$$

$$\text{with } \mathbf{C}_{ii} = \begin{bmatrix} 1 & f_i \\ f_i & 1 \end{bmatrix}, i = 2, \dots, n, \quad (3)$$

where f_i is the conditional probability that 2 homologous alleles at the MQTL in individual i are identical by decent, given observed marker genotypes \mathbf{M} (conditional inbreeding coefficient of individual i for the MQTL, given \mathbf{M}), which can be calculated according to formula (11) in [21], and

$$\mathbf{G}_1^{-1} = (\mathbf{G}_1)^{-1} = \mathbf{I}_2 \quad \text{and} \quad \mathbf{G}_i^{-1} = \left[\begin{array}{c|c} \mathbf{G}_{i-1}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] + \left[\begin{array}{c|c} \mathbf{A}'_i\mathbf{D}_i^{-1}\mathbf{A}_i & -\mathbf{A}'_i\mathbf{D}_i^{-1} \\ \hline -\mathbf{D}_i^{-1}\mathbf{A}_i & \mathbf{D}_i^{-1} \end{array} \right],$$

$$\text{with } \mathbf{D}_i = (\mathbf{C}_{ii} - \mathbf{A}_i\mathbf{G}_{i-1}\mathbf{A}'_i), i = 2, \dots, n. \quad (4)$$

\mathbf{A}_i is a $(2 \times 2[i-1])$ -dimensional matrix constructed by setting the $(2s-1)$ th and $(2s)$ th column equal to the first and second column of \mathbf{Q}_i and the $(2d-2)$ th and $(2d)$ th column equal to the third and fourth column of \mathbf{Q}_i , all other elements of \mathbf{A}_i are zero, where s and d are the numbers of the sire and the dam of individual i in the ordered pedigree.

Abdel-Azim and Freeman [1] gave an algorithm for the decomposition of \mathbf{G} by $\mathbf{G} = \mathbf{BDB}'$, where \mathbf{B} is a lower triangular matrix and \mathbf{D} is a block diagonal matrix with (2×2) -matrices \mathbf{D}_i from (4) in the i th block. \mathbf{B} can be recursively computed as

$$\mathbf{B}_1 = \mathbf{I}_2 \quad \text{and} \quad \mathbf{B}_i = \begin{bmatrix} \mathbf{B}_{i-1} & \mathbf{0} \\ \mathbf{A}_i\mathbf{B}_{i-1} & \mathbf{I}_2 \end{bmatrix}, i = 2, \dots, n, \quad (5)$$

where \mathbf{I}_2 is an identity matrix and \mathbf{A}_i is the same matrix as in (3) and (4). The inverse of \mathbf{G} can be calculated as $\mathbf{G}^{-1} = (\mathbf{B}')^{-1}\mathbf{D}^{-1}\mathbf{B}^{-1}$, with

Table I. Example pedigree, marker genotypes from [1] and \mathbf{Q}_i^* (bold numbers) from (2b), in \mathbf{Q}_i notation (2a).

Animal (i)	Sire (s)	Dam (d)	Marker genotype	\mathbf{Q}_i^* in \mathbf{Q}_i notation (2a) (recombination rate: $r = 0.1$)			
1	0	0	A ₁ A ₁	-	-	-	-
2	0	0	A ₂ A ₂	-	-	-	-
3	0	0	A ₁ A ₂	-	-	-	-
4	1	2	A ₁ A ₂	0.50	1 - 0.50	0.00	0.00
				0.00	0.00	0.50	1 - 0.50
5	3	4	A ₁ A ₁	0.50	1 - 0.50	0.00	0.00
				0.00	0.00	0.90	1 - 0.90
6	1	4	A ₁ A ₂	0.50	1 - 0.50	0.00	0.00
				0.00	0.00	0.10	1 - 0.10
7	5	6	A ₁ A ₂	0.50	1 - 0.50	0.00	0.00
				0.00	0.00	0.10	1 - 0.10

$\mathbf{D}^{-1} = \text{diag}(\mathbf{D}_1^{-1}, \dots, \mathbf{D}_n^{-1})$ and recursively computed \mathbf{B}^{-1} :

$$\mathbf{B}_1^{-1} = \mathbf{I}_2 \quad \text{and} \quad \mathbf{B}_i^{-1} = \begin{bmatrix} \mathbf{B}_{i-1}^{-1} & \mathbf{0} \\ -\mathbf{A}_i & \mathbf{I}_2 \end{bmatrix}, i = 2, \dots, n, \tag{6}$$

[1] proposed efficient computational techniques using this decomposition and a sparse storage scheme for \mathbf{G}^{-1} .

$\mathbf{G}^{-1} = (\mathbf{B}')^{-1} \mathbf{D}^{-1} \mathbf{B}^{-1}$ can be computed if and only if the (2×2) -matrices \mathbf{D}_i^{-1} exist for each individual i ($i = 1, \dots, n$), that means all determinants $\det(\mathbf{D}_i) \neq 0$.

The example of Abdel-Azim and Freeman (see Tab. I in [1]) can be used to demonstrate \mathbf{G} (Fig. 1 in [1]) and \mathbf{G}^{-1} (p. 162 in [1]) for complete marker data, linkage equilibrium and a recombination rate of 0.10 under gamete identification by markers.

3.2. Gametes are identified by parental origin of the marker alleles

When the gametes α_i^1, α_i^2 are identified by the parental origin of the marker alleles, the first MQTL allele of animal i is defined as its paternal ($\alpha_i^1 =_{\text{def}} \alpha_i^s$) and the second as its maternal allele ($\alpha_i^2 =_{\text{def}} \alpha_i^d$). Consequently (2a) becomes

$$\mathbf{Q}_i = \begin{bmatrix} P(\alpha_i^s \leftarrow \alpha_s^s | \mathbf{M}) & P(\alpha_i^s \leftarrow \alpha_s^d | \mathbf{M}) & 0 & 0 \\ 0 & 0 & P(\alpha_i^d \leftarrow \alpha_d^s | \mathbf{M}) & P(\alpha_i^d \leftarrow \alpha_d^d | \mathbf{M}) \end{bmatrix},$$

and with the fact that the row sums of \mathbf{Q}_i are equal to 1

$$P(\alpha_i^s \leftarrow \alpha_s^d | \mathbf{M}) = 1 - P(\alpha_i^s \leftarrow \alpha_s^s | \mathbf{M})$$

and

$$P(\alpha_i^d \leftarrow \alpha_d^d | M) = 1 - P(\alpha_i^d \leftarrow \alpha_d^s | M),$$

i.e. only two parameters $P(\alpha_i^s \leftarrow \alpha_s^s | M)$ and $P(\alpha_i^d \leftarrow \alpha_d^s | M)$ are to be calculated and therefore \mathbf{Q}_i reduces to

$$\mathbf{Q}_i^* = \left(P(\alpha_i^s \leftarrow \alpha_s^s | M) \quad P(\alpha_i^d \leftarrow \alpha_d^s | M) \right)' = \left(\mathbf{Q}_i^{*1} \quad \mathbf{Q}_i^{*2} \right)'. \quad (2b)$$

\mathbf{Q}_i^{*1} and \mathbf{Q}_i^{*2} are known as transition probabilities in QTL analysis.

In contrast to gamete identification by markers (2a), the gametes of base animals cannot be uniquely identified and the paternal or maternal origin of the marker alleles of all base animals remains uncertain when (2b) is applied. With a probability of 0.5 the first marker allele may be of paternal or maternal origin, and the second, too. This fact creates differences in the \mathbf{Q}_i matrices and, as a consequence, differences in \mathbf{G} and its inverse if gamete identification by parental origin is used. The same is true for heterozygous offspring of uninformative matings. For illustration, let us consider animal 5 in Table I in [1] and Table I of this paper. Animal 5 has a marker genotype A_1A_1 and is offspring of animal 3 (sire, A_1A_2) and animal 4 (dam, A_1A_2). It is evident that animal 5 has inherited A_1 from both parents. With definition (2a), this is the first allele of the sire and the first of the dam, but because of the homozygosity, each of the A_1 in animal 5, A_1 can be the first or the second marker allele. Thus under (2a), \mathbf{Q}_5 must be determined as

$$\mathbf{Q}_5 = \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1-r & r & 0 & 0 \\ r & 1-r & 0 & 0 \\ 0 & 0 & 1-r & r \\ 0 & 0 & r & 1-r \end{bmatrix} = \begin{bmatrix} 0.45 & 0.05 & 0.45 & 0.05 \\ 0.45 & 0.05 & 0.45 & 0.05 \end{bmatrix},$$

where the first matrix of the product is the matrix with the probabilities of descent for the marker alleles and the second is the matrix with the recombination rate $r = 0.1$ in both formulas for \mathbf{Q}_5 .

Now we use definition (2b), and the fact that the sire of 5 is base animal 3. Hence in individual 3 A_1 can be maternal or paternal with probability 0.5. The dam of animal 5 is no base animal. So it is clear that A_1 is the paternal allele of the dam, and

$$\mathbf{Q}_5 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1-r & r & 0 & 0 \\ r & 1-r & 0 & 0 \\ 0 & 0 & 1-r & r \\ 0 & 0 & r & 1-r \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 \end{bmatrix}$$

or in (2b) notation $\mathbf{Q}_5^* = (0.5 \ 0.9)'$.

The complete set of \mathbf{Q}_i^* s (2b) in their \mathbf{Q}_i notation (2a) for Table I data in [1] for gamete identification by parental origin can be found in Table I.

With \mathbf{Q}_i -notation of the \mathbf{Q}_i^* the algorithm of [21] and [1] can also be applied for computing the conditional gametic relationship matrix (non-zero elements of this matrix see (E 1) and its inverse (non-zero elements of the inverse see (E 2)).

1.00	-	-	-	-	-	.500	-	-	.450	.500	.050	.225	.095
-	1.00	-	-	-	-	.500	-	-	.450	.500	.050	.225	.095
-	-	1.00	-	-	-	.500	-	-	.050	-	.450	.025	.405
-	-	-	1.00	-	-	.500	-	-	.050	-	.450	.025	.405
-	-	-	-	1.00	-	-	-	.500	-	-	-	.250	-
-	-	-	-	-	1.00	-	-	.500	-	-	-	.250	-
.500	.500	-	-	-	-	1.00	-	-	.900	.500	.100	.450	.140
-	-	.500	.500	-	-	-	1.00	-	.100	-	.900	.050	.810
-	-	-	-	.500	.500	-	-	1.00	-	-	-	.500	-
.450	.450	.050	.050	-	-	.900	.100	-	1.00	.450	.180	.500	.207
.500	.500	-	-	-	-	.500	-	-	.450	1.00	.050	.225	.145
.050	.050	.450	.450	-	-	.100	.900	-	.180	.050	1.00	.090	.905
.225	.225	.025	.025	.250	.250	.450	.050	.500	.500	.225	.090	1.00	.1035
.095	.095	.405	.405	-	-	.140	.810	-	.207	.145	.905	.1035	1.00

(E 1)

2.000	1.000	-	-	-	-	-1.000	-	-	-	-1.000	-	-	-
1.000	2.000	-	-	-	-	-1.000	-	-	-	-1.000	-	-	-
-	-	1.500	.500	-	-	-	-1.000	-	-	-	-	-	-
-	-	.500	1.500	-	-	-	-1.000	-	-	-	-	-	-
-	-	-	-	1.500	.500	-	-	-1.000	-	-	-	-	-
-	-	-	-	.500	1.500	-	-	-1.000	-	-	-	-	-
-1.000	-1.000	-	-	-	-	6.556	1.000	-	-5.000	-	-5.56	-	-
-	-	-1.000	-1.000	-	-	1.000	6.556	-	-5.56	-	-5.000	-	-
-	-	-	-	-1.000	-1.000	-	-	2.500	.500	-	-	-1.000	-
-	-	-	-	-	-	-5.000	-5.56	.500	6.056	-	-	-1.000	-
-1.000	-1.000	-	-	-	-	-	-	-	-	2.058	.526	-	-5.85
-	-	-	-	-	-	-5.56	-5.000	-	-	.526	10.292	-	-5.263
-	-	-	-	-	-	-	-	-1.000	-1.000	-	-	2.000	-
-	-	-	-	-	-	-	-	-	-	-5.85	-5.263	-	5.848

(E 2)

Comparing Figure 1 in [1] and (E 1) or the matrix at page 162 in [1] and (E 2), there are some differences in \mathbf{G} and \mathbf{G}^{-1} . The \mathbf{G} -matrix [1] is of full rank and has 128 non-zero elements, \mathbf{G} in (E 1) is of full rank, too, but it only has 106 non-zero elements. The numbers of non-zeros in the corresponding inverses are 74 (p. 162 in [1]) versus 58 (E 2).

With the $\mathbf{w} = \mathbf{T}\mathbf{v}$, model (1) can be written as MQTL genotypic effects of model $\mathbf{y} = \mathbf{X}\mathbf{f} + \mathbf{Z}\mathbf{u} + \mathbf{Z}\mathbf{w} + \mathbf{e}$, with $(n \times 1)$ -vector \mathbf{w} of genotypic effects at the MQTL of the n animals, $E(\mathbf{w}) = \mathbf{0}$, $\text{Cov}(\mathbf{w}) = \sigma_w^2 \mathbf{Q}_{(n \times n)}^G$ with $\sigma_w^2 = 2\sigma_v^2$. It turns out that the relation $\sigma_v^2 \cdot \mathbf{Q}_{(n \times n)}^G = \sigma_v^2 \cdot 0.5 \cdot \mathbf{T}_{(n \times 2n)} \mathbf{G}_{(2n \times 2n)} \mathbf{T}'_{(2n \times n)}$ leads

to the same conditional genotypic relationship matrix [19] (non-zero elements in (E 3))

$$\begin{bmatrix} 1.0000 & - & - & 0.5000 & 0.4500 & 0.5500 & 0.3200 \\ - & 1.0000 & - & 0.5000 & 0.0500 & 0.4500 & 0.4300 \\ - & - & 1.0000 & - & 0.5000 & - & 0.2500 \\ 0.5000 & 0.5000 & - & 1.0000 & 0.5000 & 0.7500 & 0.7250 \\ 0.4500 & 0.0500 & 0.5000 & 0.5000 & 1.0000 & 0.3150 & 0.6035 \\ 0.5500 & 0.4500 & - & 0.7500 & 0.3150 & 1.0500 & 0.6825 \\ 0.3200 & 0.4300 & 0.2500 & 0.7250 & 0.6035 & 0.6825 & 1.1035 \end{bmatrix} \quad (\text{E } 3)$$

for both different conditional gametic relationship matrices Figure 1 in [1] and (E 1). As a consequence the resulting genotypic effects \mathbf{w} are independent of the variant of \mathbf{G} and the same is true for polygenic effects and the total breeding values of all animals.

4. LINEAR DEPENDENCIES IN \mathbf{G} AND RULES FOR ELIMINATING THEM

As already mentioned, the recombination rate r between MQTL and the marker may be zero for certain applications. Therefore we re-examine the example from Table I in [1] using gamete identification by markers, but now with a recombination rate of $r = 0$. The corresponding \mathbf{Q}_i s can be found in Table II.

With the Abdel-Azim and Freeman algorithm [1] the \mathbf{G} -matrix can be calculated, but it has dependent rows and columns (*e.g.* identical rows/columns 8, 12 and 14, see (E 4)).

$$\begin{bmatrix} 1.00 & - & - & - & - & - & .50 & - & .25 & .25 & .50 & - & .25 & - \\ - & 1.00 & - & - & - & - & .50 & - & .25 & .25 & .50 & - & .25 & - \\ - & - & 1.00 & - & - & - & .50 & - & - & - & .50 & - & .50 & - \\ - & - & - & 1.00 & - & - & .50 & - & - & - & .50 & - & .50 & - \\ - & - & - & - & 1.00 & - & - & - & - & - & - & - & - & - \\ .50 & .50 & - & - & - & 1.00 & - & .50 & .50 & .50 & - & .50 & - & - \\ - & - & .50 & .50 & - & - & 1.00 & - & - & - & 1.00 & - & 1.00 & - \\ .25 & .25 & - & - & .50 & - & .50 & - & 1.00 & - & .25 & - & .50 & - \\ .25 & .25 & - & - & .50 & - & .50 & - & - & 1.00 & .25 & - & .50 & - \\ .50 & .50 & - & - & - & - & .50 & - & .25 & .25 & 1.00 & - & .25 & - \\ - & - & .50 & .50 & - & - & 1.00 & - & - & - & 1.00 & - & 1.00 & - \\ .25 & .25 & - & - & .50 & - & .50 & - & .50 & .50 & .25 & - & 1.00 & - \\ - & - & .50 & .50 & - & - & 1.00 & - & - & - & 1.00 & - & 1.00 & - \end{bmatrix} \quad (\text{E } 4)$$

The computation of \mathbf{G}^{-1} fails because of the dependencies in \mathbf{G} . These dependencies are indicated by $\det(\mathbf{D}_i) = 0$ for individuals $i = 5, 6, 7$, and consequently, \mathbf{D}_i^{-1} in (4) or (6) does not exist for these individuals. The dependencies in \mathbf{G} are caused by the configuration of \mathbf{Q}_i s. Problem-creating \mathbf{Q}_i -matrices in the example are \mathbf{Q}_5 , \mathbf{Q}_6 and \mathbf{Q}_7 in Table II. \mathbf{Q}_6 and \mathbf{Q}_7 imply

Table II. Example pedigree, marker genotypes from [1] and \mathbf{Q}_i (recombination rate: $r = 0.0$).

Animal (i)	Sire (s)	Dam (d)	Marker genotype	\mathbf{Q}_i according (2a)				For calculation of (E 6), (E 7):	
								f_i	$\widehat{\mathbf{D}}_i^*$
1	0	0	A_1A_1	-	-	-	-	0	\mathbf{I}_2
2	0	0	A_2A_2	-	-	-	-	0	\mathbf{I}_2
3	0	0	A_1A_2	-	-	-	-	0	\mathbf{I}_2
4	1	2	A_1A_2	0.50	0.50	0.00	0.00	0	$\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$
5	3	4	A_1A_1	0.50	0.00	0.50	0.00	-	-
6	1	4	A_1A_2	0.50	0.50	0.00	0.00	-	0.5
7	5	6	A_1A_2	0.50	0.50	0.00	0.00	-	0.5
				0.00	0.00	0.00	1.00		

that the second MQTL-alleles of individuals 6 and 7 are identical with the second MQTL-alleles of their dams, *i.e.* animals 4, 6 and 7 have identical second MQTL-alleles and this results in identical effects $v_4^2 = v_6^2 = v_7^2$ in model (1). \mathbf{Q}_5 contains the information that animal 5 has received the sire’s first MQTL-allele and the dam’s first MQTL-allele, but it is not known which of these alleles is the first and which is the second in animal 5. Therefore \mathbf{Q}_5 can be written as the average $\mathbf{Q}_5 = 0.5 \cdot \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \right]$, and the corresponding effects as $v_5^1 = v_5^2 = 0.5 \cdot (v_3^1 + v_4^1)$. Hence the number of gametic effects in model (1) can be reduced to a smaller set of different effects without dependencies in a corresponding ‘condensed’ gametic relationship matrix \mathbf{G}^* . How the configuration of the \mathbf{Q}_i s can be used in a ‘condensing’ algorithm for the gametic effects and the computing of the ‘condensed’ gametic relationship matrix \mathbf{G}^* and its inverse is outlined in detail in the following section.

Let \mathbf{v}^* denote the n^* -dimensional vector of the n^* remaining components of \mathbf{v} and let \mathbf{L} be a $(2n \times n^*)$ -dimensional matrix with row sums equal to 1 in such a manner that $\mathbf{v} = \mathbf{L}\mathbf{v}^*$. Therewith, model (1) can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{f} + \mathbf{Z}\mathbf{u} + \mathbf{Z}\mathbf{T} \cdot \mathbf{L}\mathbf{v}^* + \mathbf{e},$$

with $E(\mathbf{v}^*) = \mathbf{0}$ and $Cov(\mathbf{v}^*) = \sigma_v^2 \mathbf{G}^*$. The determination of the n^* remaining components of \mathbf{v} is part of the condensing algorithm. It is assumed that the \mathbf{Q}_i

matrices (2a) have already been computed for all animals and the pedigree is ordered such that parents precede their progeny.

Let further $\mathbf{SQ}_i = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \mathbf{Q}_i = \begin{bmatrix} \text{SQ}_i^1 & \text{SQ}_i^2 & \text{SQ}_i^3 & \text{SQ}_i^4 \end{bmatrix}$ define the (1×4) -vector of the column sums of \mathbf{Q}_i . $\text{SQ}_i^1 = 1$ for example means that animal i has received the first MQTL-allele of its sire and therefore $\text{SQ}_i^2 = 0$. If there is a one in the first or second row of the first column of \mathbf{Q}_i the place of this allele in i is the number of that row containing the one.

Define $\mathbf{N} = ((\mathbf{N}_{i,j}))$, $i = 1, \dots, n$; $j = 1, 2$ a $(n \times 2)$ -dimensional integer matrix with the indices of the remaining gametic effects \mathbf{v}^* of n animals and $\mathbf{N}_i = (\mathbf{N}_{i,1}; \mathbf{N}_{i,2})$ the i th row of \mathbf{N} and let n_b be the number of base animals at the top of the pedigree which are considered to be unrelated and non inbred, and $n_{i-1}^{\max} = \max_{\substack{j=1, \dots, i-1 \\ k=1, 2}} \{N_{j,k}\}$.

The algorithm consists of four parts: the generation of the index matrix \mathbf{N} , the determination of matrix \mathbf{L} , the calculation of the condensed gametic relationship matrix \mathbf{G}^* , and finally, the computation of its inverse. It is independent of the mode of gamete identification and can be used with \mathbf{Q}_i definition (2a) as well as with \mathbf{Q}_i definition (2b).

First part of the algorithm: Generation of the index matrix N

For $i \leq n_b$ (base animals):

$$\mathbf{N}_i = (2i - 1; 2i). \quad (7a)$$

For $i > n_b$ (non base animals) and $k, j = 1, 2$:

$$\mathbf{N}_i = \begin{cases} (\mathbf{N}_{s(i),k}; \mathbf{N}_{d(i),j}) & ; \text{if } \langle \mathbf{Q}_i(1, k) = 1 \wedge \mathbf{Q}_i(2, j + 2) = 1 \rangle \\ & \vee \langle \text{SQ}_i^k = 1 \wedge \text{SQ}_i^{2+j} = 1 \rangle \\ (\mathbf{N}_{d(i),j}; \mathbf{N}_{s(i),k}) & ; \text{if } \langle \mathbf{Q}_i(2, k) = 1 \wedge \mathbf{Q}_i(1, j + 2) = 1 \rangle \\ (\mathbf{N}_{s(i),k}; n_{i-1}^{\max} + 1) & ; \text{if } \langle \mathbf{Q}_i(1, k) = 1 \wedge \text{SQ}_i^{2+j} \neq 1, \forall j \rangle \\ (\mathbf{N}_{d(i),j}; n_{i-1}^{\max} + 1) & ; \text{if } \langle \mathbf{Q}_i(1, 2 + j) = 1 \wedge \text{SQ}_i^k \neq 1, \forall k \rangle \\ (n_{i-1}^{\max} + 1; \mathbf{N}_{s(i),k}) & ; \text{if } \langle \mathbf{Q}_i(2, k) = 1 \wedge \text{SQ}_i^{2+j} \neq 1, \forall j \rangle \\ (n_{i-1}^{\max} + 1; \mathbf{N}_{d(i),j}) & ; \text{if } \langle \mathbf{Q}_i(2, 2 + j) = 1 \wedge \text{SQ}_i^k \neq 1, \forall k \rangle \\ (n_{i-1}^{\max} + 1; n_{i-1}^{\max} + 2) & ; \text{if } \langle \text{SQ}_i^k \neq 1, \forall k \wedge \text{SQ}_i^{2+j} \neq 1, \forall j \rangle \end{cases} \quad (7b)$$

where $\mathbf{N}_{s(i),k}$ is the index of the k th MQTL-allele ($k = 1, 2$) of the sire $s(i)$ and $\mathbf{N}_{d(i),j}$ is the index of the j th MQTL-allele ($j = 1, 2$) of the dam $d(i)$ of

Table III. Example from Table II – computation of index matrix **N**.

Animal	Sire	Dam	$N_{i,1}$	$N_{i,2}$	n_i^{\max}	Remarks	
(i)	(s)	(d)					
1	0	0	1	2	2	} base animals	
2	0	0	3	4	4		
3	0	0	5	6	6		
4	1	2	7	8	8	$SQ_4^1 = SQ_4^2 = SQ_4^3 = SQ_4^4 = 0.5 \neq 1$	} non-base animals
5	3	4	5	7	8	$SQ_5^1 = 1 \wedge SQ_5^3 = 1$	
6	1	4	9	8	9	$Q_6(2,4) = 1 \wedge SQ_6^1 = SQ_6^2 = 0.5 \neq 1$	
7	5	6	10	8	-	$Q_7(2,4) = 1 \wedge SQ_7^1 = SQ_7^2 = 0.5 \neq 1$	

animal i , $Q_i(o, t)$ ($o = 1, 2; t = 1, \dots, 4$) denotes the t th element of the o th row in Q_i , ‘ \wedge ’/‘ \vee ’ are the logical ‘and’/‘or’ and ‘ \forall ’ is used in the meaning ‘for all’.

The computation of **N** is demonstrated with the example from Table II. Let us consider animal 4. Animal 4 is a non-base animal. Hence (7b) must be used to calculate $N_{4,1}$ and $N_{4,2}$.

$Q_4 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$ and thus all four column sums of Q_4 are equal to $0.5 \neq 1$, and therefore, $N_4 = (n_3^{\max} + 1; n_3^{\max} + 2) = (7; 8)$ where $n_3^{\max} = 6$. For the complete **N** see Table III.

Second part of the algorithm: Determination of the incidence matrix L

For each animal i ($i = 1, \dots, n$) there are two rows in **L**. Let $L_{2i-1,t}$ denote the elements of the first row and $L_{2i,t}$ ($t = 1, \dots, n^*$) those of the second. The following algorithm determines the non zero elements of **L**:

$$L_{2i-1,t} = \begin{cases} \begin{Bmatrix} Q_i(1,k) \\ Q_i(1,j) \end{Bmatrix} & ; \text{if } \langle SQ_i^k = 1 \wedge SQ_i^j = 1 \rangle \wedge \langle 1 \notin Q_i \rangle \text{ for } \begin{cases} t = N_{i,1} \\ t = N_{i,2} \end{cases} \\ 1 & ; \text{else for } t = N_{i,1} \end{cases} \tag{8}$$

and

$$L_{2i,t} = \begin{cases} \begin{Bmatrix} Q_i(2,k) \\ Q_i(2,j) \end{Bmatrix} & ; \text{if } \langle SQ_i^k = 1 \wedge SQ_i^j = 1 \rangle \wedge \langle 1 \notin Q_i \rangle \text{ for } \begin{cases} t = N_{i,1} \\ t = N_{i,2} \end{cases} \\ 1 & ; \text{else for } t = N_{i,2} \end{cases}$$

where $1 \notin Q_i$ means that no element of Q_i equals one.

For demonstration, Table II data and index matrix **N** from Table III are used. The complete **L**-Matrix, **v***, **v** and **T · v** for this example are

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{v}^* = \begin{bmatrix} v_1^* \\ \vdots \\ v_{10}^* \end{bmatrix}, \mathbf{v} = \mathbf{L}\mathbf{v}^* = \begin{bmatrix} v_1^* \\ v_2^* \\ v_3^* \\ v_4^* \\ v_5^* \\ v_6^* \\ v_7^* \\ v_8^* \\ \hline 0.5(v_5^* + v_7^*) \\ 0.5(v_5^* + v_7^*) \\ \hline v_9^* \\ v_8^* \\ v_{10}^* \\ v_8^* \end{bmatrix},$$

$$\mathbf{T}\mathbf{v} = \begin{bmatrix} v_1^* + v_2^* \\ v_3^* + v_4^* \\ v_5^* + v_6^* \\ v_7^* + v_8^* \\ \hline v_5^* + v_7^* \\ \hline v_9^* + v_8^* \\ v_{10}^* + v_8^* \end{bmatrix} \quad (\text{E } 5)$$

Let us consider $L_{9,t}$ and $L_{10,t}$ ($t = 1, \dots, n^*$) in more details, with the elements of the rows \mathbf{L}_9 and \mathbf{L}_{10} of **L** belonging to animal 5. From $\mathbf{Q}_5 = \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$ follows $SQ_1^1 = 1$ and $SQ_1^3 = 1$ and $1 \notin \mathbf{Q}_5$. With $\mathbf{N}_5 = (N_{5,1}; N_{5,2}) = (5; 7)$ (see Tab. III) the non zero elements of \mathbf{L}_9 are $L_{9,5} = Q_5(1; 1) = 0.5$ and with $L_{9,7} = Q_5(1; 3) = 0.5$ the non zero elements of \mathbf{L}_{10} are $L_{10,5} = Q_5(2; 1) = 0.5$ and $L_{10,7} = Q_5(2; 3) = 0.5$.

*Third part of the algorithm: Calculation of the condensed gametic relationship matrix **G****

The condensed gametic relationship matrix **G*** with full rank n^* can be computed by the use of the following generalization of the tabular method

of [21]: $\mathbf{G}_1^* = \mathbf{G}_1$ and for $i = 2, \dots, n$

$$\mathbf{G}_i^* = \begin{cases} \mathbf{G}_{i-1}^* & , \text{ if } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ & \text{(no row/column is added)} \\ \left[\begin{array}{c|c} \mathbf{G}_{i-1}^* & \mathbf{G}_{i-1}^* \mathbf{A}_i^{*k'} \\ \mathbf{A}_i^{*k} \mathbf{G}_{i-1}^* & 1 \end{array} \right] & , \text{ if } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ & \text{(+ 1 row/column)} \\ \left[\begin{array}{c|c} \mathbf{G}_{i-1}^* & \mathbf{G}_{i-1}^* \mathbf{A}_i^{*j'} \\ \mathbf{A}_i^{*j} \mathbf{G}_{i-1}^* & \mathbf{C}_{ii} \end{array} \right] & , \text{ if } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \text{ (+ 2 rows/column)} \end{cases} \quad (9)$$

where \mathbf{A}_i^* is a $(2 \times n_{i-1}^{\max})$ -dimensional matrix with columns $N_{s(i),1}, N_{s(i),2}, N_{d(i),1}, N_{d(i),2}$ being identically with the first, second, third, fourth column of \mathbf{Q}_i in this order and zero elements otherwise. \mathbf{A}_i^{*k} is the k th row of \mathbf{A}_i^* , where $k = 1$ if $j = 2$ or $k = 2$ if $j = 1$.

The example from Table II illustrates the computation of \mathbf{G}^* (see its non-zero elements in (E 6))

$$\mathbf{G}^* = \begin{bmatrix} 1.00 & - & - & - & - & - & .50 & - & .50 & .25 \\ & -1.00 & - & - & - & - & .50 & - & .50 & .25 \\ & & -1.00 & - & - & - & .50 & - & - & - \\ & & & -1.00 & - & - & .50 & - & - & - \\ & & & & -1.00 & - & - & - & - & .50 \\ & & & & & -1.00 & - & - & - & - \\ .50 & .50 & - & - & - & -1.00 & - & .50 & .50 & - \\ - & - & .50 & .50 & - & - & -1.00 & - & - & - \\ .50 & .50 & - & - & - & - & .50 & -1.00 & .25 & - \\ .25 & .25 & - & - & .50 & - & .50 & - & .25 & 1.00 \end{bmatrix} \quad (E 6)$$

and more detailed \mathbf{G}_1^* for animal $i = 7$.

\mathbf{G}_6^* is the matrix with the first 9 rows and columns in (E 6) with $\text{rank}(\mathbf{G}_6^*) = n_6^{\max} = 9$. From Table III we get $\mathbf{N}_7 = (N_{7,1}; N_{7,2}) = (10; 8)$ for animal 7, $\mathbf{N}_{s(7)} = \mathbf{N}_5 = (5; 7)$ for its sire and $\mathbf{N}_{d(7)} = \mathbf{N}_6 = (9; 8)$ for its dam. $N_{7,1} = 10 > n_6^{\max} = 9$ and $N_{7,2} = 8 \leq n_6^{\max} = 9$ implies that the first row of \mathbf{A}_7^* must be used, containing the elements $(0.5 \ 0.5 \ 0 \ 0)$ of the first row of \mathbf{Q}_7 (Tab. II) at the places 5, 7, 9, 8: $\mathbf{A}_7^{*1} = (0 \ 0 \ 0 \ 0 \ 0.5 \ 0 \ 0.5 \ 0 \ 0)$. Therefore $\mathbf{A}_7^{*1} \mathbf{G}_6^*$ results in $(.25 \ .25 \ .00 \ .00 \ .50 \ .00 \ .50 \ .00 \ .25)$ and $\mathbf{G}_7^* = \left[\begin{array}{c|c} \mathbf{G}_6^* & \mathbf{G}_6^* \mathbf{A}_7^{*1'} \\ \mathbf{A}_7^{*1} \mathbf{G}_6^* & 1 \end{array} \right] = \mathbf{G}^*$ (see (E 6)).

*Fourth part of the algorithm: Computation of the inverse of the condensed gametic relationship matrix \mathbf{G}^**

The inverse \mathbf{G}^{*-1} of the condensed gametic relationship matrix \mathbf{G}^* can be determined by generalization of the tabular method of [21] in an analogous way: $\mathbf{G}_1^{*-1} = \mathbf{G}_1^{-1}$ and for $i = 2, \dots, n$

$$\mathbf{G}_i^{*-1} = \begin{cases} \mathbf{G}_{i-1}^{*-1} & , \text{ if } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \\ \left[\begin{array}{c|c} \mathbf{G}_{i-1}^{*-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] + \left[\begin{array}{c|c} \frac{1}{d_i^*} \mathbf{A}_i^{*k'} \mathbf{A}_i^{*k} & -\frac{1}{d_i^*} \mathbf{A}_i^{*k'} \\ \hline -\frac{1}{d_i^*} \mathbf{A}_i^{*k} & \frac{1}{d_i^*} \end{array} \right] & , \text{ if } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}, \\ & k \neq j; k, j = 1, 2 \\ \left[\begin{array}{c|c} \mathbf{G}_{i-1}^{*-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] + \left[\begin{array}{c|c} \mathbf{A}_i^{*'} \mathbf{D}_i^{*-1} \mathbf{A}_i^{*'} & -\mathbf{A}_i^{*'} \mathbf{D}_i^{*-1} \\ \hline -\mathbf{D}_i^{*-1} \mathbf{A}_i^{*'} & \mathbf{D}_i^{*-1} \end{array} \right] & , \text{ if } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \end{cases} \quad (10)$$

where $d_i^* = (1 - \mathbf{A}_i^{*k} \mathbf{G}_{i-1}^* \mathbf{A}_i^{*k'})$ if one row and column is added and $\mathbf{D}_i^* = (\mathbf{C}_{ii} - \mathbf{A}_i^* \mathbf{G}_{i-1}^* \mathbf{A}_i^{*'})$ if two rows and columns are added. Calculation of d_i^* and \mathbf{D}_i^* can be simplified by using $d_i^* = (1 - \mathbf{Q}_i^k \widehat{\mathbf{G}}_i \mathbf{Q}_i^{k'})$ and $\mathbf{D}_i^* = (\mathbf{C}_{ii} - \mathbf{Q}_i \widehat{\mathbf{G}}_i \mathbf{Q}_i')$, where \mathbf{Q}_i^k is the k th row of \mathbf{Q}_i and

$$\widehat{\mathbf{G}}_i = \begin{bmatrix} \mathbf{G}_{s(i)} & \mathbf{G}_{s(i)d(i)} \\ \mathbf{G}_{d(i)s(i)} & \mathbf{G}_{d(i)} \end{bmatrix},$$

with

$$\mathbf{G}_{s(i)d(i)} = \begin{bmatrix} \mathbf{G}_{i-1}^* (N_{s(i),1}; N_{d(i),1}) & \mathbf{G}_{i-1}^* (N_{s(i),1}; N_{d(i),2}) \\ \mathbf{G}_{i-1}^* (N_{s(i),2}; N_{d(i),1}) & \mathbf{G}_{i-1}^* (N_{s(i),2}; N_{d(i),2}) \end{bmatrix} = \mathbf{G}'_{d(i)s(i)},$$

$$\mathbf{G}_{s(i)} = \begin{bmatrix} 1 & f_{s(i)} \\ f_{s(i)} & 1 \end{bmatrix}, \quad f_{s(i)} = \mathbf{G}_{i-1}^* (N_{s(i),1}; N_{s(i),2}), \quad \mathbf{G}_{d(i)} = \begin{bmatrix} 1 & f_{d(i)} \\ f_{d(i)} & 1 \end{bmatrix},$$

$f_{d(i)} = \mathbf{G}_{i-1}^* (N_{d(i),1}; N_{d(i),2})$ and $\mathbf{G}_{i-1}^*(o; t)$ is the t th element in the o th row of \mathbf{G}_{i-1}^* .

We continue with animal 7 for illustration of (10). With \mathbf{G}_6^{*-1} we only have to compute d_7^* to get $\mathbf{G}_7^{*-1} = \mathbf{G}^{*-1}$ (see (E 7) for non-zero elements of the

inverse \mathbf{G}^{*-1}) with $\mathbf{A}_7^{*1} = (0 \ 0 \ 0 \ 0 \ 0.5 \ 0 \ 0.5 \ 0 \ 0)$ from above.

$$\mathbf{G}^{*-1} = \begin{bmatrix} 2.00 & 1.00 & - & - & - & - & -1.00 & - & -1.00 & - \\ 1.00 & 2.00 & - & - & - & - & -1.00 & - & -1.00 & - \\ - & - & 1.50 & .50 & - & - & - & -1.00 & - & - \\ - & - & .50 & 1.50 & - & - & - & -1.00 & - & - \\ - & - & - & - & 1.50 & - & .50 & - & - & -1.00 \\ - & - & - & - & - & 1.00 & - & - & - & - \\ -1.00 & -1.00 & - & - & .50 & - & 2.50 & - & - & -1.00 \\ - & - & -1.00 & -1.00 & - & - & - & 2.00 & - & - \\ -1.00 & -1.00 & - & - & - & - & - & - & 2.00 & - \\ - & - & - & - & -1.00 & - & -1.00 & - & - & 2.00 \end{bmatrix} \tag{E 7}$$

with

$$f_{s(7)} = G_6^*(N_{s(7),1}; N_{s(7),2}) = G_6^*(5, 7) = 0,$$

$$f_{d(7)} = G_6^*(N_{d(7),1}; N_{d(7),2}) = G_6^*(9; 8) = 0$$

and

$$\begin{aligned} \mathbf{G}_{s(7)d(7)} &= \begin{bmatrix} G_6^*(N_{s(7),1}; N_{d(7),1}) & G_6^*(N_{s(7),1}; N_{d(7),2}) \\ G_6^*(N_{s(7),2}; N_{d(7),1}) & G_6^*(N_{s(7),2}; N_{d(7),2}) \end{bmatrix} = \begin{bmatrix} G_6^*(5; 9) & G_6^*(5; 8) \\ G_6^*(7; 9) & G_6^*(7; 8) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0.5 & 0 \end{bmatrix} \end{aligned}$$

and the first row ($k = 1$) of \mathbf{Q}_7 (Tab. II) is $d_7^* = (1 - \mathbf{Q}_7^1 \widehat{\mathbf{G}}_7 \mathbf{Q}_7^{1'}) = (1 - 0.5) = 0.5$. Thus $-\frac{1}{d_7^*} \mathbf{A}_7^{*1} = -2 \cdot \mathbf{A}_7^{*1}$ is equal to $(0 \ 0 \ 0 \ 0 \ -1 \ 0 \ -1 \ 0 \ 0)$ (see (E 7)).

Finally, it is straightforward to verify that $\mathbf{L}\mathbf{G}^*\mathbf{L}' = \mathbf{G}$, where \mathbf{L} is from (E 5), \mathbf{G}^* from (E 6) and \mathbf{G} from (E 4).

Decomposition of \mathbf{G}^* from (9) into $\mathbf{G}^* = \mathbf{B}^*\mathbf{D}^*\mathbf{B}'$ can be done in an analogy to (4) with $\mathbf{D}^* = \text{diag}(\widehat{\mathbf{D}}_1^*, \dots, \widehat{\mathbf{D}}_n^*)$, where

$$\widehat{\mathbf{D}}_i^* = \begin{cases} - & , \text{ if } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} & \text{(no row/column is added)} \\ d_i^* & , \text{ if } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 & \text{(+ 1 row/column)} \\ \mathbf{D}_i^* & , \text{ if } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} & \text{(+ 2 rows/column)} \end{cases} \tag{11}$$

and d_i^* and D_i^* are from (10) recursively calculated B^* : $B_1^* = I_2$ and for $i = 2, \dots, n$:

$$B_i^* = \begin{cases} B_{i-1}^* & , \text{ if } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \quad (\text{no row/column is added}) \\ \left[\begin{array}{c|c} B_{i-1}^* & \mathbf{0} \\ \hline A_i^{*k} B_{i-1}^* & 1 \end{array} \right] & , \text{ if } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ & (+ 1 \text{ row/column}) \\ \left[\begin{array}{c|c} B_{i-1}^* & \mathbf{0} \\ \hline A_i^{*k} B_{i-1}^* & I_2 \end{array} \right] & , \text{ if } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \quad (+ 2 \text{ rows/columns}) \end{cases} \quad (12)$$

Consequently, the inverse of the condensed gametic relationship matrix G^* can be calculated as $G^{*-1} = (B^{*'})^{-1} D^{*-1} B^{*-1}$, with $D^{*-1} = \text{diag}(\widehat{D}_1^{*-1}, \dots, \widehat{D}_n^{*-1})$ and recursively determined B^{*-1} : $B_1^{*-1} = I_2$ and for $i = 2, \dots, n$

$$B_i^{*-1} = \begin{cases} B_{i-1}^{*-1} & , \text{ if } N_{i,1} \leq n_{i-1}^{\max} \wedge N_{i,2} \leq n_{i-1}^{\max} \quad (\text{no row/column is added}) \\ \left[\begin{array}{c|c} B_{i-1}^{*-1} & \mathbf{0} \\ \hline -A_i^{*k} & 1 \end{array} \right] & , \text{ if } N_{i,k} > n_{i-1}^{\max} \wedge N_{i,j} \leq n_{i-1}^{\max}; k \neq j; k, j = 1, 2 \\ & (+ 1 \text{ row/column}) \\ \left[\begin{array}{c|c} B_{i-1}^{*-1} & \mathbf{0} \\ \hline -A_i^{*k} & I_2 \end{array} \right] & , \text{ if } N_{i,1} > n_{i-1}^{\max} \wedge N_{i,2} > n_{i-1}^{\max} \quad (+ 2 \text{ rows/columns}) \end{cases} \quad (13)$$

in analogy to (5).

With $w = Tv$, $v = Lv^*$, the relation $Q^G = 0.5 \cdot TGT' = 0.5 \cdot T(LG^*L')T'$ between conditional genotypic relationship matrix Q^G and conditional gametic relationship matrices G and G^* , it is easy to verify that G^* from (E 6) and G from (E 4) result in the same conditional genotypic relationship matrix

$$Q^G = \begin{bmatrix} 1.000 & - & - 0.500 & 0.500 & 0.500 & 0.250 \\ & - 1.000 & - 0.500 & - 0.500 & 0.500 & \\ & - & - 1.000 & - 0.500 & - 0.250 & \\ 0.500 & 0.500 & - 1.000 & 0.500 & 0.750 & 0.750 \\ 0.500 & - 0.500 & 0.500 & 1.000 & 0.250 & 0.500 \\ 0.500 & 0.500 & - 0.750 & 0.250 & 1.000 & 0.625 \\ 0.250 & 0.500 & 0.250 & 0.750 & 0.500 & 0.625 & 1.000 \end{bmatrix} \cdot \quad (E 8)$$

Again, we consider the situation that the gametes α_1^1, α_1^2 are identified by parental origin for Table II data and calculate Q_i^* according to (2b). The Q_i^* s for

the non-base animals are $\mathbf{Q}_4^* = (0.5 \ 0.5)'$, $\mathbf{Q}_5^* = (0.5 \ 1.0)'$, $\mathbf{Q}_6^* = (0.5 \ 0.0)'$ and $\mathbf{Q}_7^* = (0.5 \ 0.0)'$. Applying the condensing algorithm for this data, the condensed gametic relationship matrix \mathbf{G}^* (see (E 9) for non-zero elements)

$$\begin{bmatrix} 1.00 & - & - & - & - & - & .50 & - & - & .50 & .25 \\ - & 1.00 & - & - & - & - & .50 & - & - & .50 & .25 \\ - & - & 1.00 & - & - & - & .50 & - & - & - & - \\ - & - & - & 1.00 & - & - & .50 & - & - & - & - \\ - & - & - & - & 1.00 & - & .50 & - & - & .50 & .25 \\ .50 & .50 & - & - & - & - & 1.00 & - & - & .50 & .50 \\ - & - & .50 & .50 & - & - & - & 1.00 & - & - & - \\ - & - & - & - & .50 & .50 & - & - & 1.00 & - & .50 \\ .50 & .50 & - & - & - & - & .50 & - & - & 1.00 & .25 \\ .25 & .25 & - & - & .25 & .25 & .50 & - & .50 & .25 & 1.00 \end{bmatrix} \quad (\text{E } 9)$$

and its inverse \mathbf{G}^{*-1} (see (E 10) for non-zero elements)

$$\begin{bmatrix} 2.00 & 1.00 & - & - & - & - & -1.00 & - & - & -1.00 & - \\ 1.00 & 2.00 & - & - & - & - & -1.00 & - & - & -1.00 & - \\ - & - & 1.50 & .50 & - & - & - & -1.00 & - & - & - \\ - & - & .50 & 1.50 & - & - & - & -1.00 & - & - & - \\ - & - & - & - & 1.50 & .50 & - & - & -1.00 & - & - \\ - & - & - & - & .50 & 1.50 & - & - & -1.00 & - & - \\ -1.00 & -1.00 & - & - & - & - & 2.50 & - & .50 & - & -1.00 \\ - & - & -1.00 & -1.00 & - & - & - & 2.00 & - & - & - \\ - & - & - & - & -1.00 & -1.00 & .50 & - & 2.50 & - & -1.00 \\ -1.00 & -1.00 & - & - & - & - & - & - & - & 2.00 & - \\ - & - & - & - & - & - & -1.00 & - & -1.00 & - & 2.00 \end{bmatrix} \quad (\text{E } 10)$$

can be calculated recursively. In contrast to (E 5) (10 remaining effects) there are now 11 effects left. The differences in the condensed gametic relationship matrices ((E 6) vs. (E 9)) and their inverses ((E 7) vs. (E 10)) are evident. \mathbf{G}^* in (E 6) is of rank 10 and has 34 non-zero elements, \mathbf{G}^* in (E 9) is of rank 11 and has 43 non-zero elements. The corresponding inverses are of rank 10 with 32 non-zeros (E 7) versus rank 11 with 39 non-zero elements (E 10). But again the matrices of (E 6) and (E 9) result in the identical genotypic relationship matrix (E 8), which can easily be verified. This means that the number and size of estimates of gametic MQTL effects depend on the mode of gamete identification, but the sum of these effects remains unaffected for each animal.

5. DISCUSSION

Gamete identification by parental origin and gamete identification by markers have already been used earlier in the literature without a clear distinction [1, 2, 14, 21]. In this article it was shown for the first time, at least to the authors' knowledge, that both identification methods result in different conditional gametic relationship matrices and different estimates of gametic effects in a MA-BLUP model. It could, however, be demonstrated that the MA-BLUP breeding value – the estimated sum of QTL-gamete effects and the polygenic effect – remain the same irrespective of the method of gamete identification.

A practical advantage of identification by parental origin is that both the sire-block and the dam-block of the \mathbf{Q}_i -matrices (2a) can each be represented by a single number, namely the probability that the paternal allele of the sire and the dam have been transmitted to the descendant, respectively. The reason for this is that each block (sire and dam) of the \mathbf{Q}_i -matrices has only two non-zero entries which sum up to one, if the \mathbf{Q}_i -matrices reflect gamete identification by parental origin. The alternative mode of gamete identification needs three probabilities for each block, because the number of non-zero entries per block (again summing up to one) is four when gamete identification is done by marker. It may be worthwhile to store \mathbf{Q}_i -matrices for all animals additionally to marker raw data as an essence of marker information and intermediate result in computing the \mathbf{G} matrix and its inverse and also for other purposes as *e.g.* the computation of measures of marker information content. Though six numbers per animal will not be prohibitive to store even with tens of thousands of animals, identification by parental origin needs only two and is therefore easier to administer.

The genotypic relationship matrix \mathbf{Q}^G in the MQTL genotypic effect model may either be determined by deterministic methods [1, 2, 9, 13, 21] or by Markov chain Monte Carlo (MCMC) [5, 12, 16–18, 20] and their advantages and disadvantages have been investigated by several authors [11, 13]. MCMC has been implemented in the LOKI program [6] in order to compute \mathbf{Q}^G by MCMC, but currently it cannot be used to compute \mathbf{G} for a MQTL allelic effects model. Though not primarily designed for this purpose SimWalk2 [15] can be employed to achieve this goal: it reports MCMC estimates of all 15 detailed identity state probabilities [4] between any pair of animals in the pedigree. \mathbf{Q}_i matrices can be derived from the SimWalk2 output implicitly using gamete identification by parental origin. This follows from the definition of identity states in SimWalk2 software [15].

The MA-BLUP breeding value of each animal in (1) equals the sum of the MQTL genotypic effect and the polygenic effect. Thus it would be sufficient to

use a MQTL genotypic effects model and to save one equation per genotyped animal. Since the MQTL genotypic effect is the sum of the first and the second MQTL allele effects, one positive and one negative gametic effect may give rise to the same genotypic effect as two gametic effects of average size. It would be interesting for breeders to know how an MQTL genotypic effect is composed and therefore it is desirable to have estimates of the available gametic effects. A conclusion of the considerations above is, that a certain mode of gamete identification has to be chosen. It may be more natural to animal breeders to think in pedigrees and parental origin rather than in marker haplotypes and therefore gamete identification by parental origin may be preferred for this purpose.

It has been proposed by [11] to estimate MQTL genotypic effects \mathbf{w} and to transform these into allelic effects \mathbf{v} . If such a transformation exists, it would have to be specific for a certain definition of \mathbf{v} , *i.e.* it would depend on the mode of gamete identification. The conclusion of [11] that $\mathbf{v} = 0.5 \cdot \mathbf{GT}'(\mathbf{Q}_{(n \times n)}^G)^{-1} \mathbf{w}$ follows from $\mathbf{Tv} = \mathbf{w} = \mathbf{TGT}'(\mathbf{TGT}')^{-1} \mathbf{w} = 0.5 \cdot \mathbf{TGT}'(\mathbf{Q}_{(n \times n)}^G)^{-1} \mathbf{w}$ is however not possible, because a left-inverse, say $\mathbf{T}^{(\text{left})-}$, of \mathbf{T} with $\mathbf{T}_{(2n \times n)}^{(\text{left})-} \mathbf{T}_{(n \times 2n)} = \mathbf{I}_{2n}$ does not exist. For the proof we only have to consider the first two elements of the first row of $\mathbf{T}_{(2n \times n)}^{(\text{left})-} \mathbf{T}_{(n \times 2n)}$

$$\mathbf{T}_{(2n \times n)}^{(\text{left})-} \mathbf{T}_{(n \times 2n)} = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{2n,1} & t_{2n,2} & \cdots & t_{2n,n} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix} = \begin{bmatrix} t_{1,1} & t_{1,1} & | & \cdots \\ \vdots & \vdots & | & \ddots \end{bmatrix} \neq \mathbf{I}_{2n},$$

where $t_{1,1}$ cannot take the values 1 and 0 at the same time, and consequently, $\mathbf{T}_{(2n \times n)}^{(\text{left})-} \mathbf{T}_{(n \times 2n)}$ never equals the identity matrix.

Both modes of gamete identification fail for some animals. The gametes of homozygous individuals (both base and non-base) cannot be distinguished by markers. On the contrary, gamete identification by parental origin fails in all founders and in all non-founders from non-informative matings, where from marker analysis it cannot be deduced whether the marker was inherited from the dam or from the sire. As the number of markers is increased, the probability for a mating to be non-informative for all markers as well as the probability for an individual to be homozygous at all marker loci becomes smaller and smaller. Multiple markers will therefore help to identify nearly all gametes unequivocally. This is, however, not true for founder animals, when gamete identification is done by parental origin. This problem can, however, be resolved by identifying founder gametes by markers and by arbitrarily denoting one of

both gametes of each founder animal as maternal and the other as paternal. By applying this rule to the example pedigree [1] it can be shown that a third variant of a conditional gametic relationship matrix results and differs from the two others demonstrated above (data not shown), but again transforms to the same \mathbf{Q}^G -matrix and, consequently, leads to the same MA-BLUP breeding values.

Gamete identification by marker may be of interest in applications where the intention is to test a certain polymorphism for linkage disequilibrium with a QTL. With two alleles and in using this polymorphism for gamete identification, all gametes with allele 1 are treated as the first and all gametes with allele 2 as the second gamete of heterozygous animals. The expectation is that, if the polymorphism is in strong linkage disequilibrium with the QTL, differences between the first and the second gametic effect of heterozygotes will exhibit the same sign and roughly the same size, provided there is a sufficient accuracy of both gametic estimates.

A reduction of the size of the conditional gametic relationship matrix has already been proposed by [10]: parents and offspring sharing the same marker haplotype were treated as sharing the same QTL allele, by assuming a zero probability of double recombinations. In these cases the same gametic QTL effect was assigned to the parent and the offspring. When the offspring has received a recombinant marker haplotype a new gametic QTL effect was defined as the average of both gametes of the parent animal. Treating these two gametes as parents of the new gamete allows to set up a pedigree of gametic effects and to compute the conditional gametic relationship matrix and its inverse simply by applying the Henderson rules [7, 8]. The condensing algorithm will, of course, lead to identical results, if desired: sire- and dam-blocks of progeny with non-recombinant parental haplotypes have to carry zeros and ones only, and in the recombinant case the corresponding sire- and dam-blocks are assembled by fifty-percent transition probabilities as in the case without markers. The Meuwissen and Goddard proposal [10] therefore combines a special case of the condensing algorithm with gamete identification by parental origin. Assuming the QTL in the middle of a certain marker interval and rounding the QTL-transition probabilities to either one or zero if they are closer to these values as a predefined threshold, *e.g.* 0.02, will give the same results as in [10] for those animals which are informative at the markers flanking this particular interval. For the other animals information from markers more remote and asymmetrically distributed around the QTL's assumed home interval may be available. In these cases the probability of double recombinations may be too high to be neglected and, furthermore, the assumption of an equal transmission

probability of the first and second parental allele may be unrealistic in the light of the markers transmitted. These animals can easily be combined with the former group by maintaining the original transition probabilities in the Q_i -matrices without rounding and then applying the condensing algorithm to all pedigree members in the same way, no matter of previous rounding or not.

In conclusion, the condensing algorithm is a generalization of the Abdel-Azim and Freeman algorithm [1] for computing the conditional gametic relationship matrix and its inverse. Although suggested before by other authors, computing this inverse cannot be avoided if estimates of gametic QTL effects are desired.

The condensing algorithm can be applied to different modes of gamete identification, situations with and without markers including X-linkage, clones and haplodiploid pedigrees. Treatment of haplotypes according to the proposals of [10] are covered as a special case and can be combined with exact treatment of more remote marker information.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Fritz Reinhardt and Dr. Hauke Thomsen (Vereinigte Informationssysteme Tierhaltung w. V. Verden, Germany), Dr. Jörn Bennewitz (Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität Kiel, Germany) for their helpful discussion and Dr. Gertraude Freyer for useful hints on SimWalk2 software.

REFERENCES

- [1] Abdel-Azim G., Freeman A.E., A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked Quantitative Trait Locus, *Genet. Sel. Evol.* 33 (2001) 153–173.
- [2] Fernando R.L., Grossman M., Marker-assisted selection using best linear unbiased prediction, *Genet. Sel. Evol.* 21 (1989) 467–477.
- [3] Georges M., Recent progress in livestock genomics and potential impact on breeding programs, *Theriogenology* 55 (2001) 15–21.
- [4] Gillois M., La relation d'identité en génétique, Thèse Faculté des Sciences de Paris (1964).
- [5] Grignola F.G., Hoeschele I., Tier B., Mapping quantitative trait loci *via* residual maximum likelihood, I. Methodology, *Genet. Sel. Evol.* 28 (1996) 479–490.
- [6] Heath S.C., Markov chain Monte Carlo segregation and linkage analysis for oligogenic models, *Am. J. Hum. Genet.* 61 (1997) 748–760.
- [7] Henderson C.R., Best linear unbiased estimation and prediction under a selection model, *Biometrics* 31 (1975) 423–439.

- [8] Henderson C.R., A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, *Biometrics* 32 (1976) 69–83.
- [9] Knott S.A., Elsen J.M., Haley C.S., Methods for multiple marker mapping of quantitative trait loci in half-sib populations, *Theor. Appl. Genet.* 93 (1996) 71–80.
- [10] Meuwissen T.H.E., Goddard M.E., The use of marker-haplotypes in animal breeding schemes, *Genet. Sel. Evol.* 28 (1996) 161–176.
- [11] Nagamine Y., Knott S.A., Visscher P.M., Haley C.S., Simple deterministic identity-by-descent coefficients and estimation of QTL allelic effects in full and half sibs, *Genet. Res. Camb.* 80 (2002) 237–243.
- [12] Pérez-Enciso M., Varona L., Rothschild M.F., Computation of identity by descent probabilities conditional on DNA markers *via* a Monte Carlo Markov chain method, *Genet. Sel. Evol.* 32 (2000) 467–482.
- [13] Pong-Wong R., George A.W., Woolliams J.A., Haley C.S., A simple and rapid method for calculating identity-by-descent matrices using multiple markers, *Genet. Sel. Evol.* 33 (2001) 453–471.
- [14] Schaeffer L.R., Kennedy B.W., Gibson J.P., The inverse of the gametic relationship matrix, *J. Dairy Sci.* 72 (1989) 1266–1272.
- [15] Sobel E., Lange K., Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics, *Am. J. Hum. Genet.* 58 (1996) 1323–1337.
- [16] Thompson E.A., Monte Carlo likelihood in genetic mapping, *Stat. Sci.* 9 (1994) 355–366.
- [17] Thompson E.A., Inferring gene ancestry: estimating gene descent, *Int. Stat. Rev.* 66 (1998) 29–40.
- [18] Thompson E.A., Heath S.C., Estimation of conditional multilocus gene identity among relatives, in: Seiller-Moseiwitch F., Donnelly P., Watermann M. (Eds.), *Statistics in Molecular Biology*, Springer-Verlag IMS lecture note series, New York, 1999.
- [19] van Arendonk J.A.M., Tier B., Kinghorn B.P., Use of multiple genetic markers in prediction of breeding values, *Genetics* 137 (1994) 319–329.
- [20] Visscher P.M., Haley C.S., Heath S.C., Muir W.J., Blackwood DHR Detecting QTLs for uni- and bipolar disorder using a variance component method, *Psych. Genet.* 9 (1999) 75–84.
- [21] Wang T., Fernando R.L., van der Beek S., Grossman M., van Arendonk J.A.M., Covariance between relatives for a marked quantitative trait locus, *Genet. Sel. Evol.* 27 (1995) 251–274.