Theor Appl Genet (2001) 102:1107–1112

R. van Berloo · P. Stam

Simultaneous marker-assisted selection for multiple traits in autogamous crops

Received: 10 September 1999 / Accepted: 24 August 2000

Abstract A method is presented for the selection of parents with the aim of obtaining improved genotypes in the progeny of a cross. The procedure is designed to select in several unrelated traits simultaneously and is based on the selection of molecular markers that are linked to QTLs. The method was compared with conventional phenotypic selection in simulation experiments for a number of genetic structures underlying the traits and several types of parental populations. Although the method in general provides good results, some of the underlying assumptions may be violated quite easily, thereby reducing the applicability of the procedure in practice.

Keywords Computer simulation · Marker-assisted selection · Quantitative trait loci

Introduction

Since most agronomically important traits are quantitative and controlled by several genetic factors, the ability to obtain more control of the behaviour of these genetic factors by the introduction of linked molecular markers has been very welcome. In recent years substantial contributions have been made to improve the identification of the loci that underlie important quantitative traits (quantitative trait loci, QTLs). Although QTL mapping methods remain an object of continued study and improvement (e.g. Henshall and Goddard 1999), several fairly reliable procedures have been established and implemented (e.g. Lincoln et al. 1992; Holloway and Knapp 1994; Van Ooijen and Maliepaard 1996). The next issue to be addressed is the efficient use of information on QTLs, which is now readily becoming available for many crops and populations. Several simulation studies have been published on the efficiency of using QTL and marker information for selection (Gimelfarb and Lande 1994a, b; Hospital et al. 1997; Romagosa et al. 1999). Most studies show that marker-assisted selection (MAS) yields an improved selection result in continued selection for several generations, especially in the first generations. A combined index of marker and phenotypic information typically yielded the best response (Lande and Thompson 1990). Experimental results of applying MAS have been discussed in several papers (e.g. Stuber 1995; Van Berloo and Stam 1999). Both Moreau et al. (1998) and Van Berloo and Stam (1998) argued that population size and trait heritability are the key factors influencing MAS results. Tanksley (1997) advocated a slightly different use of marker and QTL information. He proposed a selective enrichment of the gene pools currently used for the production of commercial varieties with minor QTL alleles that still reside undetected in wild relatives or unadapted germplasm. Tanksley and Nelson (1996) described a procedure for the simultaneous discovery and introgression of QTLs from unadapted germplasm. A successful application of this method was presented by Bernacchi et al. (1998a, b). Nevertheless, conventional breeding has shown continued success while using elite material, demonstrating that there is still room for improvement even within the currently used germplasm. In this paper we try to evaluate how a knowledge of QTL positions can be of use to speed up and increase selection results. The procedure used in this study builds on the method described in our earlier simulation and experimental work (Van Berloo and Stam 1998, 1999) and extends this work beyond one trait. We also applied it here to other common population types besides recombinant inbred lines (RILs).

Communicated by H.C. Becker

R. van Berloo (✉) · P. Stam
C.T. de Wit Graduate school for Resource Conservation
and Production Ecology
Laboratory of Plant Breeding,
Wageningen University and Research Centre, P.O. Box 386,
6700 AJ Wageningen, the Netherlands
e-mail: Ralph.vanBerloo@pv.dpw.wag-ur.nl
Fax: +31-317-483457

Table 1 Specification of the genetic design

Trait	Heritability	Number of QTLs	Effect QTL-1	Effect QTL-2	Effect QTL-3	Effect QTL-4
Trait 1	0.7	3	3.0	2.0	1.0	
Trait 2	0.5	2	2.0	1.0		
Trait 3	0.2	4	4.0	3.0	2.0	1.0
Trait 4	0.1	4	4.0	3.0	2.0	1.0
Trait 5	0.1	4	4.0	3.0	2.0	1.0

Materials and methods

The procedure used for marker-assisted selection uses available information on markers and QTLs in a mapping population. Information on QTL-flanking markers is used to assess pairs of lines for their ability to give rise, in a progeny derived after crossing the line pair, to genotypes with accumulated beneficiary QTL alleles. Such genotypes can be called 'superior' genotypes. Progeny of line pairs that were selected in this way were compared to progeny obtained by crossing parents that were selected using phenotypic selection. This procedure for MAS of parental pairs has been implemented in a computer package. The previously implemented selection method (Van Berloo and Stam 1998) was modified and extended in three areas: (1) selection was applied to several unrelated traits simultaneously, with individual trait values being combined into a single index value by assigning weights to each individual trait; (2) increased algorithm efficiency and computing power reduced the need for a 'pre-selection' of possible promising line-pairs; (3) the selected objects can also consist of a combination of three or four lines (in these cases a pre-selection may still be required).

Genetic architecture

The starting point in this simulation study are mapping populations in which QTLs have already been identified and located on the genome. Thus, we start with a set of plants or lines that have been genotyped with respect to markers. For each QTL the probability (p_m) that the advantageous allele is present is inferred from the genotypes of the flanking markers. The distance between flanking markers, i.e. the size of the QTL supporting marker interval, is also used in this assessment. The 'genetic architecture' for each population was such that a number of traits were segregating. We simulated the segregation of 17 QTLs, affecting five traits, according to the specification given in Table 1. The genome consisted of ten chromosome pairs, each 100 cM in length. The QTLs were dispersed randomly over the genome. Also, for each locus, the parent contributing the advantageous allele was selected at random. In each simulation experiment three to five of these randomly created genomes were used to avoid bias caused by a randomly created genome.

Trait weighing

A weight value was assigned to each trait, and through weighed summing over traits a general index was created (Eq. 1).

$$GI = \sum_{Traits} (W_T \times \sum_{QTLs} (p_m \bullet q_{T,I}))$$

where GI is the general index value, W_T is the weight factor for trait T, p_m is the probability of the presence of the favourable allele, given flanking marker genotypes and $q_{T,I}$ is the QTL effect of QTL_I affecting trait T.

In most simulations all values for W_T were set to 1 – i.e. all traits were considered equally important – but in other cases unequal weights were assigned to each trait. It should be noted that the above index is different from the commonly used indices in index selection theory that predict maximum genetic gain with selection (e.g. Yamada et al., 1975).

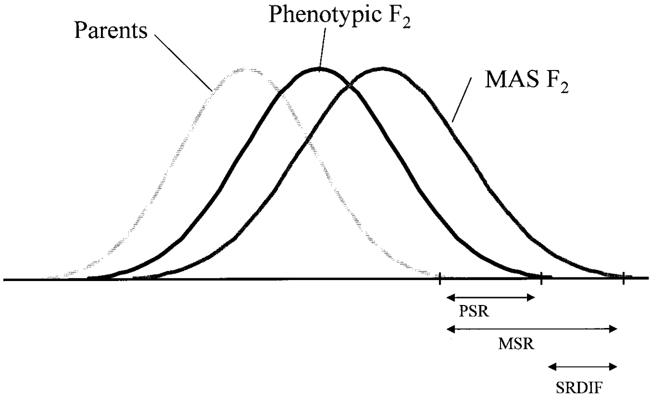


Fig. 1 Different measures used in the evaluation of selection success. *PSR*. Phenotypic Selection response, *MSR* MAS selection response; *SRDIF* Selection response difference

Selection procedure

The marker-assisted selection procedure is started in a mapping population in which markers have been scored and QTLs identified. The genetic constitution of each plant with regard to the QTLs under study is inferred from QTL-flanking markers. Because many QTLs are involved, the probability that, in an average-sized population, all advantageous QTL alleles will be present in a single plant or line is very small. However, the chance that such an individual will be present among the progeny of a well-selected pair of lines is fairly high. Theoretically, it would be possible, using a probabilistic approach, to determine which line-pairs are the most promising. Here we followed a different approach: for all possible line-pairs a test cross is simulated, and from this F_1 a progeny of sufficient size is derived by selfing. The most superior genotype that is observed among the resulting F_2 population of size 100 is recorded and an average over five or ten replicates of this value is used as the selection parameter, giving an indication of the potential quality of a line-pair. This value is compared with the most superior parental genotype present among the lines. We define the selection response (MSR, Fig. 1) as the difference between these two values, usually expressed as a percentage. The selection response obtained by applying MAS is then compared with the selection response obtained through phenotypical selection (PSR, Fig. 1), and the difference (SRDIF, Fig. 1) was chosen to be the main parameter for evaluation of selection success.

The phenotype for each individual trait was derived from the trait-genotype, supplemented with a random error term to represent environmental noise. The size of the error term was derived from the trait heritability and the observed genetic variance among the parental lines. In this way phenotypic values for all traits were determined. Next, Eq. 1 was applied to obtain the 'phenotypic' value of the general index.

A procedure, similar to one that could be used to improve existing elite material for a single trait, was used to simulate phenotypic selection: line combinations were made by combining the lines with the highest phenotype for the general index with lines that showed the highest phenotype for a particular trait. These line-pairs were then processed in the same way as the MAS derived line-pairs. A set of 100 line-pairs was selected in this way.

Types of populations

In our previous paper (Van Berloo and Stam 1998) we focussed on the effects of heritability and population size in a RIL population.

Table 2 Comparison of MAS and phenotypic selection results: left column, difference between MAS response and phenotypic selection response; right column; MAS selection response. All traits were weighed equally

	***n*=50**		***n*=100**		***n*=200**	
RILs	6.7%	11.9%	6.8%	10.6%	11.2%	10.4%
DH	5.6%	6.5%	7.5%	11.2%	7.5%	10.8%
BC_1	8.1%	15.4%	7.3%	16.9%	8.5%	17.7%
F_2	6.9%	21.6%	10.8%	27.9%	12.1%	23.6%

Table 3 Comparison of MAS and phenotypic selection results: left column, difference between MAS response and phenotypic selection response; right column, MAS selection response. Trait weights were chosen inversely proportional to trait heritability

	***n*=50**		***n*=100**		***n*=200**	
RILs	7.1%	19.3%	4.3%	7.1%	5.9%	4.2%
DH	3.1%	2.0%	5.2%	4.4%	6.2%	3.5%
BC_1	9.0%	25.4%	9.4%	23.9%	16.5%	29.9%
F_2	7.6%	21.9%	9.1%	21.2%	7.8%	20.3%

In this paper we also discuss simulations of the application of MAS to F_2, BC_1 and doubled haploid (DH) populations. Parental populations were composed of 50, 100 or 200 plants/lines. When populations were used that were still to a large extent heterozygous (F_2, BC_1), and therefore able to produce a larger variety of gametes, the number of replications was doubled to account for the higher genetic sampling variance.

In the first experiment MAS and phenotypic selection were compared for the different population types assuming that all traits could be considered equally important (all traits received equal weights). In the second, similar experiment, some traits were regarded as being more important than others. The size of the trait-weight was inversely proportional to the heritability of the trait, thereby assuming that traits that are more difficult with regard to selection because they inherit with a lower heritability are also more important.

Undetected QTLs

In most QTL mapping studies, some QTLs are detected, but even when the same populations are used, different QTLs may be found in replicated trials (e.g. Beavis 1999). Some QTLs, also of larger effects, will remain undetected due to the limited detection power available in common mapping populations. This limitation is mainly due to the population size. Accurate mapping of many QTLs depends on the occurrence of rare crossovers. Since this is a process of chance, only very large populations are likely to contain individuals in which several rare crossover events have occurred. We studied the effect of missing QTL information due to incomplete QTL mapping by means of data removal. Selection was based on a subset of the QTLs, but genotype and phenotype were still constructed using all 17 QTLs. A random subset was removed from the list of detected QTLs, rendering selection for these QTLs through linked markers impossible. This was repeated five times to avoid bias. Again, the same procedure for determining the quality of pairs of lines was applied, and results were compared with phenotypic selection. Simulations were run for cases in which three, five, seven or nine QTLs had been deleted from the list of detected QTLs.

Results

Table 2 shows the results for experiment 1, with the use of equal weights for all traits. The left columns show the observed difference in selection response after applying MAS and phenotypic selection; the right columns show the selection response obtained by using MAS. The response is impressive in the F_2 and BC_1 parental populations. It is likely that the amount of heterozygosity still present in these lines is responsible for this success. Apparently MAS is very effective in taking advantage of the larger amount of available genetic diversity present in more heterozygous population types.

In all cases MAS outperforms phenotypic selection, as was expected. When RILs or F_2 plants are used for parents, a larger parental population increases the difference between MAS and phenotypic selection, i.e. marker-assisted selection uses the extra genetic diversity present in larger populations more efficiently.

The results described above were obtained assuming equal importance of all traits. However, most of the time some traits can be regarded to be more important than others. Such a situation was reflected by the second simulation experiment. The results of this experiment are displayed in Table 3; the same layout as for Table 2 was used for displaying the results. For weighed trait-selection, we expected MAS to show an extra benefit, since lower heritability traits, which can be selected better by MAS than by phenotypic selection, are considered to be more important. However, the results do not confirm this expectation. Results of MAS seem to drop for the more homozygous population types (DH, RIL), while only the results for the BC_1 population are better than the situation with equal trait-weights. We concluded, in line with Hospital *et al.* (1997), that the focus, which is usually put on low heritability when application of MAS is considered, does not have to be justified in more complex case dealing with several unrelated but genetically linked traits.

Undetected QTLs

Simulations that involve MAS based on incomplete QTL data were run for the RIL and BC_1 populations. The results of these simulations are summarised in Figs. 2 and 3. In Fig. 2 the observed selection response is plotted as a function of the QTL fraction. The QTL fraction is the proportion of the QTLs that were detected and used by the MAS procedure. As expected, the lower QTL-fractions result in decreased selection results, eventually

Fig. 2 MAS Selection response as a function of the "QTL fraction" (fraction of QTLs, present in the model, that were linked to markers

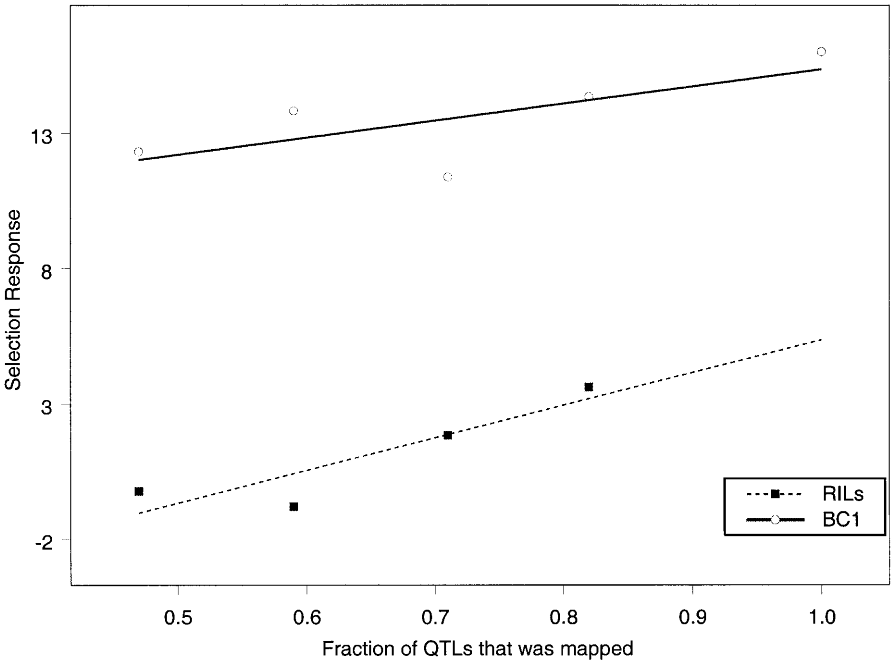
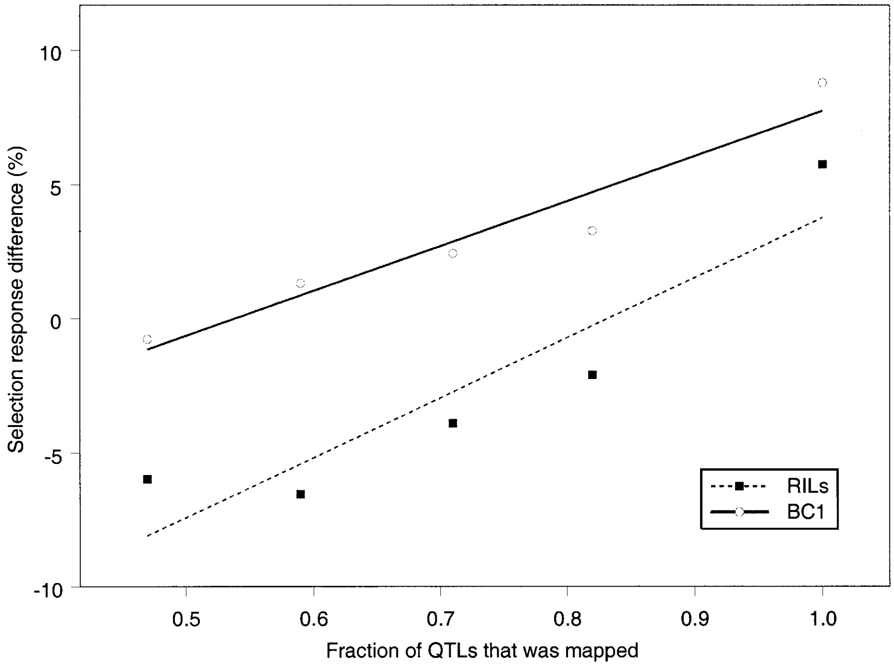


Fig. 3 Difference in selection response between MAS and phenotypic selection as a function of the "QTL fraction" (fraction of QTLs, present in the model, that were linked to markers



reaching the point where selection and crossing do not yield better genotypes than the genotypes already present in the parental population. This point is reached earlier for RILs than for BC_1 populations. In Fig. 3 the difference in selection response between MAS and phenotypic selection is plotted as a function of the QTL fraction. This gives us information in which cases application of MAS may yield better results than phenotypic selection, even when QTL information is incomplete. For RILs we see that a small number of 'missed' QTLs already has a profound influence on the efficiency of MAS. If more than 20% of the acting QTLs are missed MAS may already become less efficient than phenotypic selection. This effect is also seen for BC_1 populations, but to a lesser extent.

Discussion

With the ever-increasing amount of genomic information becoming available to breeders and scientists ways must be found to exploit this information in order to obtain more efficient methods for breeding and selection. In this paper we discuss a method that is based on molecular markers that are linked to target genes. The method is able to predict superior parental combinations, with regard to the genotype of their offspring, for several traits of interest. In general, the proposed method using marker-assisted selection gives better selection results than selection based on phenotype. It appears that the best results are obtained in populations that are heterozygous by nature. However, such populations are difficult to maintain and reproduce, thereby reducing the practical value of this observation. Still, in some cases vegetative propagation of heterozygous material could be an option for a successful application of the discussed method.

The genetic design, used in this study, is meant to represent a typical situation in which a breeder has to deal with several traits, each trait being inherited with a different heritability and controlled by several QTLs of unequal effect. In contrast with our previous simulation work (Van Berloo and Stam 1998) and the work of other authors, this design does not assume that interest lies solely with low heritability traits (i.e. a favourable situation with respect to the application of MAS). Hospital et al. (1997) noted that the response to MAS for low heritability traits, although more favourable, also is more variable than the response at higher heritabilities. In addition, the ability to select in later generations without any reevaluation of phenotypes may make MAS more attractive for high heritability traits in economic terms, even when direct selection *improvements* will be small.

In this paper we assume a given and fixed heritability for each trait. The value for the heritability is used, together with the genetic variance observed in the parental population, to obtain a value for the environmental (error) variance. For a given set of QTL effects the resulting genetic variance will differ between population types. Also, heritability is not a fixed quantity, by repeating and enlarging trails a higher heritability is often observed. However, we have chosen not to correct for deviations of our initial assumptions about trait heritability since in this study the environmental variance is used only for the creation of parental phenotypes from the genotypes, while the phenotype is used solely for phenotypic selection of potential parents.

Another point we discuss is the effect of incomplete QTL information. Beavis (1999) found up to 60 QTLs in a very large experiment. When using a subset of the data, representative in size to a commonly used mapping population, he only detected about 15 QTLs. This example illustrates the common knowledge shared by quantitative geneticists that any single QTL study will usually not be conclusive. Beavis therefore suggested the pooling of available experimental results to get a better power of QTL detection. We showed that the sensitivity of the proposed method to missing QTL information reduces its practical value. Only when extremely good molecular and field data are available and QTLs can be mapped reliably, so that only a small fraction of the genetic variance remains unaccounted for, can one expect real benefits from this type of selection. On the other hand, in many ‘difficult’ types of populations (e.g. species with a long juvenile period, a long generation time or traits that are difficult to measure) a procedure like the one described in this paper may be employed successfully, and other factors besides increased selection efficiency may be important to consider.

A possible way to compensate for QTLs that have gone unnoticed in the mapping procedure is to combine marker information and phenotypic values into the index in a way similar to the index proposed by Lande and Thompson (1990). This method basically assigns weights to markers and phenotype relative to the proportion of variance explained by the markers. Such an approach will be the subject of a future study.

Another factor that may limit the application of MAS in practice is the type of population being used. The method assumes the availability of a mapping population derived from a single cross. In general, a breeder will use material from diverse sources and origins. A strategy that might be followed to benefit from MAS while using non-mapping types of breeding populations could be to take two distinct members out of the elite gene-pool used for breeding and use these to create a new mapping population. MAS could then be applied to this mapping population, and the final result after selection, which will be similar to an improved version of the original elite material, could then be used to replace this material in a conventional breeding programme.

This paper only deals with the selection of line-pairs, but the model was extended to also allow the selection of combinations of three or four lines. Simulations that included three of four line combinations were run, but any difference with the results from selected line-pairs was barely discernible (data not shown). This is because the selection of line-pairs was already quite successful in accumulating superior sets of QTLs. In most simulations the most superior member of the progeny had obtained the advantageous allele for 16 out of 17 QTLs (either in homozygous or heterozygous state). Adding an extra line to the procedure therefore does not add much, although it significantly increases the number of required calculations (and in practice, the breeding work). Still, in more complex cases, when more traits and QTLs are involved, the exploration of sets of three or four lines may be a fruitful exercise.

In these simulations QTLs are assumed to act additively. However, in heterozygous populations it is also possible to detect QTLs with a dominant effect. Previous studies have shown a larger advantage of MAS over phenotypic selection when dominant QTL alleles play a role. But, since the final goal is to obtain homozygous genotypes that contain accumulated advantageous QTL alleles, dominance effects would be lost in the end. One could think of similar selection strategies in order to predict pairs of parents for the production of a hybrid variety. In such a case dominance would be very important and the expected benefits of MAS are expected to be larger than those observed in this study.

Another complicating factor, interaction between QTLs, is usually neglected, although a recent study by Charmet et al. (1999), using an approach very similar to the approach presented in this paper, did take QTL interactions into account. To date most QTL mapping software is not yet equipped to detect QTL interactions. However, more and more information on genes that underlie QTLs will be obtained; for instance, from genome sequencing projects. We expect that interaction between QTLs and also QTL× Environment interaction will become more important in the future. A method based on the selection of sets of genes through linked markers may be an efficient way to make sure that favourably interacting sets of genes are brought together or remain together. This is another aspect in which the extra information on markers and linked genes that currently can be made available can be put to use.

References

Beavis WD (1999) Integrating the central dogma with plant breeding. Abstr Plant Anim Genome 7:28

Bernacchi D, Beck-Bunn T, Eshed Y, lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998a) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. Theor Appl Genet 97:318–397

Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D, Tanksley S (1998b) Advanced backcross QTL analysis of tomato II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:170–180

Gimelfarb A, Lande R (1994a) Simulation of marker assisted selection in hybrid populations. Genet Res 63:39–47

Gimelfarb A, Lande R (1994b) Simulation of marker assisted selection for non-additive traits. Genet Res 64:127–136

Henshall JM, Goddard ME (1999) Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. Genetics 151:885–894

Holloway JL, Knapp SJ (1994) G-Mendel users guide. Oregon State University, Corvallis, Ore.

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181–1189

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

Lincoln S, Daly M, Lander E (1992) Mapping genes controlling quantitative traits with MAPMAKER/QTL. Whitehead Institute Technical Report, 2nd edn, Whitehead Institute, Mass.

Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-assisted selection efficiency in populations of finite size. Genetics 148:1353–1365

Romagosa I, Han F, Ullrich SE, Hayes PM, Wesenberg DM (1999) Verification of yield QTL through realized molecular marker-assisted selection responses in a barley cross. Mol Breed 5:143–152

Stuber CW (1995) Successes in the use of molecular markers for yield enhancement in corn. In: Proc 49th Annu Corn Sorghum Industry Res Conf. Am Seed Trade Assoc, Washington, D.C., pp 232–238

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksely SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063–1066

Van Berloo R, P Stam (1998) Marker assisted selection in autogamous RIL populations: a simulation study. Theor Appl Genet 96:147–154

Van Berloo R, P Stam (1999) Comparison between marker-assisted selection and phenotypical selection in a set of *Arabidopsis thaliana* recombinant inbred lines. Theor Appl Genet 98: 113–118

Van Ooijen JW, Maliepaard C (1996) MAPQTL™ version 3.0: software for the calculation of QTL positions on genetic maps. CPRO-DLO, Wageningen

Yamada Y, Yokouchi K, A Nishida (1975) Selection index when genetic gains of individual traits are of primary concern. Jpn J Genet 50:33–41