Theor Appl Genet (1991) 81:212–220

THEORETICAL AND APPLIED GENETICS

Quantitative genetic variance associated with chromosomal markers in segregating populations

J. C. M. Dekkers* and M. R. Dentine**

Department of Dairy Science, University of Wisconsin, Madison, WI 53706, USA

Received February 12, 1990; Accepted July 13, 1990
Communicated by D. Van Vleck

Summary. Use of chromosomal markers can accelerate genetic progress for quantitative traits in pedigree selection programs by providing early information on Mendelian segregation effects for individual progeny. Potential effectiveness of selection using markers is determined by the amount of additive genetic variance traced from parents to progeny by the markers. Theoretical equations for the amount of additive genetic variance associated with a marker were derived at the individual level and for a segregating population in joint linkage equilibrium. Factors considered were the number of quantitative trait loci linked to the marker, their individual effects, and recombination rates with the marker. Subsequently, the expected amount of genetic variance associated with a marker in a segregating population was derived. In pedigree selection programs in segregating populations, a considerable fraction of the genetic variance on a chromosome is expected to be associated with a marker located on that chromosome. For an average chromosome in the bovine, this fraction is approximately 40% of the Mendelian segregation variance contributed by the chromosome. The effects of interference and position of the marker on this expectation are relative small. Length of the chromosome has a large effect on the expected variance. Effectiveness of MAS is, however, greatly reduced by lack of polymorphism at the marker and inaccuracy of estimation of chromosome substitution effects. The size of the expected amount of genetic variance associated with a chromosomal marker indicates that, even when the marker is not the active locus, large chromosome substitution effects are not uncommon in segregating populations.

Key words: Genetic marker – Genetic variance – Marker-assisted selection – Quantitative traits

Introduction

In pedigree selection programs for quantitative traits in segregating populations, estimates of breeding values of parents are used for selection of progeny. These estimates provide information on conditional means of additive genetic values of gametes and progeny, but not on Mendelian segregation effects for individual gametes or progeny. According to established quantitative genetic theory, the breeding value of progeny is equal to the sum of the additive genetic values of uniting gametes, one from the female parent and one from the male. The additive genetic value of gametes produced by each parent has an expected value equal to one-half the breeding value of the parent and a variance equal to one-quarter of the genetic variance, due to Mendelian segregation. In pedigree selection programs, half of the genetic variance (within full sib families) is, therefore, unavailable for selection prior to availability of information on individual or progeny performance. In sire selection programs for sex-limited traits, the latter information is not available until males are older, e.g., 5 years of age in dairy cattle.

Availability of a polymorphic genetic marker, located on a chromosome segment (Geldermann 1975) containing one or more quantitative trait loci (QTLs), can provide early information on Mendelian segregation for one or both parents. For a parent heterozygous at the marker, which of the two homologous chromosome segments a progeny received can be determined from identification of the marker allele transmitted. Progeny can then be selected based on the marker received, if the

* *Current address:* Center for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada
** To whom correspondence should be addressed

marker is associated with a difference in expected breeding value for the selection goal. This form of selection is called marker-assisted selection (MAS), and has been discussed by Soller (1978), Soller and Beckmann (1982), Stam (1986), Smith and Simpson (1986), and Lande and Thompson (1990). Currently, several classes of molecular markers are available (Botstein et al. 1980; Jeffreys et al. 1985) that offer great potential for MAS.

MAS involves estimation of chromosome substitution effects (Geldermann 1975; Dentine and Cowan 1990) for parents heterozygous at the marker. A chromosome substitution effect can be defined as the average or expected difference in breeding value between progeny receiving one marker and those receiving the alternative marker from a common heterozygous parent (Dentine and Cowan 1990), and is due to differences in value between homologous alleles at one or more QTLs linked to the marker. If the marker is not itself a QTL, sign and size of the chromosome substitution effect depend on linkage phase and size of the linked QTL effects. In this case, unless the marker is in linkage disequilibrium with the QTLs, a separate effect must be estimated for each parent (Smith and Simpson 1986). Statistical procedures for estimating marker effects have been developed by Geldermann (1975), Soller and Beckmann (1982), Fernando and Wang (1989), Lander and Botstein (1989), and Dentine and Cowan (1990).

Potential effectiveness of MAS in pedigree selection programs is determined by the amount of additive genetic variance that can be traced by a marker from parent to offspring, and by the accuracy of estimation of chromosome substitution effects (Soller 1978). For each parent (male or female), the maximum amount of variance that can be utilized through MAS is the single-parent Mendelian segregation variance, or one-quarter of the genetic variance. When MAS is used for both parents, one-half of the genetic variance is available. The benefits from MAS in dairy cattle selection programs were examined by Soller and Beckmann (1982), Stam (1986), and Smith and Simpson (1986) for a given amount of genetic variance associated with a marker. Stam (1986) found that MAS could increase genetic response from sire selection in dairy cattle by up to 40% when the marker explained all of the sire's Mendelian segregation variance.

Little is known about the size and constitution of additive genetic effects and genetic variance associated with markers in segregating populations. Factors involved are the number of linked QTLs, the size of each QTL, and recombination rates with the marker. Soller et al. (1979) investigated the a priori distribution and composition of quantitative effects associated with a marker in crosses between inbred lines. In the current paper these aspects are examined for pedigree selection programs in segregating populations. Formulae for additive genetic variance associated with a marker are derived for individual parents and at the population level. Subsequently, the expected or a priori amount of population variance associated with a marker, with unknown position of the marker and QTLs, is derived and examined.

Theory

A marker can only trace segregation of QTLs located on the same chromosome. Consequently, for each parent (male or female), the maximum amount of genetic variance for which a marker can provide information is the single-parent Mendelian segregation variance for that chromosome (V_m). In a population in joint linkage equilibrium (Falconer 1981), V_m is one-quarter of the additive genetic variance contributed by the QTLs on the chromosome. In this section, theoretical aspects of the amount of variance that can be traced by a codominant marker are examined for parents heterozygous at the marker. Chromosome substitution effects are assumed to be estimated without error, and in each progeny the marker allele passed on by the parent is identified with certainty. Effects of parents homozygous at the marker and uncertainty of marker transmission are examined in a later section.

One QTL

Let locus A be the only QTL segregating on a chromosome that is marked at a polymorphic locus M. The recombination fraction between M and A is r_{MA}. Following Stam (1986), denote the i^{th} allele at the locus by A_i ($i=1, 2, \ldots, n_A$) and its average effect (Falconer 1981), by α_i. Let p_i be the population frequency of A_i. In a population in Hardy-Weinberg equilibrium, the additive genetic variance contributed by this QTL, and consequently by this chromosome, is $2\sum_{i=1}^{n_A} p_i \alpha_i^2$ (Kempthorne 1955), which is equal to $4V_m$.

Consider a parent heterozygous at both M and A: $M_x A_i/M_y A_j$, where "/" indicates separation between homologous chromosomes, and M_x and M_y indicate two alternative alleles at the marker locus. The breeding value of this individual for the marked chromosome is $\alpha_i+\alpha_j$. Gametes are produced containing either A_i or A_j in equal proportions, with additive genetic values α_i and α_j. With complete information on the breeding value of the parent, the estimated contribution of the QTL to the breeding value of a progeny without individual performance is $\frac{1}{2}(\alpha_i+\alpha_j)$ (pedigree value), regardless of which QTL allele the progeny received. However, based on knowledge of the marker allele received, gametes and offspring can be separated into two groups (Table 1). Offspring receiving the M_x marker can be assigned an expected breeding value of $(1-r_{MA})\alpha_i+r_{MA}\alpha_j$ and those receiving the M_y allele a value of $r_{MA}\alpha_i+(1-r_{MA})\alpha_j$. This amounts to deviations from the pedigree value of $\frac{1}{2}(\alpha_i+\alpha_j)$ by $\pm\frac{1}{2}(1-2r_{MA})(\alpha_i-\alpha_j)$ for the $M_x(+)$ and $M_y(-)$ marker. These deviations can be interpreted as estimates of Mendelian segregation for the marked chromosome. The chromosome substitution effect is given by $\mu_{x-y}=(1-2r_{MA})(\alpha_i-\alpha_j)$ (Soller 1978; Dentine and Cowan 1990).

Let $V_m(A_i/A_j)$ be the Mendelian segregation variance at the marked chromosome for genotype $M_x A_i/M_y A_j$. $V_m(A_i/A_j)$ is the variance among the two groups of gametes receiving the A_i or the A_j allele and is equal to $\frac{1}{4}(\alpha_i-\alpha_j)^2$. This variance is not utilized by pedigree selection. However, based on marker genotype, $V_m(A_i/A_j)$ can be partitioned into variance between marker alleles ($V_b(A_i/A_j)$), which can be utilized by MAS, and variance within marker alleles ($V_w(A_i/A_j)$), which remains unexplained. $V_b(A_i/A_j)$ is the variance in additive genetic value between the

Table 1. Gametes produced by individuals with genotype $M_x A_i/M_y A_j$

Marker allele	QTL allele	Frequency	Additive genetic value	Average additive genetic value	Average segregation effect [a]
M_x	A_i	$0.5\,(1-r_{MA})$	α_i	$(1-r_{MA})\,\alpha_i + r_{MA}\,\alpha_j$	$0.5\,(1-2r_{MA})\,(\alpha_i-\alpha_j)$
	A_j	$0.5\,r_{MA}$	α_j		
M_y	A_i	$0.5\,r_{MA}$	α_i	$r_{MA}\,\alpha_i + (1-r_{MA})\,\alpha_j$	$-0.5\,(1-2r_{MA})\,(\alpha_i-\alpha_j)$
	A_j	$0.5\,(1-r_{MA})$	α_j		

[a] Deviation from pedigree value of $0.5\,(\alpha_i+\alpha_j)$

two groups receiving the alternative marker alleles in Table 1 (Soller 1978). $V_w(A_i/A_j)$ is the variance among gametes within alternative marker alleles. From Table 1 these variances can be derived as:

$$V_b(A_i/A_j) = \tfrac{1}{4}(1-2r_{MA})^2(\alpha_i-\alpha_j)^2 = (1-2r_{MA})^2\,V_m(A_i/A_j), \quad (1)$$

$$V_w(A_i/A_j) = r_{MA}(1-r_{MA})\,(\alpha_i-\alpha_j)^2 = 4r_{MA}(1-r_{MA})\,V_m(A_i/A_j). \quad (2)$$

Note that $V_b(A_i/A_j)+V_w(A_i/A_j)=V_m(A_i/A_j)$.

From Eq. 1 it follows that, for a given genotype, the amount of variance that is traced by a marker from parent to progeny depends on the magnitude of the difference between the effects of the two alleles at the QTL and on the recombination fraction of the QTL with the marker. However, regardless of genotype at the QTL, the variance between marker alleles is a fraction $(1-2r_{MA})^2$ of the Mendelian segregation variance. Note that this fraction is zero when M and A are unlinked ($r_{MA}=\frac{1}{2}$) and unity when A is the same as the marker ($r_{MA}=0$).

Results obtained for individual genotypes can be extended to the population level. For a population in linkage equilibrium for M and A, frequency of the ordered genotype A_i/A_j is $p_i\,p_j$. The total single-parent Mendelian segregation variance in the population at the chromosome is:

$$V_m = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j\,V_m(A_i/A_j) = \tfrac{1}{2}\sum_{i=1}^{n_A} p_i\,\alpha_i^2 . \quad (3)$$

Equation 3 remains valid for the subpopulation of individuals heterozygous for the marker. From Eq. 1, variances between (V_b) and within (V_w) marker alleles for this subpopulation are:

$$V_b = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j\,V_b(A_i/A_j) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j(1-2r_{MA})^2\,V_m(A_i/A_j) = (1-2r_{MA})^2\,V_m , \quad (4)$$

$$V_w = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j\,V_w(A_i/A_j) = 4r_{MA}(1-r_{MA})\,V_m . \quad (5)$$

Thus, on the population level, as in the single-genotype case, the fraction of the Mendelian segregation variance at the marked chromosome that is available for MAS is equal to $(1-2r_{MA})^2$. In the next section, population parameters V_i from Eqs. 3, 4, and 5 will be denoted by $V_i(A)$ to indicate variances due to the A locus.

In a population in linkage equilibrium, the average chromosome substitution effect for a given set of marker alleles M_x and M_y is zero, since the chromosome substitution effect for the equally frequent genotypes $M_x A_i/M_y A_j$ and $M_x A_j/M_y A_i$ are of equal size but opposite sign, i.e.:

$$E(\mu_{x-y}) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j(1-2r_{MA})\,(\alpha_i-\alpha_j) = 0 . \quad (6)$$

The variance of chromosome substitution effects in the population is related to the variance between marker alleles, V_b, as follows:

$$\mathrm{Var}(\mu_{x-y}) = \sum_{i=1}^{n_A}\sum_{j=1}^{n_A} p_i\,p_j(1-2r_{MA})^2\,(\alpha_i-\alpha_j)^2 = 4V_b . \quad (7)$$

Two QTLs

Let B be an additional QTL segregating on the marked chromosome with recombination fraction r_{MB} with M and r_{AB} with A. Locus B has alleles B_i ($i=1, 2, \ldots, n_B$) with average effects β_i and frequencies q_i. Epistatic effects between A and B are assumed to be absent.

An individual with genotype $A_i B_k/A_j B_l$ produces gametes $A_i B_k$, $A_i B_l$, $A_j B_k$, and $A_j B_l$ with frequencies $\frac{1}{2}(1-r_{AB})$, $\frac{1}{2}r_{AB}$, $\frac{1}{2}r_{AB}$, and $\frac{1}{2}(1-r_{AB})$, and additive genetic values $\alpha_i+\beta_k$, $\alpha_i+\beta_l$, $\alpha_j+\beta_k$, and $\alpha_j+\beta_l$. The variance in additive genetic value among these gametes is the single-parent Mendelian segregation variance, denoted by $V_m(A_i B_k/A_j B_l)$, and can be expressed in the form of a variance due to each QTL independently and a covariance (C_m) between QTLs:

$$\begin{aligned} V_m(A_i B_k/A_j B_l) &= \tfrac{1}{4}(\alpha_i-\alpha_j)^2 + \tfrac{1}{4}(\beta_k-\beta_l)^2 + \tfrac{1}{2}(1-2r_{AB})\,(\alpha_i-\alpha_j)(\beta_k-\beta_l) \\ &= V_m(A_i/A_j) + V_m(B_k/B_l) + 2C_m(A_i B_k/A_j B_l) . \end{aligned} \quad (8)$$

In a population in joint genetic equilibrium, the frequency of the ordered genotype $A_i B_k/A_j B_l$ is $p_i\,p_j\,q_k\,q_l$. The total Mendelian segregation variance at the marked chromosome is:

$$\begin{aligned} V_m &= \sum_{i=1}^{n_A}\sum_{j=1}^{n_A}\sum_{k=1}^{n_B}\sum_{l=1}^{n_B} p_i\,p_j\,q_k\,q_l\,\{V_m(A_i/A_j) + V_m(B_k/B_l) + 2C_m(A_i B_k/A_j B_l)\} \\ &= V_m(A) + V_m(B) + 2\sum_{i=1}^{n_A}\sum_{j=1}^{n_A}\sum_{k=1}^{n_B}\sum_{l=1}^{n_B} p_i\,p_j\,q_k\,q_l\,C_m(A_i B_k/A_j B_l) \\ &= V_m(A) + V_m(B) . \end{aligned} \quad (9)$$

The last equality in Eq. 9 results from cancelation of covariance terms due to opposite linkage relationships; for example, $C_m(A_i B_k/A_j B_l) = -C_m(A_i B_l/A_j B_k)$ and their frequencies are equal.

Equation 9 shows that in the total population, the Mendelian segregation variance consists of independent contributions (Eq. 3) from the two QTLs, despite their linkage. The same result was obtained with a different approach by Kempthorne (1955) and Turner and Young (1969).

To obtain variances between and within markers for individual genotypes, initially no interference will be assumed, i.e., events of crossing-over in adjacent regions on the chromosome

Table 2. Gametes produced by individuals with genotypes $A_i M_x B_k / A_j M_y B_l$ and $M_x A_i B_k / M_y A_j B_l$

Marker allele	QTL haplotype	Frequency for gene order		Additive genetic value	Average additive genetic value	Average segregation effect [c]
		AMB [a]	*MAB* [b]			
M_x	$A_i B_k$	$0.5(1-r_{MA})(1-r_{MB})$	$0.5(1-r_{MA})(1-r_{AB})$	$\alpha_i+\beta_k$	$(1-r_{MA})\alpha_i+r_{MA}\alpha_j$ + $(1-r_{MB})\beta_k+r_{MB}\beta_l$	$0.5(1-2r_{MA})(\alpha_i-\alpha_j)$ + $0.5(1-2r_{MB})(\beta_k-\beta_l)$
	$A_i B_l$	$0.5(1-r_{MA})r_{MB}$	$0.5(1-r_{MA})r_{AB}$	$\alpha_i+\beta_l$		
	$A_j B_k$	$0.5r_{MA}(1-r_{MB})$	$0.5r_{MA}r_{AB}$	$\alpha_j+\beta_k$		
	$A_j B_l$	$0.5r_{MA}r_{MB}$	$0.5r_{MA}(1-r_{AB})$	$\alpha_j+\beta_l$		
M_y	$A_i B_k$	$0.5r_{MA}r_{MB}$	$0.5r_{MA}(1-r_{AB})$	$\alpha_i+\beta_k$	$r_{MA}\alpha_i+(1-r_{MA})\alpha_j$ + $r_{MB}\beta_k+(1-r_{MB})\beta_l$	$-0.5(1-2r_{MA})(\alpha_i-\alpha_j)$ + $-0.5(1-2r_{MB})(\beta_k-\beta_l)$
	$A_i B_l$	$0.5(1-r_{MA})r_{MB}$	$0.5(1-r_{MA})r_{AB}$	$\alpha_i+\beta_l$		
	$A_j B_k$	$0.5(1-r_{MA})r_{MB}$	$0.5(1-r_{MA})r_{AB}$	$\alpha_j+\beta_k$		
	$A_j B_l$	$0.5(1-r_{MA})(1-r_{MB})$	$0.5(1-r_{MA})(1-r_{AB})$	$\alpha_j+\beta_l$		

[a] $r_{AB}=r_{MA}+r_{MB}-2r_{MA}r_{MB}$
[b] $r_{MB}=r_{MA}+r_{AB}-2r_{MA}r_{AB}$
[c] Deviation from pedigree value of $0.5(\alpha_i+\alpha_j+\beta_k+\beta_l)$

are independent. This assumption is later relaxed. Also, two gene orders are considered: *AMB* and *MAB*.

Gene order AMB. Gametes produced by individuals with genotype $A_i M_x B_k / A_j M_y B_l$ are shown in Table 2. The chromosome substitution effect consists of independent contributions from the QTLs: $\mu_{x-y}=(1-2r_{MA})(\alpha_i-\alpha_j)+(1-2r_{MB})(\beta_k-\beta_l)$. Similar to the one-locus case, the Mendelian segregation variance, $V_m(A_i B_k/A_j B_l)$, can be partitioned into a variance between ($V_b(A_i B_k/A_j B_l)$) and within marker alleles ($V_w(A_i B_k/A_j B_l)$). Using the relationship $r_{AB}=r_{MA}+r_{MB}-2r_{MA}r_{MB}$, corresponding to no interference (Haldane 1919), these variances can be expressed in terms of variances due to each QTL independently and a covariance term:

$$\begin{aligned}V_b(A_i B_k/A_j B_l) &= \tfrac{1}{4}(1-2r_{MA})^2(\alpha_i-\alpha_j)^2+\tfrac{1}{4}(1-2r_{MB})^2(\beta_k-\beta_l)^2\\ &\quad+\tfrac{1}{2}(1-2r_{AB})(\alpha_i-\alpha_j)(\beta_k-\beta_l)\\ &= (1-2r_{MA})^2 V_m(A_i/A_j)+(1-2r_{MB})^2 V_m(B_k/B_l)\\ &\quad+2C_m(A_i B_k/A_j B_l),\end{aligned} \tag{10}$$

$$\begin{aligned}V_w(A_i B_k/A_j B_l) &= r_{MA}(1-r_{MA})(\alpha_i-\alpha_j)^2+r_{MB}(1-r_{MB})(\beta_k-\beta_l)^2\\ &= 4r_{MA}(1-r_{MA})V_m(A_i/A_j)+4r_{MB}(1-r_{MB})V_m(B_k/B_l).\end{aligned} \tag{11}$$

Gene order MAB. Gametes produced by an individual with genotype $M_x A_i B_k / M_y A_j B_l$ are also shown in Table 2. Using the relationship $r_{MB}=r_{MA}+r_{AB}-2r_{MA}r_{AB}$, corresponding to no interference for this situation, the chromosome substitution effect is equivalent to that for gene order *AMB*. Variances between and within markers are:

$$\begin{aligned}V_b(A_i B_k/A_j B_l) &= (1-2r_{MA})^2 V_m(A_i/A_j)+(1-2r_{MB})^2 V_m(B_k/B_l)\\ &\quad+2(1-2r_{MA})^2 C_m(A_i B_k/A_j B_l),\end{aligned} \tag{12}$$

$$\begin{aligned}V_w(A_i B_k/A_j B_l) &= 4r_{MA}(1-r_{MA})V_m(A_i/A_j)+4r_{MB}(1-r_{MB})V_m(B_k/B_l)\\ &\quad+8r_{MA}(1-r_{MA})C_m(A_i B_k/A_j B_l).\end{aligned} \tag{13}$$

Equations 10 and 12 show that, regardless of gene order, the variance between marker alleles contain a fraction $(1-2r)^2$ of the Mendelian sampling variance for each QTL independently, where r is the recombination fraction between QTL and marker. This is identical to the one QTL case (Eq. 1). However, the variance between markers also contains a fraction of the covariance between QTLs that contributes to the Mendelian sampling variance. This proportion depends on gene order.

Total population. Variances between and within markers for the subpopulation of individuals heterozygous for the marker, assuming joint genetic equilibrium, are obtained from Eqs. 10 through 13 by pooling across genotypes. Similar to Eq. 9, covariance terms in Eqs. 10 through 13 cancel when summing over genotypes, and only the independent contributions of the QTLs remain. Regardless of gene order, variances between and within markers in the population simplify to:

$$V_b=(1-2r_{MA})^2 V_m(A)+(1-2r_{MB})^2 V_m(B)=V_b(A)+V_b(B), \tag{14}$$

$$\begin{aligned}V_w &= 4r_{MA}(1-r_{MA})V_m(A)+4r_{MB}(1-r_{MB})V_m(B)\\ &= V_w(A)+V_w(B).\end{aligned} \tag{15}$$

Similar to the one QTL case (Eqs. 6 and 7), the mean and variance of chromosome substitution effects in a population in joint linkage equilibrium are equal to zero and $4V_b$.

Interference. Given r_{MA} and r_{MB}, interference with respect to crossing-over affects only the covariance terms in Eqs. 10 through 13. Variance terms remain unchanged. Direction and size of changes in covariance terms depend on the amount of interference and gene order. Variances in the total population (Eqs. 11 and 12) are also unaffected by interference, since all covariance terms cancel.

More than two QTLs.

Equations derived in the previous section for two QTLs can readily be extended to more than two QTLs by adding terms for variance due to each additional QTL, and covariance terms between the additional QTL and each original QTL. With t QTLs segregating on the chromosome, denoted by QTL_i ($i=1, 2, \ldots, t$), with recombination fractions r_i with the marker, variances in the total population of individuals heterozygous for the marker, assuming joint linkage equilibrium and no epistatic effects, are given by:

$$V_m=\sum_{i=1}^{t} V_m(QTL_i) \tag{16}$$

where $V_m(QTL_i)$ is given by Eq. 3,

$$V_b = \sum_{i=1}^{t} (1-2r_i)^2 V_m(QTL_i), \quad (17)$$

$$V_w = \sum_{i=1}^{t} 4r_i(1-r_i) V_m(QTL_i). \quad (18)$$

Expected variance between markers

Consider a chromosome of length L Morgans with a marker situated L_1 and L_2 Morgans from the ends of the chromosome ($L_1+L_2=L$). Assume one QTL (A) is segregating on the chromosome. Then V_b, the variance between markers for the subpopulation of parents heterozygous for the marker, is a fraction $(1-2r_{MA})^2$ of the Mendelian segregation variance (V_m), as shown previously. Without knowledge of r_{MA}, prior information can be used to obtain an expected value of $(1-2r_{MA})^2$. This prior information consists of the prior probability density function $f(x)$ of the map distance x (Morgans) between marker and QTL, and a mapping function $r=g(x)$, which relates recombination fractions r to map distances x.

Given $f(x)$ and $g(x)$, the expectation of V_b is given by the following integral:

$$E(V_b) = \int_{x=0}^{L_1} f(x)(1-2g(x))^2 V_m \, dx + \int_{x=0}^{L_2} f(x)(1-2g(x))^2 V_m \, dx. \quad (19)$$

If the QTL can be located anywhere on the chromosome with equal probability, and map distances are proportional to physical distances, then $f(x)$ is uniform and equal to $1/L$. Two extreme mapping functions, those for no and complete interference, are considered (Haldane 1919). For no interference, $g(x)=\frac{1}{2}(1-e^{-2x})$, and the solution to Eq. 19 is:

$$E(V_b) = \frac{1}{4L}\{(1-e^{-4L_1})+(1-e^{-4L_2})\} V_m. \quad (20)$$

With complete interference, $g(x)=x$ when $x\leq\frac{1}{2}$, and $g(x)=\frac{1}{2}$ when $x\geq\frac{1}{2}$. Consequently, V_b is zero when x is larger than $\frac{1}{2}$ Morgan. Letting L_1 be the shorter of the two arms, solutions to Eq. 19 are:

if $L_1\leq\frac{1}{2}$ and $L_2\leq\frac{1}{2}$:

$$E(V_b) = \left\{\frac{L_1}{L}(\tfrac{4}{3}L_1^2-2L_1+1)+\frac{L_2}{L}(\tfrac{4}{3}L_2^2-2L_2+1)\right\} V_m;$$

if $L_1\leq\frac{1}{2}$ and $L_2\geq\frac{1}{2}$:

$$E(V_b) = \left\{\frac{L_1}{L}(\tfrac{4}{3}L_1^2-2L_1+1)+\frac{1}{6L}\right\} V_m;$$

if $L_1\geq\frac{1}{2}$ and $L_2\geq\frac{1}{2}$:

$$E(V_b) = \frac{1}{3L} V_m. \quad (21)$$

With multiple QTLs segregating on the chromosome, Eqs. 20 and 21 remain valid, because of the independence of QTLs in a population in linkage equilibrium. In this case, Eqs. 20 and 21 apply to each QTL individually, as well as to the entire chromosome, with V_b and V_m interpreted as variances contributed by, respectively, an individual QTL and the total chromosome. Therefore, regardless of the number of QTLs segregating on the chromosome, the expectation of the fraction of the variance on the chromosome that can be traced by a single marker depends only on the length of the chromosome, the position of the marker, and the amount of interference. $E(V_b)$ is equal to the actual variance when genetic variance on the chromosome is determined by the usual quantitative model of an infinite number of QTLs, each with an equal contribution (Appendix).

Relationships of expected variance between markers for the subpopulation of individuals heterozygous at the marker with chromosome length, marker position, and interference are illustrated in Figs. 1 and 2. $E(V_b)$ decreases with increasing length of the chromosome (Fig. 1). When the marker is at the center and there is no interference, $E(V_b)$ ranges from 0.63 V_m for chromosomes of 0.5 Morgans to 0.17 V_m for chromosomes of 3 Morgans. Interference decreases $E(V_b)$ (Fig.1). With complete interference, $E(V_b)$ is reduced by up to 0.1 V_m, or 30%. The effect of interference is largest for chromosomes of 1–1.5 Morgans and negligible for chromosomes smaller than 0.5 Morgans. $E(V_b)$ is maximum when the marker is at the center of the chromosome and decreases as the marker is positioned further away from the center by up to 0.2 V_m (Figs. 1 and 2). The effect of the position of the marker is largest for shorter chromosomes. For all chromosome lengths, the rate of reduction in $E(V_b)$ increases with the

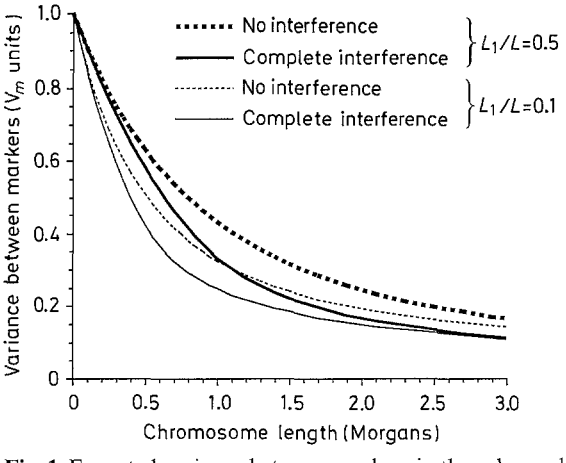

Fig. 1. Expected variance between markers in the subpopulation heterozygous for a marker located L_1 Morgans from the end of a chromosome of L Morgans. Graphs are shown with the marker at the center ($L_1/L=0.5$) or 0.1 L Morgans from the end ($L_1/L=0.1$) of the chromosome. Single-parent Mendelian segregation variance at the chromosome is V_m

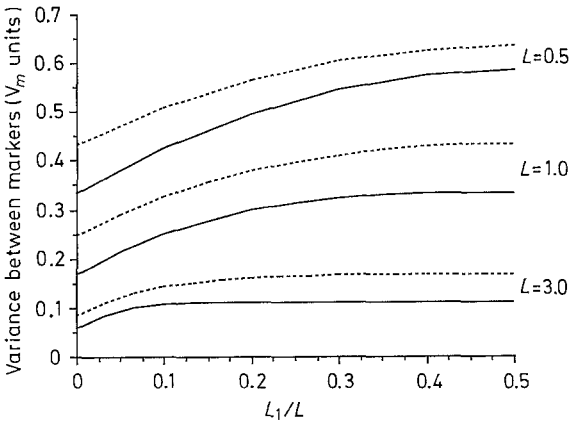

Fig. 2. Effect of position of the marker on expected variance between markers in the subpopulation heterozygous for a marker located L_1 Morgans from the end of a chromosome of L Morgans. – – – – no interference; —— complete interference. Single parent Mendelian segregation variance at the chromosome is V_m

Table 3. Identification of marker allele received by offspring of an $M_1^* M_2^*$ parent[a] for a marker with two alleles M_1, M_2 with frequencies m_1, m_2 in the population of mates

Genotype mate	Frequency	Genotype offspring	Frequency	Allele received from $M_1^* M_2^*$ parent identified	
				Mates genotyped	Mates not genotyped
$M_1 M_1$	m_1^2	$M_1^* M_1$	0.50	yes	yes
		$M_1 M_2^*$	0.50	yes	no
$M_1 M_2$	$2m_1 m_2$	$M_1^* M_1$	0.25	yes	yes
		$M_1^* M_2$	0.25	no	no
		$M_1 M_2^*$	0.25	no	no
		$M_2^* M_2$	0.25	yes	yes
$M_2 M_2$	m_2^2	$M_1^* M_2$	0.50	yes	no
		$M_2 M_2^*$	0.50	yes	yes

[a] Alleles of the common parent labelled with * for ease of interpretation

distance of the marker from the center (Fig. 2); differences in $E(V_b)$ are small between markers located within 0.25 L from the center. This comprises 50% of markers located at random on the chromosome.

With a total genome size of 30 Morgans and 30 chromosome pairs (Fries and Ruddle 1986), the average chromosome length for the bovine is approximately 1 Morgan. With no interference, $E(V_b)$ for a chromosome of 1 Morgan is 43% of V_m for a central marker (Fig. 1). Little is known about map lengths of individual chromosomes in the bovine. However, with respect to physical length, the smallest and largest chromosomes are respectively 50% smaller and 65% larger than the average (Lin et al. 1977). If physical lengths are directly related to map lengths, the latter may range from 0.5 to 1.65 Morgans.

Effect of marker polymorphism

The variance between markers, V_b, examined in previous sections, considered only the subpopulation of parents heterozygous at the marker, and assumed marker transmission could be determined with certainty for all progeny. In the total population, the amount of variance that can be utilized by MAS for a single marker is, therefore, only a fraction of V_b (Smith and Simpson 1986). The size of this fraction is determined by the extent of polymorphism at the marker, and will be named Polymorphism Information Content (PIC), after Botstein et al. (1980). For a population in joint linkage equilibrium, PIC is examined for a marker with n alleles M_i ($i=1, 2, \ldots, n$) with frequencies m_i. All alleles are assumed to be codominant (i.e., individuals heterozygous at the marker can be distinguished). First, PIC for two alleles at the marker ($n=2$) is derived.

Two alleles at the marker

Mates genotyped. For parents heterozygous at the marker (frequency $2m_1 m_2$), V_b cannot be utilized for progeny for which marker transmission from the parent cannot be determined. When mates are genotyped (G), this comprises a fraction $m_1 m_2$ of the offspring, of which one-half received the M_1 marker and one-half the M_2 allele (Table 3). Consequently, the fraction of V_b that can be utilized for heterozygous parents is $1-m_1 m_2$. Adding the absence of marker information for homozygous parents, the fraction of V_b that can be utilized when the total population is considered is:

$$\mathrm{PIC}_G = 2m_1 m_2 (1-m_1 m_2)\,. \tag{22}$$

Mates not genotyped. When mates are not genotyped (NG), marker transmission can be determined with certainty only for progeny homozygous at the marker (Table 3). One-half of the offspring of randomly mated heterozygous parents are homozygous, regardless of gene frequencies. Therefore,

$$\mathrm{PIC}_{NG} = 2m_1 m_2 \left(\tfrac{1}{2}\right) = m_1 m_2\,. \tag{23}$$

However, of the heterozygous progeny from heterozygous parents, a fraction m_2 received the M_1 marker and a fraction m_1 the M_2 marker (Table 3). This information can be used to obtain an estimate of the average segregation effect for these progeny. For a parent with chromosome substitution effect μ_{1-2}, this estimate is $m_2(\frac{1}{2}\mu_{1-2}) + m_1(-\frac{1}{2}\mu_{1-2}) = (m_2 - m_1)\frac{1}{2}\mu_{1-2}$. The amount of Mendelian segregation variance that the marker accounts for in these progeny is $(m_2-m_1)^2 V_b$. For all progeny of heterozygous parents, the usable variance is $\{\frac{1}{2}+\frac{1}{2}(m_2-m_1)^2\} V_b$

$=(1-2m_1m_2)\,V_b$. Consequently,

$$PIC_{NG'}=2m_1m_2(1-2m_1m_2). \quad (24)$$

For two alleles at the marker, PIC reduces the amount of variance available for MAS considerably (Fig. 3). PIC is maximum (0.375 for PIC_G and 0.25 for PIC_{NG} and $PIC_{NG'}$) when gene frequencies are equal, and decreases as gene frequencies diverge from $\frac{1}{2}$. Knowledge of genotype of mates (PIC_G versus $PIC_{NG'}$) increases PIC, regardless of gene frequency. However, the difference between PIC_G and $PIC_{NG'}$ is small when gene frequencies are extreme. When mates are not genotyped, knowledge of gene frequencies to estimate a segregation effect for heterozygous progeny ($PIC_{NG'}$ versus PIC_{NG}) is most beneficial when gene frequencies are ca. 0.25 or 0.75.

Multiple alleles at the marker

Mates genotyped. For progeny of a heterozygous parent, marker transmission is uncertain when marker genotype of the progeny is identical to that of both parents. This amounts to one-half of the progeny from matings between parents with identical marker genotypes. The frequency of matings between parents with genotype M_iM_j is $(2m_im_j)^2$. Subtracting an additional term for parents homozygous at the marker (frequency $\sum_{i=1}^{n} m_i^2$), PIC in the total population is:

$$PIC_G=1-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2m_i^2m_j^2-\sum_{i=1}^{n} m_i^2. \quad (25)$$

PIG_G in Eq. 25 is equivalent to the index derived by Botstein et al. (1980) for the fraction of offspring that is informative in determining linkage of a marker to a disease locus.

Mates not genotyped. When mates of a heterozygous parent are not genotyped, marker transmission is uncertain for progeny that have the same marker genotype as the parent. This is a fraction $\frac{1}{2}(m_i+m_j)$ of the progeny of a parent with genotype M_iM_j; $\frac{1}{2}m_j$ receiving the M_i marker and $\frac{1}{2}m_i$ the M_j marker. Without use of the latter information, the variance between markers that can be utilized for heterozygous parents is $\{1-\frac{1}{2}(m_i+m_j)\}\,V_b$, and

$$PIC_{NG}=1-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} m_im_j(m_i+m_j)-\sum_{i=1}^{n} m_i^2. \quad (26)$$

When an average segregation effect is assigned to progeny with the same marker genotype as the parent, according to the proportion of each marker, the usable variance between markers for M_iM_j parents is $\left\{1-\frac{2m_im_j}{m_i+m_j}\right\}V_b$, and

$$PIC_{NG'}=1-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{4m_i^2m_j^2}{m_i+m_j}-\sum_{i=1}^{n} m_i^2. \quad (27)$$

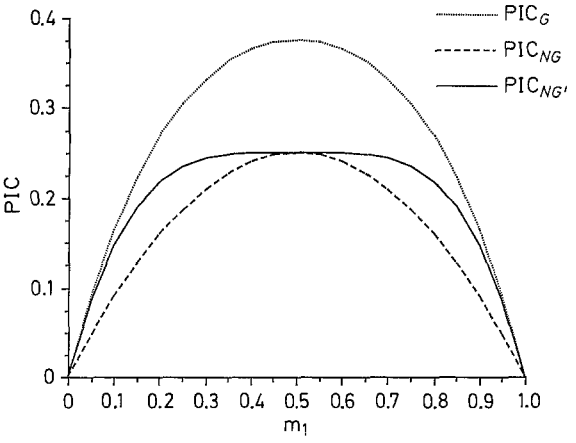

Fig. 3. Effect of gene frequency (m_1) at a diallelic marker locus on PIC. PIC_G = mates genotyped; PIC_{NG} = mates not genotyped and gene frequency unknown; $PIC_{NG'}$ = mates not genotyped but gene frequency known

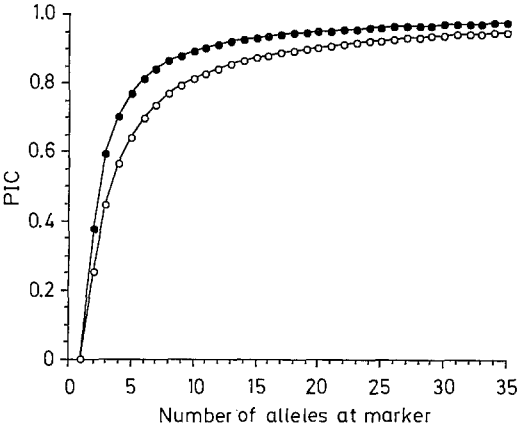

Fig. 4. Effect of number of alleles at a marker locus on PIC when all alleles have equal frequency, depending on genotyping of mates. —●— mates genotyped; —o— mates not genotyped

PIC increases rapidly with number of alleles at the marker. This is shown in Fig. 4 for alleles with equal frequency ($1/n$), in which situation all PICs are maximum, and PIC_{NG} is equal to $PIC_{NG'}$. Benefit from genotyping mates decreases as the marker becomes more polymorphic (Fig. 4).

Discussion

Before individual estimates of genetic parameters for specific markers are available, prior knowledge can be used to determine the potential for accelerating genetic progress with MAS. For pedigree selection programs in cattle, prior estimates of the amount of genetic variance associated with a single marker on a chromosome of average length (1 Morgan) are 30–40% of the Mendelian segregation variance at the chromosome (V_m), depending

on interference and position of the marker. Therefore, with one highly polymorphic marker on each of the 30 chromosomes pairs, 30–40% of the total Mendelian segregation variance may be explained.

The actual value of the variance associated with a single marker may range from zero, when all QTLs have a recombination rate of 50% with the marker, to V_m, when all QTLs on the chromosome are at the marker. However, as the number of QTLs segregating on the chromosome increases, deviations of the actual variance from its expectation are expected to become smaller. When the number of QTLs is infinite and each QTL has an equal contribution to genetic variance, the actual variance becomes equal to its expectation (Appendix).

Derivation of the a priori variance between markers was based on several assumptions. First, position of QTLs with respect to the marker was assumed random. In the search for markers, the expectation of the variance between markers can be increased by searching for polymorphisms within or around a known active gene, as was done by Cowan et al. (1990) for the prolactin gene in dairy cattle. In Figs. 1 and 2 this is equivalent to decreasing the effective length of the chromosome, which results in a higher expected variance between markers. Second, recombination was assumed to be homogeneous along the chromosome. However, "hotspots" for recombination have been found (Chakravarti et al. 1986). Location of a marker near a hotspot would decrease expected variance between markers.

Derivations at the population level were under the assumption of joint linkage equilibrium. Directional selection, which is practiced in most livestock species, induces linkage disequilibrium, resulting from a negative covariance between QTLs (Bulmer 1971). When the marker is not a QTL, the assumption of linkage equilibrium between the marker and each QTL remains appropriate. Consequently, derivations for one QTL on the chromosome are not affected by selection. However, with multiple-linked QTLs on the chromosome, both V_m and V_b may be reduced by selection. Further study is needed.

PIC decreases the available variance between markers by over 60% when there are two alleles at the marker. For an average chromosome in the bovine, the expected population variance between markers is then less than 15% of V_m. However, with an increasing number of alleles at the locus, PIC becomes less of a factor. Development of unique markers with DNA fingerprinting techniques (Jeffreys et al. 1985) and combination of markers into haplotypes (Lander and Botstein 1989) deserve further attention. With those techniques, PIC approaches unity. Alternatively, the number of progeny from a mating could be increased to such an extent (e.g., with superovulation and embryo transfer in cattle) that selected progeny have inherited the superior chromosome segment with certainty.

Utilization of MAS in segregating populations to full potential requires accurate estimation of chromosome substitution effects for each parent. This aspect has not been addressed in the present study, but could reduce effectiveness of MAS considerably (Soller and Beckmann 1982; Smith and Simpson 1986). Accurate estimation of chromosome substitution effects requires evaluation of a large number of progeny per parent for the marker and the quantitative trait [>1,000 daughters in dairy cattle (Soller 1978; Soller and Beckmann 1982)]. Progeny groups of such magnitude are available for sires in dairy cattle progeny testing schemes; however, costs associated with genotyping each daughter for the marker are large (Soller and Beckmann 1982).Weller et al. (1988) and Dentine and Cowan (1990) proposed using progeny tests of sons for estimation of chromosome substitution effects in dairy sires. With this approach, only a limited number of sons per sire needs to be genotyped, but accurate estimation may require use of information from all relatives.

Chromosome substitution effects must be reestimated each generation, depending on recombination rates between the marker and QTLs. With negligible recombination rates, estimates of chromosome substitution effects can be used in subsequent generations and selection can be directly on marker genotype (Soller 1978).

The presented theory does not preclude the existence of considerable marker-linked effects when many QTLs with only small effects are linked to the marker. When the marker is not a QTL, the distribution of chromosome substitution effects in a population in joint linkage equilibrium has a mean of zero (Eq. 6) and a variance equal to four times the variance associated with the marker (Eq. 17). With an expected variance between markers of 40% of V_m, the expected variance of chromosome substitution effects is, therefore, 1.6 V_m, or 0.4 times the additive genetic variance on the chromosome. Given a nonuniform distribution of total additive genetic variance across chromosomes (Paterson et al. 1988), large individual chromosome substitution effects would, therefore, not be uncommon in a segregating population.

Acknowledgements. This research was supported by a grant from a consortium of artificial insemination organizations including 21st Century Genetics, Sire Power, Inc., Atlantic Breeders Cooperative, and Noba, Inc.

Appendix

Variance between markers for an infinite number of QTLs with small effect

Consider $2n$ QTLs (n large) equally spaced along a chromosome of length L (Morgans) with a marker at the center. Let V_m be the single parent Mendelian segregation variance at the chromosome and let all QTLs have an equal contribution of $\frac{1}{2n} V_m$ to this variance. QTLs are numbered in pairs according to their

distance x_i from the marker. For the i^{th} pair, $x_i = \frac{iL}{2n}$. Then V_b, the variance between marker alleles for the subpopulation of parents heterozygous for the marker, is:

$$V_b = \sum_{i=1}^{n} \{1-2g(x_i)\}^2 \frac{1}{n} V_m, \quad \text{where } g(x) \text{ is the mapping function.}$$

For no interference, using $g(x)=\frac{1}{2}(1-e^{-2x})$ (Haldane 1919),

$$V_b = \sum_{i=1}^{n} e^{-4x_i} \frac{1}{n} V_m .$$

Let Δx be the map distance between two adjacent QTLs: $\Delta x = \frac{L}{2n}$, such that $\frac{1}{n} = \frac{2}{L}\Delta x$. Substitution in the previous equation gives:

$$V_b = V_m \frac{2}{L} \sum_{i=1}^{n} e^{-4x_i} \Delta x .$$

Taking the limit of V_b for n at infinity, and realizing that x_1 and Δx tend to zero as n goes to infinity and that $x_n = L/2$:

$$V_b = \lim_{n\to\infty} \left(V_m \frac{2}{L} \sum_{i=1}^{n} e^{-4x_i} \Delta x \right)$$

$$= V_m \frac{2}{L} \int_{x=0}^{L/2} e^{-4x} dx = \frac{1}{2L}(1-e^{-2L}) V_m .$$

References

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bulmer MG (1971) The effect of selection on natural variability. Am Nat 105:201–211

Chakravarti A, Elbein SC, Permutt MA (1986) Evidence for increased recombination near the human insulin gene: implication for disease association studies. Proc Natl Acad Sci USA 83:1045–1049

Cowan CM, Dentine MR, Ax RL, Schuler LA (1990) Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. Theor Appl Genet 79:577–582

Dentine MR, Cowan CM (1990) An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous at a segregating marker locus. Theor Appl Genet 79:775–780

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman, London

Fernando RL, Wang TW (1989) Marker-assisted selection using best linear unbiased prediction. J Dairy Sci 72 [Suppl. 1]:31

Fries R, Ruddle FH (1986) Gene mapping in domestic animals. In: John JS (ed) 10th Beltsville Symp Agric Res: Biotechnology for solving agricultural problems, vol 10, pp 19–37

Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. 1. Methods. Theor Appl Genet 46:319–330

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

Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific "fingerprints" of human DNA. Nature 316:76–79

Kempthorne O (1955) The theoretical values of correlations between relatives in random mating populations. Genetics 40:153–167

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

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lin CC, Newton DR, Church RB (1977) Identification and nomenclature for G-banded bovine chromosomes. Can J Genet Cytol 19:271–282

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Smith C, Simpson SP (1986) The use of genetic polymorphisms in livestock improvement. J Anim Breed Genet 103:205–217

Soller M (1978) The use of loci associated with quantitative effects in dairy cattle improvement. Anim Prod 27:133–139

Soller M, Beckmann JS (1982) Restriction fragment length polymorphisms and genetic improvement. In: Proc 2nd World Congr Genet Appl Livestock, Madrid, pp 396–404

Soller M, Brody T, Genizi A (1982) The expected distribution of marker-linked quantitative effects in crosses between inbred lines. Heredity 43:179–190

Stam P (1986) The use of marker loci in selection for quantitative characters. In: Smith C, King JWB, McKay JC (eds) Exploiting new technologies in animal breeding genetic developments. Oxford University Press, Oxford, pp 170–182

Turner HN, Young SY (1969) Quantitative genetics in sheep breeding. Macmillan, Melbourne

Weller JI, Kashi Y, Soller M (1988) Estimation of the sample size necessary for genetic mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 71 [Suppl. 1]:142