Original article

Efficient computation of the inverse of gametic relationship matrix for a marked QTL

Mehdi SARGOLZAEI[a*], Hiroaki IWAISAKI[a,b],
Jean-Jacques COLLEAU[c]

[a] Course of Environmental Management Science, Graduate School of Science and Technology,
Niigata 950-2181, Japan
[b] Department of Agro-biology, Faculty of Agriculture, Niigata University,
Niigata 950-2181, Japan
[c] Station de génétique quantitative et appliquée, Institut national de la recherche agronomique,
78352 Jouy-en-Josas Cedex, France

Abstract – Best linear unbiased prediction of genetic merits for a marked quantitative trait locus (QTL) using mixed model methodology includes the inverse of conditional gametic relationship matrix ($\mathbf{G}^{-1}$) for a marked QTL. When accounting for inbreeding, the conditional gametic relationships between two parents of individuals for a marked QTL are necessary to build $\mathbf{G}^{-1}$ directly. Up to now, the tabular method and its adaptations have been used to compute these relationships. In the present paper, an indirect method was implemented at the gametic level to compute these few relationships. Simulation results showed that the indirect method can perform faster with significantly less storage requirements than adaptation of the tabular method. The efficiency of the indirect method was mainly due to the use of the sparseness of $\mathbf{G}^{-1}$. The indirect method can also be applied to construct an approximate $\mathbf{G}^{-1}$ for populations with incomplete marker data, providing approximate probabilities of descent for QTL alleles for individuals with incomplete marker data.

BLUP / indirect method / inverse of gametic relationship matrix / marked QTL / tabular method

1. INTRODUCTION

For best linear unbiased prediction of genetic merits, Fernando and Grossman [3] proposed a mixed linear model to account for covariance of

---

* Corresponding author: f03m301f@mail.cc.niigata-u.ac.jp

a marked quantitative trait locus (QTL) between relatives. However, this approach requires the inverse of conditional gametic relationship matrix ($\mathbf{G}^{-1}$) for the marked QTL of interest. Therefore, efficient computation of $\mathbf{G}^{-1}$ is an important issue in large populations. This computation is more important in simulation studies, because $\mathbf{G}^{-1}$ must be obtained once for each replicate.

Computation of $\mathbf{G}$ and its inverse for a marked QTL, given completely ordered marker genotypes, was first described by Fernando and Grossman [3]. Goddard [4] extended this method to multiple markers where each QTL is flanked by two adjacent markers and double recombination between two flanking markers is ignored. Ruane and Colleau [8] constructed $\mathbf{G}^{-1}$ following the rules of Fernando and Grossman [3] except that as an approximation, inbreeding coefficients of parents based on pedigree information were used. van Arendonk *et al.* [14] showed how partitioned matrix theory can be applied to directly compute $\mathbf{G}^{-1}$ for a marked QTL. Wang *et al.* [15] extended the work of Fernando and Grossman [3] to the situation where parental origin of haplotypes may not be known (*i.e.*, unordered marker genotypes). Inbred parents were taken into account, and $\mathbf{G}^{-1}$ was directly calculated using the partitioned matrix theory. Based on the development by Wang *et al.* [15], Abdel-Azim and Freeman [1] calculated $\mathbf{G}^{-1}$ directly using the decomposition of $\mathbf{G}$. In constructing $\mathbf{G}^{-1}$ directly by the methods of Wang *et al.* [15] and Abdel-Azim and Freeman [1], accounting for inbreeding, the only nontrivial task is to compute a matrix which is proportional to the conditional Mendelian sampling covariance matrix for a marked QTL ($\mathbf{D}$). In order to compute $\mathbf{D}$, however, conditional relationships between two parents of all individuals are necessary [1, 15]. Wang *et al.* [15, 16] applied the tabular method to compute $\mathbf{D}$, whereas Abdel-Azim and Freeman [1] applied an adaptation of the tabular method [12] in which a small portion of $\mathbf{G}$ is enough to compute $\mathbf{D}$. Although the latter method provides a great computational advantage for large populations, its efficiency depends on the size of the required portion of $\mathbf{G}$.

Recently, a fast indirect method was developed for computing vectors involving the numerator relationship matrix in the polygenic model [2]. It was applied for computing relationship statistics and inbreeding coefficients [2,10]. The objective of this paper was to examine the potential, in terms of computational efficiency, of this indirect method when relationships are to be calculated at the gametic level in large populations in order to obtain $\mathbf{G}^{-1}$. To focus on this question, complete marker data with either known or unknown parental origins of haplotypes were assumed; however, possible application of the indirect method for populations with incomplete marker data will be discussed.

2. METHODS

2.1. Theoretical considerations

For simplicity, we first describe how to compute the inverse of $\mathbf{G}$ directly and what requirements have to be fulfilled. Let $\mathbf{G}$ denote the conditional gametic relationship matrix, $\mathbf{G} = \{\mathbf{G}_{ij}\}$, where $\mathbf{G}_{ij}$ is a 2×2 matrix that includes four gametic relationships between individuals i and j. Thus, $\mathbf{G}_{ii}$ contains 1s on the diagonal and the inbreeding coefficient, conditional on markers, of individual i on the off-diagonals. Note that individuals are ordered by birth date and numbered consecutively from 1 to n, where n is the number of individuals. Assuming complete marker information, two identified parents for non-founders and nonzero recombination rates between marker(s) and QTL, $\mathbf{G}$ can be decomposed, similar to decomposing the numerator relationship matrix [5] as:

$$\mathbf{G} = \mathbf{LDL}', \tag{1}$$

where $\mathbf{L} = \{\mathbf{L}_{ij}\}$ is a lower block triangular matrix which can be recursively formed from pedigree and marker data, and $\mathbf{D} = \mathrm{diag}\{\mathbf{D}_{ii}\}$ is a block diagonal matrix proportional to the conditional Mendelian sampling covariance matrix, or $\mathbf{M} = \mathbf{D}\sigma_v^2$, where σ_v^2 is the genetic variance contributed by the marked QTL [1, 15]. Matrix $\mathbf{D}_{ii}$ can be obtained as:

$$\mathbf{D}_{ii} = \mathbf{G}_{ii} - \mathbf{B}_i \begin{bmatrix} \mathbf{G}_{s_i s_i} & \mathbf{G}_{s_i d_i} \\ \mathbf{G}_{d_i s_i} & \mathbf{G}_{d_i d_i} \end{bmatrix} \mathbf{B}'_i, \tag{2}$$

where $\mathbf{B}_i$ is a 2×4 matrix containing the conditional probabilities that QTL alleles in individual i descended from parental alleles (PDQ). Details of computation of PDQ for a QTL linked to a single marker and to flanking markers have been given by Wang et al. [15] and Liu et al. [6], respectively. Off-diagonal terms of $\mathbf{G}_{ii}$ are expressed as $tr(\mathbf{T}'\mathbf{G}_{s_i d_i})$. Matrix $\mathbf{T}$ is a 2×2 matrix with term t_{mn} representing the probability that the pair of QTL 1 and 2 in individual i corresponds to QTL m of its sire and QTL n of its dam. For calculating $\mathbf{T}$, see Wang et al. [15]. Matrix $\mathbf{B}_i$ can be partitioned as $\begin{bmatrix} \mathbf{B}_i^s & \mathbf{B}_i^d \end{bmatrix}$, where $\mathbf{B}_i^s$ and $\mathbf{B}_i^d$ contain PDQ from paternal and maternal origins, respectively. Furthermore, $\mathbf{G}^{-1}$ using (1), is expressed as:

$$\mathbf{G}^{-1} = (\mathbf{L}^{-1})'\mathbf{D}^{-1}\mathbf{L}^{-1},$$

where $\mathbf{L}^{-1}$ is a lower block triangular matrix which contains $-\mathbf{B}_i^s$ and $-\mathbf{B}_i^d$ at positions linking individual i to its sire and dam and 1s on the diagonal. Therefore, $\mathbf{L}^{-1}$ can be directly and quickly obtained without computing $\mathbf{L}$.

Calculating matrix $\mathbf{D}$ requires knowing the conditional gametic relationships between the parents of all individuals. Note that $\mathbf{G}^{-1}$ can also be calculated by using the partitioned matrix theory [15].

From (2), it is clear that for non-inbred populations, $\mathbf{D}$ and consequently $\mathbf{G}^{-1}$ can be calculated simply since there are no gametic relationships between any individual's parents. However, for inbred populations, these relationships must be determined to compute $\mathbf{D}$ and this could be computationally costly if the conventional tabular method is used.

Abdel-Azim and Freeman [1] implemented Tier's algorithm [12], or adaptation of the tabular method, at the gametic level to compute $\mathbf{D}$ quickly. The idea behind the Tier algorithm is that only a small part of $\mathbf{G}$ is necessary to compute the conditional gametic relationships between two parents of individuals. The required elements of $\mathbf{G}$ are first flagged and then calculated [1, 12]. Although the required subset is a small part of $\mathbf{G}$, this subset might require a large amount of memory to be stored in large populations.

Colleau [2] presented an approach to multiply the numerator relationship matrix ($\mathbf{A}$) by a vector, say $\mathbf{x}$, regarding a planned mating design (*i.e.*, $\mathbf{y} = \mathbf{Ax}$). This method is called the "indirect" method, because it computes a group of elements of $\mathbf{A}$ simultaneously. With the indirect method, solutions are obtained from considering $\mathbf{A}^{-1}$ instead of $\mathbf{A}$. It is assumed that inbreeding coefficients of the ancestors are known. Colleau [2] also showed that inbreeding coefficients themselves can be obtained by the indirect method efficiently. Inbreeding coefficients for paternal sibs, either half or full, having the same pseudo-generation number were obtained simultaneously from relationships between the sire and relevant mates. The pseudo-generation number or equivalently longest ancestral path number was set to 0 for founders and subsequently incremented by 1 after considering the maximal value found for parents [2,9]. Recently, the algorithm for computing inbreeding coefficients by the indirect method was modified in order to facilitate the implementation of this method and to speed up the computation of the coefficients, especially with overlapping generations [10].

Now consider implementation of the indirect method at the gametic level to compute $\mathbf{D}$ for a marked QTL. It is assumed here that PDQ have already been calculated, and thus the only requirement to compute $\mathbf{D}$ is the conditional gametic relationships between the parents of every individual. The conditional gametic relationships between an individual and other individuals ($\mathbf{Y}$) can be expressed as:

$$\mathbf{Y} = \mathbf{GX}, \tag{3}$$

where $\mathbf{X}$ is a $n \times 2$ matrix consisting of a 2×2 identity matrix at the position of the individual of concern and null otherwise. Substituting (1) into (3) leads to:

$$\mathbf{Y} = \mathbf{LDL}'\mathbf{X}.$$

Denoting $\mathbf{L}'\mathbf{X}$ by $\mathbf{Z}$, we have:

$$(\mathbf{L}^{-1})'\mathbf{Z} = \mathbf{X} \qquad (4)$$

and

$$\mathbf{L}^{-1}\mathbf{Y} = \mathbf{DZ}. \qquad (5)$$

As mentioned earlier, $\mathbf{L}^{-1}$ is a lower block triangular matrix with special structure that contains only $-\mathbf{B}_i^s$ and $-\mathbf{B}_i^d$ at positions where individual i is linked to its sire and dam and 1s on the diagonal, and therefore (4) and subsequently (5) can be solved for $\mathbf{Z}$ and $\mathbf{Y}$, respectively, using Gaussian elimination. The elimination for (4) can be simply done by tracing up the pedigree that is to update $\mathbf{X}$ upwards as: $\mathbf{X}_s \Leftarrow \mathbf{X}_s + (\mathbf{B}_i^s)'\mathbf{X}_i$ and $\mathbf{X}_d \Leftarrow \mathbf{X}_d + (\mathbf{B}_i^d)'\mathbf{X}_i$. This provides solutions for $\mathbf{Z}$ which should be multiplied by $\mathbf{D}$ ($\mathbf{X}_i \Leftarrow \mathbf{D}_{ii}\mathbf{X}_i$). Note that $\mathbf{X}$ is now overwritten by $\mathbf{DZ}$. Thus, the elimination for (5) can be simply done by tracing down the pedigree – that is to update $\mathbf{X}$ downwards as: $\mathbf{X}_i \Leftarrow \mathbf{X}_i + \mathbf{B}_i^s\mathbf{X}_s$ and $\mathbf{X}_i \Leftarrow \mathbf{X}_i + \mathbf{B}_i^d\mathbf{X}_d$. This finally provides solutions for $\mathbf{Y}$. It is assumed that $\mathbf{D}_{ii}$ is known for the individual of concern and its ancestors. Therefore, conditional gametic relationships between a sire (or a dam) and its mates can be derived by this method in two simple steps, or tracing up and tracing down the pedigree. To compute all conditional gametic relationships of concern, the indirect method should be run as many times as the number of sires or dams. In practice, the number of sires is much less than that of dams.

2.2. Algorithm and processing strategies

The modified algorithm of the indirect method to compute inbreeding coefficients [10] can be applied by some changes to compute $\mathbf{D}$, as follows:

(1) compute PDQ for all individuals;

(2) sort identification of individuals by pseudo-generation numbers of their sires, and then by identification of their sires in ascending order;

(3) for j = 1 to max (pseudo-generation numbers):

(3-1) extract the pedigree of sires with pseudo-generation number j and their mates and recode them sequentially starting from 1. Thus, for the extracted pedigree, we have $\mathbf{Y}^* = \mathbf{G}^*\mathbf{X}^*$;

(3-2) calculate $\mathbf{D}_{ii}$ for the sires and their ancestors if not already calculated, and then form $\mathbf{D}^*$;

(3-3) process the indirect method sire after sire:

(3-3-1) set $\mathbf{X}^*$ to a 2×2 identity matrix at the position of the sire considered and null otherwise;

(3-3-2) solve $(\mathbf{L}^{*-1})'\mathbf{Z}^* = \mathbf{X}^*$ for $\mathbf{Z}^*$ by tracing up the extracted pedigree;

(3-3-3) solve $\mathbf{L}^{*-1}\mathbf{Y}^* = \mathbf{D}^*\mathbf{Z}^*$ for $\mathbf{Y}^*$ by tracing down the extracted pedigree;

(4) finally, calculate $\mathbf{D}_{ii}$ for individuals that have not had their $\mathbf{D}_{ii}$ calculated yet.

Matrices $\mathbf{Z}^*$ and $\mathbf{D}^*\mathbf{Z}^*$ are obtained simultaneously. Since elements of vector $\mathbf{Z}^*$ excluding those for the sire considered and his ancestors are certainly null, $\mathbf{D}^*\mathbf{Z}^*$ can be obtained only knowing $\mathbf{D}^*_{ii}$ for the sire and his ancestors [10]. Note that, as shown in the previous section, updating $\mathbf{X}^*$upward and subsequently downward, using information from the extracted pedigree, give $\mathbf{D}^*\mathbf{Z}^*$and $\mathbf{Y}^*$, respectively.

Theoretically, the indirect method should be solved by tracing a pedigree containing only the sire of concern, his mates and their ancestors. However, extracting and recoding the pedigree for every sire (plus his mates) separately is time-consuming. Therefore, with our previous algorithm for computing in-breeding coefficients [10], a reduced pedigree containing all sires and dams is extracted once. With this pedigree, processing the indirect method for each sire might include some excessive computations that do not contribute to the final results. In the case of the polygenic model, it seems that these excessive computations are less costly than extracting and recoding many pedigrees. However, inheritance of QTL alleles from parents to a child involves eight possible allelic pathways. Therefore, for the same problem, solving (4) and (5) requires much more computations than solving corresponding equations for the polygenic model and consequently more time for the computations possibly in excess. Hence, in order to lower excessive computations with the gametic model, a few small pedigrees are built by extracting the pedigree for each group of sires with the same pseudo-generation number and their mates. It should be pointed out that since the pseudo-generation number represents a chronological order [9], the indirect method can be processed group after group sequentially. With such a sequential process, $\mathbf{D}_{ii}$ for the sires of the current group and their ancestors will be available when needed. Excessive computations can be further reduced by using a Boolean vector to flag non-null block matrices of $\mathbf{X}$ as they are calculated.

2.3. Simulation

A simulation study was conducted to compare the computational efficiency of the indirect method [2] and adaptation of the tabular method [1, 12] (denoted as IM and AT, respectively). A situation where a QTL located in the middle of a 10 cM interval flanked by two markers was considered. The QTL contributed 20% of the additive genetic variation of the trait with heritability of 0.3. To create a QTL with the given additive genetic variation, the average effects of the two QTL alleles were set to 6.7 and –6.7, and the remaining additive and environmental variances to 360 and 1050, respectively. Considering three factors with two levels each, eight populations with complete marker information were generated for 10, 20 and 30 years of reproduction. The levels of the three factors were discrete or overlapping generations, 50 or 300 breeding sires and 2 or 8 litter size. All eight breeding populations consisted of 1500 dams. In the base generation, two alleles at equal frequencies for the QTL and the maximum number of alleles for the two markers (*i.e.*, maximum polymorphic markers) were simulated, and all founders were assumed to be unrelated, non-inbred and unselected. For the setting of overlapping generations, the oldest 50% and 30% of sires and dams, respectively, were replaced each year with selected progeny from the previous year. Parents were selected based on individual phenotypic values and mated at random.

The efficient sparse storage scheme described by Tier [12] was applied to store the required upper triangular subset of $\mathbf{G}$. Although this scheme requires a little more memory than that described by Abdel-Azim and Freeman [1], it is much faster, because it avoids repeated searching through linked list and retrieves elements on the lower diagonal efficiently. Each entry of the sparse matrix is a 2×2 matrix representing $\mathbf{G}_{ij}$. To avoid redundant computations, PDQ are obtained once and saved in memory with both algorithms. The PDQ for flanking markers and the number of discrete-generation equivalents were calculated according to Liu *et al.* [6] and Meuwissen and Luo [7], respectively. The probabilities of linkage phases (*cis* or *trans*) were calculated using information of parents and progeny first. Then the PDQ of each individual were calculated with respect to parental origin of haplotypes, gametic frequencies, probabilities of linkage phases of itself, its sire and its dam. In all eight populations, less than 5% of parental origin of haplotypes in generation 30 was unknown. The computational cost for calculating PDQ, time required for allocating memory, and memory used to store pedigree and marker information were assessed. The algorithm described by Wang *et al.* [15] was applied to compose $\mathbf{G}^{-1}$ after calculating $\mathbf{D}$. All the algorithms were programmed in C++ language. Variable types of *float* (4 bytes), *int* (4 bytes), *short int*

(2 bytes) and *bool* (1 byte) were declared and computations were carried out on a PC Pentium 4, 3.2 GHz processor under Windows XP (1 Gb of RAM).

3. RESULTS AND DISCUSSION

The performance of the two algorithms, in terms of computing time and storage requirements, to compute matrix **D** is presented in Tables I and II, respectively. The results clearly indicate that IM could be run in less time and with less memory requirement compared to AT. The computing time of IM relative to AT ranged from 2 to 50%. With discrete generations, both AT and IM performed slower than with overlapping generations, which may be due to the larger number of ancestors and the larger equivalent number of discrete-generations. As expected, the computing time of both methods was increased substantially with increasing number of sires. A smaller number of sires indicates a smaller number of genetic origins in the population, and therefore the number of required elements of **G** was far less with 50 sires compared with 300 sires. The effect of the number of sires on IM was due to the fact that this method essentially solves (4) and (5) as many times as the number of sires. However, the relative efficiency of AT decreased more quickly with the increased number of sires. The results (not shown) from a study of the effect of the number of dams indicate that IM was significantly less sensitive to the number of dams compared to AT.

Since litter size increased from 2 to 8, the total number of individuals increased 4 times. In spite of that, AT became relatively faster with larger litter size. Similar findings were reported for polygenic models [7, 10]. With larger litter size, selection pressure is increased because the same number of breeding individuals is chosen among a higher number of candidates. Hence, the number of ancestors of the selected individuals tends to decrease, contributing to decreasing the number of required elements of **G**. In contrast, IM consumed slightly more time with larger litter size except for the discrete situation with 300 sires. This could be because increased numbers of individuals increases the overhead for other tasks.

Increasing the number of generations increased the computing time for both methods. More precisely, the rate of increase was higher for IM than for AT. While IM ran fast compared to AT in all situations considered, AT may become competitive with IM in terms of computing time when both litter size and the number of generations are large, but may still require much more memory than IM.

Table I. CPU time (in s) required by the adaptation of the tabular method (AT) and the indirect method (IM) for computing matrix **D**.

No. of sires	Litter size	Years evaluated	No. of individuals	No. of dis. gen. equiv.[1]	Discrete		Overlapping	
					AT	IM	AT	IM
50	2	0–10	31 550	6.4	1.9	0.1	0.3	0.1
		0–20	61 550	12.4	8.9	0.4	1.3	0.1
		0–30	91 550	18.3	17.8	0.9	2.5	0.2
	8	0–10	121 550	6.8	1.2	0.2	0.4	0.2
		0–20	241 550	13.5	4.3	0.5	1.2	0.4
		0–30	361 550	20.0	7.4	1.0	2.2	0.6
300	2	0–10	31 800	6.0	24.2	0.5	1.9	0.1
		0–20	61 800	11.8	97.7	3.5	9.9	0.5
		0–30	91 800	17.7	342.5[2]	9.5	17.7	1.3
	8	0–10	121 800	6.3	9.9	0.5	1.8	0.2
		0–20	241 800	12.6	37.0	2.9	7.1	0.7
		0–30	361 800	18.9	65.9	7.3	12.9	1.6

[1] Computed for individuals born in the last year with overlapping generations.
[2] Hard disk memory was used.

Table II. Memory (in Mb) required by the adaptation of the tabular method (AT) and the indirect method (IM) for computing matrix **D**.

No. of sires	Litter size	Years evaluated	Discrete			Overlapping		
			AT	IM	No. of required elements of $\mathbf{G}$[1]	AT	IM	No. of required elements of $\mathbf{G}$[1]
50	2	0–10	34.3	3.4	5 535 322	9.7	2.8	1 244 014
		0–20	125.5	7.0	21 046 382	30.1	5.4	4 375 274
		0–30	217.8	10.0	36 749 726	51.5	8.0	7 686 430
	8	0–10	30.5	10.3	3 588 682	16.7	9.7	1 183 862
		0–20	85.6	20.4	11 505 926	40.3	19.1	3 587 134
		0–30	139.1	30.6	19 135 686	64.5	28.6	6 091 406
300	2	0–10	183.3	3.6	31 584 180	29.1	2.9	4 625 232
		0–20	607.8	7.0	105 330 252	112.6	5.6	18 799 380
		0–30	1032.9	10.5	179 189 156	195.6	8.2	32 862 484
	8	0–10	105.6	10.5	16 701 572	33.9	9.8	4 174 208
		0–20	330.3	20.8	54 255 500	99.8	19.3	13 979 460
		0–30	562.6	31.1	93 132 984	168.8	28.8	24 306 928

[1] Computed for AT.

In general, the memory requirement of AT showed the same trend as was seen for computing time and was larger than that of IM in all settings examined. In the case of discrete generations with 300 sires, two progeny per dam and 30 years of reproduction, AT performed worst. This bad performance of AT could be linked to a larger number of ancestors with a more dispersed

Table III. The number of calculated elements of $\mathbf{G}^{-1}$ and memory (in Mb) required to store them as a lower triangular sparse matrix.

No. of sires	Litter size	Years evaluated	Discrete		Overlapping	
			No. of calculated elements	Memory required	No. of calculated elements	Memory required
50	2	0–10	404 172	2.1	392 144	2.0
		0–20	824 092	4.2	811 360	4.2
		0–30	1 243 972	6.4	1 230 540	6.3
	8	0–10	1 486 128	7.6	1 475 112	7.5
		0–20	2 985 980	15.2	2 974 396	15.2
		0–30	4 485 744	22.9	4 473 552	22.8
300	2	0–10	396 860	2.0	381 200	1.9
		0–20	816 792	4.2	800 748	4.1
		0–30	1 236 748	6.4	1 220 404	6.3
	8	0–10	1 479 316	7.5	1 464 528	7.5
		0–20	2 979 200	15.2	2 964 276	15.1
		0–30	4 479 124	22.9	4 464 012	22.8

pedigree, which led to an increase in the number of required elements of $\mathbf{G}$. In this case, the memory requirement for AT was 1032.9 Mb or 98 times that required by IM, and computing time was exceptionally larger due to the use of hard disk space. In contrast to computing time, storage requirements of IM did not significantly differ between overlapping and discrete generations and between 50 and 300 sires. However, the memory required by IM was greater with the larger litter size and more generations, due to increased total number of individuals in these cases. As a result, the storage requirement of IM was almost a linear function of the total number of individuals, whereas that of AT could be a function of many factors such as the number of ancestors, litter size, the number of generations, and selection strategy. Populations that require a large amount of memory with AT can be divided into smaller populations based on the related group [12]. Nevertheless, such divided populations may always have some part of their pedigrees in common, and consequently repeated computation of the same elements of $\mathbf{G}$ would often be required, which would result in a marked increase in computing time.

Table III shows the number of calculated elements of $\mathbf{G}^{-1}$ and memory required to store them as a lower triangular sparse matrix. The total CPU time consumed for creating the triangular inverse was less than 0.2 seconds in all scenarios.

Finally, a problem lies in calculating $\mathbf{G}^{-1}$ for populations with incomplete marker data. A solution for this situation is given by Wang *et al.* [15], which is not feasible for large populations. Given missing marker information,

approximate methods to compute $\mathbf{G}$ and its inverse for large populations in which PDQ are first approximated and then $\mathbf{G}$ is constructed recursively were presented [13, 15]. The approximated PDQ can also be used by AT or IM to compute $\mathbf{D}$ and consequently $\mathbf{G}^{-1}$. Although these approximate methods do not adequately utilize the existing marker information for approximating PDQ [13], the use of a more exact method such as the Monte Carlo Markov chain might improve the accuracy of the approximations. Recently, given incomplete marker data, a method for constructing a gametic covariance matrix for an equivalent mixed model was developed [11]. With this method, however, genotype probabilities for individuals with missing marker information should be calculated, and two random effects for each possible marker genotype should be considered.

In conclusion, we show that it is feasible to apply the indirect method at the gametic level for computing conditional gametic relationships between two parents of all individuals in inbred populations which are necessary for calculating matrix $\mathbf{D}$ and finally $\mathbf{G}^{-1}$. Simulation results indicate that the indirect method can compute $\mathbf{D}$ faster with significantly less storage requirements compared to the adaptation of the tabular method. The efficiency of the indirect method relies on the use of the sparseness of $\mathbf{G}^{-1}$ which can be exploited quickly in two simple steps.

ACKNOWLEDGEMENTS

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] Colleau J.J., An indirect approach to the extensive calculation of relationship coefficients, Genet. Sel. Evol. 34 (2002) 409–421.

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

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

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

[6] Liu Y., Jansen G.B., Lin C.Y., The covariance between relatives conditional on genetic markers, Genet. Sel. Evol. 34 (2002) 657–678.

[7] Meuwissen T.H.E., Luo Z., Computing inbreeding coefficients in large populations, Genet. Sel. Evol. 24 (1992) 305–313.

[8] Ruane J., Colleau J.J., Marker assisted selection for genetic improvement of animal populations when a single QTL is marked, Genet. Res. 66 (1995) 71–83.

[9] Sargolzaei M., Iwaisaki H., An efficient algorithm for computing inbreeding coefficients in large populations, Jpn. J. Biom. 25 (2004) 25–36.

[10] Sargolzaei M., Iwaisaki H., Colleau J.J., A fast algorithm for computing inbreeding coefficients in large populations, J. Anim. Breed. Genet. 122 (2005) 325–331.

[11] Thallman R.M., Hanford K.J., Kachman S.D., van Vleck L.D., Sparse inverse of covariance matrix of QTL effects with incomplete marker data, Stat. Appl. Genet. Mol. Biol. 3 (2004) Article 30.

[12] Tier B., Computing inbreeding coefficients quickly, Genet. Sel. Evol. 22 (1990) 419–430.

[13] Totir L.R., Fernando R.L., Dekkers J.C.M., Fernandez S.A., Guldbrandtsen B., The effect of using approximate gametic variance covariance matrices on marker assisted selection by BLUP, Genet. Sel. Evol. 36 (2004) 29–48.

[14] 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.

[15] 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.

[16] Wang T., Fernando R.L., Grossman M., Genetic evaluation by BLUP using marker and trait information in a multibreed population, Genetics 148 (1998) 507–515.

To access this journal online:
www.edpsciences.org