Optimal Marker-Assisted Selection to Increase the Effective Size of Small Populations

Jinliang Wang

Institute of Zoology, Zoological Society of London, London NW1 4RY, United Kingdom

Manuscript received August 4, 2000
Accepted for publication October 24, 2000

ABSTRACT

An approach to the optimal utilization of marker and pedigree information in minimizing the rates of inbreeding and genetic drift at the average locus of the genome (not just the marked loci) in a small diploid population is proposed, and its efficiency is investigated by stochastic simulations. The approach is based on estimating the expected pedigree of each chromosome by using marker and individual pedigree information and minimizing the average coancestry of selected chromosomes by quadratic integer programming. It is shown that the approach is much more effective and much less computer demanding in implementation than previous ones. For pigs with 10 offspring per mother genotyped for two markers (each with four alleles at equal initial frequency) per chromosome of 100 cM, the approach can increase the average effective size for the whole genome by ~40 and 55% if mating ratios (the number of females mated with a male) are 3 and 12, respectively, compared with the corresponding values obtained by optimizing between-family selection using pedigree information only. The efficiency of the marker-assisted selection method increases with increasing amount of marker information (number of markers per chromosome, heterozygosity per marker) and family size, but decreases with increasing genome size. For less prolific species, the approach is still effective if the mating ratio is large so that a high marker-assisted selection pressure on the rarer sex can be maintained.

A major genetic problem in maintaining small populations under captive breeding is the inescapable accumulation of inbreeding and genetic drift over generations, which puts them in jeopardy of immediate extinction due to inbreeding depression and also risks their survival in the long run due to the depletion of genetic variation and loss of evolutionary potential (Frankham 1995; Lacy 1997). Minimizing inbreeding and genetic drift is, therefore, a fundamental task in the genetic management of small conserved populations. Traditionally, the most effective method to minimize inbreeding and drift is to equalize the contribution of offspring from all potential ancestors, realized generally by selecting the breeding individuals with the smallest average coancestry among them at each generation (Ballou and Lacy 1995; Lacy 1995; Caballero and Toro 2000). In the ideal situation of equal contribution among parents mated at random at each generation, the effective population size (N_e) is maximized to ~$2N$, where N is the actual size of a monoecious population or a dioecious population with equal numbers of males and females (Crow and Kimura 1970; Caballero 1994). Therefore, minimum coancestry or equal contribution of parents is widely recommended in conservation practice (Ballou and Lacy 1995; Frankham 1995). For a diploid species, N_e cannot be increased above $2N$, except for mating schemes involving a higher than random probability of matings between close relatives [*e.g.*, population subdivision (Robertson 1964; Wang and Caballero 1999) or circular pair mating (Kimura and Crow 1963) in a single population], which are impractical for conservation populations because of the threat of inbreeding depression (Frankham 1995).

The traditional method described above for making selection decisions to minimize inbreeding utilizes only pedigree information, which describes the *expected* relationship among individuals. With the same pedigree, however, individuals of diploid species still vary greatly in their genetic makeup or in the *realized* genetic relationship among them. In this respect, genetic markers are very useful for inferring the realized genetic relationship and could be potentially utilized in increasing N_e of small populations. For using marker information to increase N_e, several approaches such as frequency-dependent selection and selection for heterozygosity at marker loci have been proposed (*e.g.*, Chevalet and Rochambeau 1986), and their effectiveness on decreasing inbreeding and genetic drift has been compared in a simulation study (Toro *et al.* 1998). Their efficiency is, however, rather limited because the marker information is not fully utilized (Wang and Hill 2000).

For a better utilization of genetic markers to increase N_e, Wang and Hill (2000) proposed a method aimed at minimizing the variation in genetic contribution between paternally and maternally derived genes within individuals under a given between-family selection

Address for correspondence: Institute of Zoology, Regent's Park, London NW1 4RY, United Kingdom. E-mail: jinliang.wang@ioz.ac.uk

scheme. The idea is that inbreeding and genetic drift for a diploid population come from the variation in contribution both among individuals and between the two homologous genes at a locus within individuals. The first can be minimized by equalizing family size, while the second can be controlled by using marker information to select, from within a family, the offspring that have the minimum average probability of identity by descent (PIBD; Wang and Hill 2000). This marker-assisted selection (MAS) method is simple in implementation and effective compared with previous marker-based approaches. It has, however, two limitations. First, it optimizes within-family selection on markers independently of between-family selection on pedigree. Though the latter and former can be optimized by minimizing the average coancestry among the selected offspring on the basis of pedigree information and by minimizing the average PIBD among selected offspring within families on the basis of marker information, respectively, the separate operations do not guarantee a globally optimized selection. Second, MAS is carried out separately for each family, while the genetic relationship among families is ignored in this MAS method. These limitations are partially removed in Toro *et al.*'s (1999) method aimed at minimizing the average coancestry of selected individuals conditional on marker information. The implementation of their method, however, has to resort to the Monte Carlo Markov chain approach, and therefore has high computational demands and may not yield the optimal solution (Toro *et al.* 1999).

In this article, I propose a simple method that optimizes the use of marker and pedigree information simultaneously to minimize inbreeding and genetic drift in a small population. Given such marker and pedigree information, the average (realized) coancestry for all loci between two diploid genomes can be estimated and then used to choose individuals that result in the minimum average (realized) coancestry among them by a standard quadratic integer programming technique. Numerical examples show that the proposed method is powerful in decreasing inbreeding and genetic drift, and also efficient in implementation compared with previous methods.

THEORY AND METHOD

First, the average realized coancestry between two homologous chromosomes given their respective marker genotypes and pedigrees is computed. Second, the mean coancestry between any two diploid genomes is obtained by averaging over chromosomes and is used in minimizing the average coancestry among all selected individuals by integer programming. I consider the simplest case of one marker locus per chromosome in detail and then extend the treatment to two and more marker loci per chromosome. Throughout this article, *genes* are used when their identities in state or identities by descent are not distinguished, while *alleles* are used to refer to different allelic variants at a locus in the population.

One marker locus per chromosome: Let us consider a diploid population consisting of N_1 males and N_2 females, each male being mated randomly with N_2/N_1 females (called mating ratio and denoted by r hereafter) at each discrete generation. Each mating is assumed to give n offspring of each sex that are marker-genotyped for selection. There are at least two codominant alleles segregating in the population at a single marker locus on each chromosome of the genome. The problem is to select, from the $2n_T = 2nN_2$ offspring available with known marker and pedigree information, N_1 males and N_2 females with the minimum average coancestry among them as the next generation. This is accomplished by the following procedures.

1. Determine the origin of an offspring homologue. Consider a single chromosome in the genome. The two homologues in an offspring, denoted as A and A*, come from separate parents. We need to determine the probability that a gene taken at random from a particular homologue (say, A) comes from homologue i ($i = 1$ or 2) in parent s ($s = 1$ or 2 for father or mother) using marker information.
   1.1. Ascertain the parental origin of the marker gene on offspring homologue A. Denote the gene at a marker locus on homologue A as A and the gene on the other homologue as A* in the offspring. If both marker genes A and A* can be found in each parent, then marker A comes equally frequently from paternal or maternal origin. The probability that marker A in the offspring originates from parent s, $P_{A,s}$, is therefore ½. If either marker A is found in parent s only, or A* in parent s^* only ($s \neq s^* = 1$ or 2 for father or mother), then marker A comes from parent s with probability 1, $P_{A,s} = 1$.
   1.2. Determine the probability that marker A comes from homologue i in parent s, $Q_{A,is}$, conditional on marker A coming from this parent. If marker A is identical to both marker genes in parent s, then $Q_{A,is} = ½$ ($i = 1, 2$). If marker A is identical to the marker on homologue i only, then $Q_{A,is} = 1$ and $Q_{A,i^*s} = 0$ ($i \neq i^* = 1, 2$).
   1.3. Determine the probability that marker A comes from homologue i in parent s. This probability is the product of probabilities P and Q, $R_{A,is} = Q_{A,is}P_{A,s}$.
   1.4. Determine the average probability that genes at all loci on offspring homologue A come from homologue i in parent s, $\overline{R}_{A,is}$. Consider the gene at a locus situated x morgans away from the marker locus on homologue A. The probabilities that the gene and marker A have not and have recombined during transmission are $y_x = ½(1 + e^{-2x})$ and $1 - y_x$, respectively, assuming

Haldane's mapping function. Assuming that loci are distributed uniformly along the chromosome and integrating y_x over intervals 0–L_1 and 0–L_2 (Wang and Hill 2000) gives the average probability that the marker locus and the other loci have not recombined, y, as

$$y = \tfrac{1}{2} + (2 - e^{-2L_1} - e^{-2L_2})/(4L), \quad (1)$$

where L_1 and L_2 are the distances from the marker locus to the two ends of the chromosome with total length $L = L_1 + L_2$. The probability that a random gene on offspring homologue A comes from homologue i in parent s is simply the weighted average

$$\begin{aligned}\overline{R}_{\mathrm{A},is} &= yR_{\mathrm{A},is} + (1 - y)R_{\mathrm{A},i^*s} \quad (i \neq i^* = 1, 2)\\ &= yR_{\mathrm{A},is} + (1 - y)(P_{\mathrm{A},s} - R_{\mathrm{A},is}). \quad (2)\end{aligned}$$

The equation is derived noting that $R_{\mathrm{A},i^*s} = (1 - Q_{\mathrm{A},is})P_{\mathrm{A},s} = P_{\mathrm{A},s} - R_{\mathrm{A},is}$ (see step 1.3 above).

2. Determine the average coancestry between two offspring homologues. For a homologue (say A′) in another offspring, we can determine similarly the probability that it comes from homologue i' ($i' = 1$ or 2) in parent s' ($s' = 1$ or 2 for father or mother), $\overline{R}_{\mathrm{A}',i's'}$. Denoting the average coancestry between homologues i (in parent s for offspring homologue A) and i' (in parent s' for offspring homologue A′) as $G_{is,i's'}$, the coancestry between offspring homologues A and A′ is

$$G_{\mathrm{A},\mathrm{A}'} = \sum_{s=1}^{2}\sum_{s'=1}^{2}\sum_{i=1}^{2}\sum_{i'=1}^{2}\overline{R}_{\mathrm{A},is}\overline{R}_{\mathrm{A}',i's'}G_{is,i's'}. \quad (3)$$

The self-coancestry of any homologue, such as A, is always 1, $G_{\mathrm{A},\mathrm{A}} \equiv 1$. The coancestry between different homologues within an offspring is calculated as

$$\begin{aligned}G_{\mathrm{A},\mathrm{A}'} = {} & \sum_{s=1}^{2}\sum_{s'\neq s=1}^{2}\sum_{i=1}^{2}\sum_{i'=1}^{2}\overline{R}_{\mathrm{A},is}\overline{R}_{\mathrm{A}',i's'}G_{is,i's'}/P_{\mathrm{A},s}\\ & (P_{\mathrm{A},s} \neq 0). \quad (4)\end{aligned}$$

Equation 4 is derived as follows. The two homologues within an offspring always come from separate parents. The probability that homologues A and A′ in the offspring originate from the parents of sexes s and s' ($s \neq s' = 1$ or 2), respectively, is $P_{\mathrm{A},s}$ or $P_{\mathrm{A}',s'}$ ($P_{\mathrm{A}',s'} \equiv P_{\mathrm{A},s}$) as calculated in step 1.1. Given $P_{\mathrm{A},s}$, the probability that A is from homologue i in parent s is $\overline{R}_{\mathrm{A},is}/P_{\mathrm{A},s}$, and the probability that A′ is from homologue i' in parent s' is $\overline{R}_{\mathrm{A}',i's'}/P_{\mathrm{A}',s'}$. The average coancestry between homologues A and A′ is then

$$G_{\mathrm{A},\mathrm{A}'} = \sum_{s=1}^{2}\sum_{s'\neq s=1}^{2}\left[P_{\mathrm{A},s}\sum_{i=1}^{2}\sum_{i'=1}^{2}\left(\frac{\overline{R}_{\mathrm{A},is}}{P_{\mathrm{A},s}}\frac{\overline{R}_{\mathrm{A}',i's'}}{P_{\mathrm{A}',s'}}G_{is,i's'}\right)\right].$$

Since $P_{\mathrm{A}',s'} \equiv P_{\mathrm{A},s}$, the above expression reduces to (4).

3. Determine the average coancestry between two offspring. Consider two offspring j and j'. Their coancestry, for the chromosome in question, is the average coancestry over the four possible pairs of homologues (one from each offspring),

$$g_{j,j'} = \tfrac{1}{4}\sum_{A=1}^{2}\sum_{A'=1}^{2}G_{Aj,A'j'} \quad (j \neq j'), \quad (5)$$

where subscript Aj ($A'j'$) refers to homologue A (A') in offspring j (j'). The self-coancestry of offspring j is

$$g_{j,j} = \tfrac{1}{2}(1 + G_{1j,2j}). \quad (6)$$

The coancestry between the two whole diploid genomes j and j', $\overline{g}_{j,j'}$, is obtained by averaging $g_{j,j'}$ over chromosomes in the genome weighted by their lengths.

4. Select the offspring with the minimum average coancestry among them. From the $2n_{\mathrm{T}}$ offspring, N_1 males and N_2 females should be selected so that the average coancestry among them, including reciprocal and self-coancestries, is minimized (Caballero and Toro 2000). Denoting the average coancestry between offspring j of sex s and j' of sex s' as $\overline{g}_{js,j's'}$, the optimization is realized by minimizing the function

$$\theta = \frac{1}{4}\sum_{s=1}^{2}\sum_{s'=1}^{2}\left(\frac{1}{N_sN_{s'}}\sum_{j=1}^{n_{\mathrm{T}}}\sum_{j'=1}^{n_{\mathrm{T}}}u_{js}u_{j's'}\overline{g}_{js,j's'}\right) \quad (7)$$

subject to the restriction

$$\sum_{j=1}^{n_{\mathrm{T}}}u_{js} = N_s \quad (s = 1 \text{ or } 2), \quad (8)$$

where indicator variable $u_{js} = 1$ or 0 if offspring j of sex s is selected or not. Equations 7 and 8 can be solved by integer quadratic programming, or by simulated annealing methods (Press *et al.* 1992). The problem can also be transformed into and solved by an integer linear programming technique (Fernández and Toro 1999).

When marker information is not available, the above marker-assisted selection procedure reduces to the method of selection based on minimizing the average coancestry of selected offspring calculated from pedigree information only (Ballou and Lacy 1995; Lacy 1995). This is obvious because when marker genotypes are unknown or uninformative (say, the marker is fixed in the population) then we always have $\overline{R}_{\mathrm{A},is} = \tfrac{1}{4}$ from the derivation of (2). From (3–5), it is clear that the coancestry between any two offspring is the average over the four pairs of parents, as expected. Therefore, the MAS procedure described above also applies to optimized between-family selection using pedigree information only. For the application of MAS in practice, missing marker genotypes for an individual or for some chromosomes within an individual can be dealt with similarly.

Two marker loci per chromosome: The amount of marker information for, and its relevance to, MAS are increased by using more than one marker locus per chromosome. With two or more marker loci per chro-

mosome, however, additional difficulty comes with determining the linkage phase. In the following, I outline the procedure for the use of two marker loci per chromosome in MAS to minimize the average coancestry among selected individuals.

1. Determine haploid marker types (linkage phases). Consider a single chromosome. If an individual is not a double heterozygote for the markers on the chromosome in question, its haploid marker genotypes are straightforward. Otherwise, its linkage phase needs to be determined.

   For a double heterozygous parent, the prior probability (β) of each linkage phase is known. At generation 0, it is reasonable to assume that the two linkage phases are equally probable and $\beta_k = 0.5$ ($k = 1$ or 2 for coupling or repulsion). At later generations, β_k for an individual can be determined from its parental genotypes (see below). Given the offspring genotypes and the priors, the posterior probability (ϕ_k) of linkage phase k can be obtained by Bayes's theorem. The linkage phase with the larger ϕ_k value is accepted for the parent. In programming, these operations to determine linkage phases are facilitated by assigning a specific value to each marker allele and each marker genotype. Although the power of this procedure to infer linkage phases decreases with the decline in the number of offspring per parent, it seems to work well because the MAS efficiency is satisfactory (compared with a single marker locus; see results in Table 1) even with only two offspring per mother (unequal sex ratio) being available for selection.

   For a double heterozygous offspring, the probabilities of the two linkage phases (β_k) are easily determined given its parental genotypes and linkage phases. For example, the probabilities of linkage phases *GH/gh* and *Gh/gH* for an offspring with genotype *GgHh* from parents *GH/gh* and *GH/Gh* are $1 - c$ and c, respectively, where c is the recombination fraction between marker loci *G* and *H*. These probabilities are priors in calculating ϕ_k in the next generation if the offspring is selected as a parent.
2. Determine the origins of the homologues in an offspring. Consider homologues denoted by A and A* of an offspring. If one or both marker genes on homologue A are found only in parent s (= 1 or 2 for father or mother), or alternatively one or both marker genes on homologue A* are found only in parent s^* ($s^* \neq s = 1, 2$), then homologue A comes from parent s with probability 1, $P_{A,s} = 1$. Otherwise, homologue A could come from either parent. The probability that it comes from parent s can be calculated, by using Bayes's theorem, as $P_{A,s} = P_{A,s}(AA^*)/[P_{A,s}(AA^*) + P_{A,s^*}(AA^*)]$, where $P_{A,s}(AA^*)$ is the probability of obtaining homologues A and A* in the offspring given that they originate from parents s and s^*, respectively. Probabilities $P_{A,s}(AA^*)$ and $P_{A,s}(AA^*)$ can be calculated easily by comparing the marker genotypes and linkage phases of the parents and the offspring. Consider the following parental and offspring haplotypes for two markers *G* and *H* as an example: *GH/gh* for the father, *Gh/gH* for the mother, and *GH/Gh* for the offspring. If the offspring homologue with marker alleles *G* and *H* on it is denoted as A, then obviously $P_{A,1}(AA^*) = \frac{1}{4}(1 - c)^2$, $P_{A,2}(AA^*) = \frac{1}{4}c^2$, and $P_{A,1} = (1 - c)^2/(1 - 2c + 2c^2)$, where c is the recombination fraction between marker loci *G* and *H*.
3. Determine the average probability that genes at all loci on homologue A of the offspring come from haploid i (= 1 or 2) in parent s, $\overline{Q}_{A,is}$, given $P_{A,s}$. The following four cases are possible.

   If both markers on homologue A are identical with those on each haploid of the parent, then obviously

   $$\overline{Q}_{A,is} = 0.5. \tag{9}$$

   If the marker at locus M (= 1 or 2 for left or right marker locus) is identical only with that on haploid i of the parent, and the marker at locus M^* ($\neq M = 1, 2$) is identical with both marker genes in the parent, then it can be derived, using a procedure similar to the single marker case, that

   $$\overline{Q}_{A,is} = \tfrac{1}{2} + (2 - e^{-2L_M} - e^{-2(L-L_M)})/(4L), \tag{10}$$

   where L_M is the distance between the left marker and the left end (if $M = 1$) or the right marker and the right end (if $M = 2$) of the chromosome.

   If both markers on homologue A are identical only with those on haploid i in parent s, then

   $$\overline{Q}_{A,is} = \tfrac{1}{2} + 1/[L(1 + e^{-2L_3})] - (e^{-2L_1} + e^{-2L_2})/(4L), \tag{11}$$

   where L_3 is the distance between the markers and $L = L_1 + L_2 + L_3$.

   If markers at loci M and M^* ($M \neq M^* = 1, 2$) on homologue A are identical only with those on haploids i and i^* ($i \neq i^* = 1, 2$) in parent s, respectively, then

   $$\overline{Q}_{A,is} = \tfrac{1}{2} + (e^{-2L_{M^*}} - e^{-2L_M})/(4L). \tag{12}$$

   For all the above cases, we always have that

   $$\overline{Q}_{A,i^*s} = 1 - \overline{Q}_{A,is} \quad (i \neq i^* = 1, 2). \tag{13}$$

   The probability that a gene taken at random on homologue A comes from haploid i in the parent of sex s is

   $$\overline{R}_{A,is} = P_{A,s}\overline{Q}_{A,is}. \tag{14}$$
4. Calculate the average coancestry between homologues of the same or different offspring using $\overline{R}_{A,is}$ and the same formula described in the single-marker case. For a double-heterozygous offspring, two linkage phases need to be considered separately using

procedures 2 and 3. Therefore, 4, 8, or 16 possible pairs of homologues (one from each offspring) need to be considered and the corresponding average coancestries calculated when none, one, or all of the two offspring are double heterozygotes for markers on the chromosome in question.

5. Determine the average coancestry between two offspring. For an offspring double heterozygous for the markers on a chromosome, its coancestry with another one is calculated separately for each linkage phase. The average weighted by the probabilities (β_k) of the two linkage phases is used in calculating the average coancestry between two diploid genomes, which is used in the mathematical programming aimed at minimizing the average coancestry of selected individuals shown in the single-marker case.

Many marker loci per chromosome: Three or more marker loci per chromosome can be treated similarly to the two-marker case shown above. With an increasing number of marker loci per chromosome, the formulations become inevitably more complicated, but the extra efficiency diminishes because the restricting factor to MAS efficiency is usually the number of genotyped offspring per parent in practice.

As the number of informative marker loci increases, the parental origin of each homologue can be inferred with increasing confidence. In the limit, the identity for every bit of a homologue can be deduced from marker information. The MAS efficiency is constrained, therefore, only by the number of marker-genotyped offspring available for selection. This extreme case is considered in the simulations described below assuming that the origin of each chromosome is completely known.

SIMULATION RESULTS

The efficiency in increasing N_e of the marker-assisted selection method developed above was investigated by stochastic simulations and compared with that of previous methods. A total of 174 loci, each with two alleles of equal initial frequency, equally spaced on each chromosome were simulated and the realized (harmonic) mean effective size was calculated from both the decrease in heterozygosity, averaged over loci and replicate runs, and the increase in the variance of allele frequency among replicate runs, averaged over loci, between generations 5 and 20. The two methods yielded essentially the same results, which were averaged as the realized mean effective size. The efficiency of MAS was expressed as percentage increase in realized mean effective size relative to the corresponding value without MAS, which was obtained by the same procedure but using pedigree information only. The simulated population consisted either of five breeding males and $5r$ breeding females when the mating ratio $r = 1$ or 3, or of three breeding males and $3r$ breeding females when $r = 6$ or 12. Each chromosome was assumed to be 100 cM in map length, and each marker locus was assumed to have four codominant alleles at equal initial frequency. Each female had an equal number of offspring (half of each sex) genotyped for selection. For a set of parameters, 100 replicates were run.

When a single-marker locus, situated at the center of a chromosome, was used for MAS, the increases in harmonic mean effective size are as shown in Table 1 (columns 2–6) for different mating ratios, numbers of chromosomes, and family sizes (half of each sex). For equal numbers of males and females, the efficiency of the present method is similar to that of Wang and Hill (2000), which considers between- and within-family selections separately, noting in the comparison that the sexes of the offspring selected from within a family are and are not restricted in the previous and the present methods, respectively. This is expected because, for the case of equal numbers of the two sexes, the genetic contributions are well balanced among parents and MAS does not affect the optimized between-family selection that results in two selected offspring per family.

When the numbers of the two sexes are different, an imbalance in genetic contribution among parents is introduced each generation. For example, a female parent whose son is selected contributes more genetically than a female parent whose daughter is selected. The imbalance in genetic contribution among parents can be minimized at each generation by using marker and pedigree information simultaneously in the present approach. Its efficiency, therefore, is much higher than our previous approach. When eight offspring (half of each sex) from each mother are genotyped for a single marker (four alleles at equal frequency) per chromosome in a haploid genome of 20 chromosomes (of 1 M each), for example, the average N_e can be increased by ~28, 31, and 35% for mating ratios 3, 6, and 12, respectively, by the present method, while the corresponding increases are about 10, 9, and 8% by the previous method (Wang and Hill 2000, Table 2).

In comparing the efficiencies, we should note that the reference selection schemes are different between the two approaches. The reference scheme in the present approach is optimized between-family selection using pedigree information, which is superior to the reference selection scheme of Gowe *et al.* (1959) used in our previous approach. Compared with Gowe *et al.*'s scheme, optimized between-family selection results in an increase in N_e of ~13–19% depending on mating ratios (data not shown), which is similar in performance to the selection scheme combined with group mating proposed by Wang (1997). It is obvious that the present approach is even better than indicated by the direct comparison of Table 1 herein and Table 2 in Wang and Hill (2000) if the difference between the two reference selection schemes is taken into account.

The MAS efficiency increases with increasing mating

TABLE 1

Percentage increase in effective size by MAS

Family size[a]	No. of markers per chromosome[b]/No. of chromosomes per haploid genome														
	1/1	1/5	1/10	1/20	1/30	2/1	2/5	2/10	2/20	2/30	m/1	m/5	m/10	m/20	m/30
							$r = 1$ ($N_{ep} = 19.0$)[c]								
4	34	21	15	13	9	56	30	24	16	12	109	47	33	25	19
8	47	34	26	19	16	105	54	37	27	22	321	87	58	44	33
16	51	44	33	23	19	115	67	49	34	30	918	136	81	59	46
32	50	45	38	30	23	121	84	57	42	33	1425	181	108	68	55
							$r = 3$ ($N_{ep} = 26.6$)								
2	30	9	5	2	1	45	15	7	5	3	72	27	20	9	6
4	42	35	23	20	17	81	47	32	23	18	244	77	55	33	29
8	65	40	30	28	20	108	59	44	34	28	571	107	75	56	44
16	64	45	37	31	27	122	71	56	42	34	968	148	96	66	54
32	65	47	41	34	30	136	86	64	46	40	1665	182	114	82	64
							$r = 6$ ($N_{ep} = 16.7$)								
2	41	17	10	6	5	65	25	18	10	7	120	45	28	20	15
4	74	43	33	22	17	114	55	40	25	20	289	102	63	44	32
8	75	49	41	31	25	143	78	54	37	29	696	141	93	63	48
16	73	58	45	32	29	143	89	68	47	37	1217	166	121	75	56
32	73	63	50	36	30	140	102	81	57	44	2328	211	140	90	72
							$r = 12$ ($N_{ep} = 16.8$)								
2	70	32	23	16	14	95	46	31	25	19	182	69	45	34	26
4	109	55	36	27	20	159	76	54	37	31	396	125	83	55	45
8	111	71	49	35	28	199	106	73	50	39	897	182	116	75	60
16	106	81	60	43	36	193	132	91	61	50	1501	230	144	93	70
32	108	90	69	48	38	193	151	104	72	56	3219	259	175	109	82

[a] Number of offspring (half of each sex) genotyped for MAS per female parent.

[b] One (situated at the center), two [left (right) marker situated ⅓ M from the left (right) end of the chromosome], or many (denoted as m) markers on a chromosome of 1 M are used in MAS. Each marker has four codominant alleles with equal initial frequency.

[c] The simulated population consists of five males and $5r$ females when mating ratio is 1 or 3, or three males and $3r$ females when mating ratio is 6 or 12. Using pedigree information only, this method can optimize between-family selection, and the effective size obtained is denoted as N_{ep}. MAS efficiency is expressed as the percentage increase over the corresponding N_{ep} value.

ratio. This is because a larger mating ratio results in a higher degree of imbalance in genetic contribution among parents and also a larger paternal family size for a given maternal family size. Both factors tend to increase the efficiency of MAS. When the mating ratio is greater than one, MAS even with two genotyped offspring per mother is still effective, and the efficiency increases rapidly with increasing mating ratios (Table 1). This relationship between mating ratio and MAS efficiency for the present method is in contrast to that for our previous one, where MAS efficiency decreases with increasing mating ratio because the within-family variation in genetic contribution on which MAS acts becomes less important compared with between-family variation.

The simulation results for the efficiency of MAS with two marker loci per chromosome, the left (right) locus being situated one-third morgan to the left (right) end of the chromosome, are listed in Table 1 (columns 7–11). Compared with a single marker, use of two marker loci per chromosome generally increases MAS efficiency, and the increase is greatest when the number of chromosomes is small and family size is large, where the restricting factor for MAS efficiency is the amount of marker information and its relevance to all loci over the whole chromosome.

Compared with our previous method, use of two markers per chromosome in the present method increases the effective size enormously, especially when mating ratio is high. This is because with an increasing mating ratio, the variation in genetic contribution among families that is ignored in the previous MAS method becomes increasingly important compared with within-family variation in determining the inbreeding and drift processes.

Other issues related to the use of two marker loci per chromosome, such as marker location and chromosome length, have been considered by Wang and Hill (2000) and the results apply also to the present method.

Columns 12–16 in Table 1 show the MAS efficiency

when the identity of each chromosome is completely known, which is realized in practice by using many informative markers per chromosome. As is clear, full knowledge of chromosome origins increases MAS efficiency enormously compared with the one- or two-marker case, especially when family size is large. The number of markers required to infer unambiguously the chromosome identity varies depending on the recombination frequency of the chromosome. With no recombination (*e.g.*, males in Drosophila), for example, one informative marker per chromosome is enough.

In Table 1, I considered populations consisting of either 5 (if $r = 1$ or 3) or 3 (if $r = 6$ or 12) males. The absolute population size, however, does not influence much of the MAS efficiency if it is not very small. When $r = 3$ and 8 offspring per mother are genotyped for a single marker (with four alleles at equal frequency) per chromosome (1 M) in a haploid genome of 20 chromosomes, for example, the increases in N_e are $\sim$ 28, 31, and 30% for male numbers being 5, 10, and 20, respectively.

DISCUSSION

In this study, an approach to optimizing within- and between-family selections simultaneously by using marker and pedigree information, and thus minimizing the rate of inbreeding or genetic drift for a small diploid population, was proposed and its efficiency was investigated by simulations. The new approach actually estimates and records the expected pedigree of each *chromosome* by using marker and pedigree (for individuals) information, and the *chromosome pedigree* is then used in a formal way to minimize the average coancestry among selected chromosomes by standard integer programming. The target of selection is chromosomes, while individuals are considered only as carriers of chromosomes and selection units. The approach is much more effective if the mating ratio is larger than one, compared with our previous MAS method that considered separately between-family selection on pedigree and within-family selection on marker information (Wang and Hill 2000).

It is also computationally simpler and more efficient compared with Toro *et al.*'s (1999) approach using the Monte Carlo Markov chains method for calculating the selection criterion. Because of the computational intensity, they considered only a specific example: a pig population with a family size of six, a mating ratio of three, and for a single chromosome of 100 cM. The increases in N_e of the specific population, calculated from the rates of inbreeding listed in their Table 3, are $\sim$40 and 70% when a single and two markers (each with four alleles) are used in the selection, respectively. In contrast, the corresponding values are $\sim$51 and 102% if the present approach is used. It is not clear why the approach proposed herein is more efficient. A direct comparison between the two approaches in a single simulation study would be helpful.

In the context of animal breeding, the probability of descent for a QTL allele (PDQ) conditional on linked marker information has been used to compute the conditional covariance of additive effects of the QTL alleles within and between individuals for the purpose of increasing the response to selection for a quantitative trait (Fernando and Grossman 1989; Wang *et al.* 1995). PDQ is similar to the conditional PIBD of alleles at an unknown locus linked to a marker, which is used in the present work to obtain (by integration) the average PIBDs between chromosomes and individuals. For multiple-marker loci, the marker linkage phase and the parental origin of marker alleles were assumed to be known for calculating PDQ (Goddard 1992), while these are inferred from the marker information of parents and offspring in this article. The average PIBD calculated herein also relates to the total allelic relationship of Nejati-Javaremi *et al.* (1997). The present approach could be adapted in the application to increase the short-term selection response by a more accurate estimate of realized genetic relationship and thus a better estimate of breeding values and to maintain genetic variation and thus increase the long-term selection response by constraining inbreeding to a certain low level or rate.

Although for convenience some simplified situations were considered in the simulation, the approach applies to a wide variety of complexities encountered in practice. These issues (*e.g.*, overlapping generations, nonrandom mating, unequal length of chromosomes, dominant markers, and different numbers and frequencies of marker alleles) as well as the potential impact on fitness and adaptation to captivity have been discussed in our previous investigation (Wang and Hill 2000). Throughout this article, I have assumed Haldane's mapping function. Any other mapping function (*e.g.*, one that allows interference and is perhaps more realistic than Haldane's mapping function) could be used without altering the general procedure and the results qualitatively. Equations 1 and 10–12, however, would have to be replaced by the appropriate integrals for averaging across the whole chromosome.

At present, the major obstacle to the practical application of MAS seems to be that there are few species with the necessary information on markers and their chromosomal distributions. Such information is, however, rapidly accumulating. It should be emphasized that, with MAS, the effective size varies over loci, depending on the location relative to marker. Loci situated close to markers have a much larger effective size than those far from markers. With the same (harmonic) mean increase in N_e by MAS, therefore, it is better to use many less informative markers scattered on a chromosome rather than a single marker with high heterozygosity.

I am grateful to Armando Caballero, Jesús Fernández, Bill Hill, Bill Jordan, Miguel Toro, and two anonymous referees for helpful comments.

LITERATURE CITED

Ballou, J. D., and R. C. Lacy, 1995 Identifying genetically important individuals for management of genetic variation in pedigreed populations, pp. 76–111 in *Population Management for Survival and Recovery. Analysis Methods and Strategies in Small Population Conservation*, edited by J. D. Ballou, M. Gilpin and T. J. Foose. Columbia University Press, New York.

Caballero, A., 1994 Developments in the prediction of effective population size. Heredity **73:** 657–679.

Caballero, A., and M. A. Toro, 2000 Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. **75:** 331–343.

Chevalet, C., and H. Rochambeau, 1986 Variabilité génétique et contrôle des souches non-consanguines. Sci. Tech. Anim. Lab. **11:** 251–257.

Crow, J. F., and M. Kimura, 1970 *Introduction to Population Genetics Theory*. Harper & Row, New York.

Fernández, J., and M. A. Toro, 1999 The use of mathematical programming to control inbreeding in selection schemes. J. Anim. Breed. Genet. **116:** 447–466.

Fernando, R. L., and M. Grossman, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467–477.

Frankham, R., 1995 Conservation genetics. Annu. Rev. Genet. **29:** 305–327.

Goddard, M. E., 1992 A mixed model for analyses of data on multiple genetic markers. Theor. Appl. Genet. **83:** 878–886.

Gowe, R. S., A. Robertson and B. D. H. Latter, 1959 Environment and poultry breeding problems. 5. The design of poultry control strains. Poultry Sci. **38:** 462–471.

Kimura, M., and J. F. Crow, 1963 On the maximum avoidance of inbreeding. Genet. Res. **4:** 399–415.

Lacy, R. C., 1995 Clarification of genetic terms and their use in the management of captive populations. Zoo Biol. **14:** 565–578.

Lacy, R. C., 1997 Importance of genetic variation to the viability of mammalian populations. J. Mammal. **78:** 320–335.

Nejati-Javaremi, A., C. Smith and J. P. Gibson, 1997 Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. **75:** 1738–1745.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes in Fortran 77*, Ed. 2. Cambridge University Press, Cambridge, UK.

Robertson, A., 1964 The effect of non-random mating within inbred lines on the rate of inbreeding. Genet. Res. **5:** 164–167.

Toro, M., L. Silió, J. Rodrigáñez and C. Rodriguez, 1998 The use of molecular markers in conservation programs of live animals. Genet. Sel. Evol. **30:** 585–600.

Toro, M., L. Silió, J. Rodrigáñez, C. Rodriguez and J. Fernández, 1999 Optimal use of genetic markers in conservation programs. Genet. Sel. Evol. **31:** 255–261.

Wang, J., 1997 More efficient breeding systems for controlling inbreeding and effective size in animal populations. Heredity **79:** 591–599.

Wang, J., and A. Caballero, 1999 Developments in predicting the effective size of subdivided populations. Heredity **82:** 212–226.

Wang, J., and W. G. Hill, 2000 Marker assisted selection to increase effective population size by reducing Mendelian segregation variance. Genetics **154:** 475–489.

Wang, T., R. L. Fernando, S. van der Beek, M. Grossman and J. A. M. van Arendonk, 1995 Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol. **27:** 251–274.

Communicating editor: J. B. Walsh