# Use of molecular markers in plant breeding

# Het gebruik van moleculaire merkers in de plantenveredeling

Promotor: dr. ir. P. Stam
Hoogleraar in de plantenveredeling,
in het bijzonder de selectiemethoden
en duurzame resistentie

Ralph van Berloo

# Use of molecular markers in plant breeding

**Proefschrift**

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
dr C.M. Karssen

In het openbaar te verdedigen
op woensdag 5 april 2000
des namiddags te vier uur in de Aula

*The experiments described in this thesis were performed at the Laboratory of Plant Breeding and at the Laboratory of Genetics, Wageningen University, The Netherlands*

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Berloo, R. van

**Use of molecular markers in plant breeding**

Thesis Wageningen University - with references - with summaries in English and Dutch.
Laboratory of Plant Breeding. P.O. Box 386, 6700 AJ, Wageningen, The Netherlands

ISBN 90-5808-167-2

**Key words:** plant breeding, selection methods, quantitative genetics, molecular markers, quantitative trait locus, introgression, marker-assisted selection, computer simulation, barley, *arabidopsis*

Printed by Ponsen & Looijen, Wageningen

*"...I hear babies cry*
*I watch them grow*
*They'll learn much more*
*than I'll ever know..."*

**Louis Amstrong**: What a Wonderful World

# Contents

# 1 Introduction

## The need for agriculture

In the past decades the number of people employed in agriculture has decreased considerably, especially in the 'developed' countries. Currently, a small part of the labour force is able to produce food and other agricultural products to provide the entire population. A huge increase in the production efficiency of human labour in agriculture can be seen over the years. Three factors are held responsible for the increase of agricultural production per hectare:

1. Overall increase in efficiency due to the merger of farms and a more efficient re-distribution of farm land.
2. More efficient cultural practises and better quality and availability of inputs (fertilizer, pesticides).
3. Genetic improvement of crops.

Thus far the horror scenarios foreseen by Malthus (1798) have not come true, and agricultural production has managed to keep up with the growing demands. At this moment over 6 billion people need to be fed and it is expected that in the year 2050 more than 10 billion people will inhabit the earth (FAO, 1996). Moreover, demands per capita will rise when standards of living in developing counties improve. As a consequence there will be a huge increase in the demand for food, and production will need to triple in the coming 40 years (Bindraban, 1997; WRR, 1995). A continued effort to secure agricultural production in the future will therefore be vital. Reduction of losses caused by a-biotic and biotic stress will be a key issue (Visser, 1999). Scenario studies on world food security generally assume that a large increase in production will be achieved by genetic improvement of crop species and biotechnology (FAO, 1996; Agrevo, 1996). History has indeed shown a continuous increase in crop yields, resulting from plant breeding efforts. However, in the light of the speed at which the human

population develops, and taking into account the expected reduction of available arable land due to climate change and human intervention, it may be necessary to accelerate the rate at which genetic improvement is achieved. Modern biotechnology provides new tools that can facilitate development of improved plant breeding methods and augment our knowledge of plant genetics. The knowledge that is obtained with these new tools can be used to contribute to an enhanced food security throughout the world. The study presented in this thesis focusses on the use of some of the modern biological tools for the improvement and acceleration of genetic crop improvement.

## Plant Breeding

Plant Breeding is a dynamic area of applied science. It relies on genetic variation and uses selection to gradually improve plants for traits and characteristics that are of interest for the grower and the consumer. Practical breeding of many economic important crops is performed by commercial companies that strive in a fierce competition for the favour of agricultural producers and consumers (Zuurbier, 1994). This competition has ensured a continued improvement of cultivated varieties (cultivars) over the past decades. These improvements were partly realised through an efficient use of existing variability, present within the available material. Another important way of improvement is the introduction of new genetic material (e.g. genes for disease resistance) from other sources, such as gene bank accessions and related plant species. In Europe the legislative system of "Breeders Rights" allows plant breeders to use genetic material that has been released by competitors in their own breeding program. This has contributed to the sharing of beneficiary genetic material among varieties but, most likely, it has also reduced the overall variability within the gene pool used for breeding. Although current breeding practises have been very successful in producing a continuous range of improved varieties, recent developments in the field of biotechnology and molecular biology can be employed to enhance plant breeding efforts and to speed up the creation of cultivars. Also, new ways and methods that allow an easier introduction of genetic material from related and unrelated plant species, without the drawbacks that are normally associated with the introduction of "wild genes" through conventional methods, become feasible.

## Quantitative traits

Some of the most difficult tasks of plant breeders relate to the improvement of traits that show a continuous range of values. Among such quantitative traits are important traits like yield, plant length and days to flowering (speed of plant development). Selection for quantitative traits is difficult, because the relation between observed trait values in the field (the phenotype) and the underlying genetic constitution (the genotype) is not straightforward. Quantitative traits are typically controlled by many genes that each contribute only a small part to the observed variation. The environmental variance resulting from differences in growing conditions further obscures the relation between phenotype and genotype. In practice, this problem is typically dealt with by evaluating large, replicated trials, which allow identification of genotypic differences through statistical analysis. Plant breeders would like to get a better grip on quantitative traits by direct selection for the genetic factors that are responsible for the observed variability in quantitative traits. This can be achieved through indirect selection: selection for other, well recognisable factors, that are linked to the target genes. Molecular markers, derived from recent bio-technological developments, can be used for this purpose.

## Molecular Markers

The discovery of restriction enzymes (Smith & Wilcox, 1970) and the polymerase chain reaction (PCR; Mullis & Faloona, 1987) have created the opportunity to visualise the composition of organisms at the DNA level, and obtain a so-called genetic fingerprint (e.g. Kearsey & Pooni, 1996). The visualisation is routinely performed by the separation, on a gel, of DNA-fragments that result from a selective digestion with enzymes or fragments that result from a selective amplification using PCR. DNA-fragments that result in different gel patterns between samples or individuals are called polymorphic markers. The visible differences on the gel result from differences at the DNA level. Not all types of markers are the same, the information content depends on the method that was used to obtain the marker data and the population in which the markers were 'scored'. For instance, it is not always possible to distinguish genome fragments that are present in homozygous condition from heterozygous fragments. In a heterogeneous population like an $F_2$, co-dominant markers like RFLPs (Botstein *et al.* 1980) and co-dominantly scored AFLPs (Vos *et al.* 1995) yield more information than dominant markers like RAPDs

(Welsh & McCleland, 1990) and dominantly scored AFLPs. Advanced tools for the retrieval of marker data and the subsequent analysis have been developed and allow a quick and reliable analysis in most plant species. These developments have opened up a new era for genetics and selection (Gallais & Charcosset, 1994; Moreau *et al.* 1998). Important information on the genetic background of individual plants and populations can be derived from linkage that is observed between markers.

## Genetic Linkage Maps

Segregation analysis can be applied to a segregating population that is derived from a common set of ancestors. Markers that co-segregate (are always present or absent together) must be linked, i.e. they must be located in each others vicinity on the genome. In some cases however, due to recombination events, the linkage between the markers may be lost. The frequency with which the linkage between co-segregating markers is broken is an indication of the genetic distance between the markers. An extensive analysis of the linkage between a large number of molecular markers yields information on their arrangement on the genome. Such analysis can finally result in the construction of a genetic map, on which all markers are arranged in separate linkage groups or chromosomes. On such a map the distances between markers reflect the degree of observed linkage. Genetic linkage maps should not be confused with physical genomic maps, which can be obtained by determining the DNA sequence of chromosomes, as is currently being done in several genome mapping projects. Linkage maps and physical maps are related, but this relation is usually not linear (e.g. Schmidt *et al.* 1995). Nowadays, software for the calculation of genetic maps has brought marker analyses, aimed at the construction of genetic maps, within the reach of many scientists.

## QTL analysis

Genetic factors that are responsible for a part of the observed phenotypic variation for a quantitative trait can be called quantitative trait loci (QTLs). Although similar to a gene, a QTL merely indicates a region on the genome, and could be comprised of one or more functional genes (Falconer & Mackay, 1996). In a process called QTL-mapping association between observed trait values and presence/absence of alleles of markers that have been mapped onto a linkage map is analysed. When it is significantly clear

that the correlation that is observed did not result from some random process, it is proclaimed that a QTL is detected. Also the size of the allelic effect of the detected QTL can be estimated. A breeder can analyse QTL occurrences and use this knowledge to his advantage, for instance by using indirect selection. When selection is (partly) based on genetic information retrieved through the application of molecular markers this is called marker-assisted selection.

## Marker-assisted selection

Marker-assisted selection (MAS), sometimes also called marker-aided selection, is a relative new tool for plant breeders. In its simplest form it can be applied to replace evaluation of a trait that is difficult or expensive to evaluate. When a marker is found that co-segregates with a major gene for an important trait, it may be easier and cheaper to screen for the presence of the marker allele linked to the gene, than to evaluate the trait. From time to time the linkage between the marker and the gene should then be verified. When more complex, polygenic controlled traits are concerned, the breeder is faced with the problem how to combine as many as possible beneficiary alleles for the QTLs that were detected. In this case the breeding material can be screened for markers that are linked to QTLs. Based on such an analysis, specific crosses can be devised for the creation of an optimal genotype, combining beneficiary QTL alleles from different sources. This situation, which is the main subject of this thesis, could also be called *marker-assisted breeding*. Successful practical application of marker-assisted selection was described by Stuber and Sisco (1992), Stuber (1994), Huang *et al.* (1997), Romagosa *et al.* (1999) and others. Marker-assisted selection, when applied within the current breeding material to enhance a breeding program, does not solve the problem of limited genetic variability that is often seen in breeding stocks (Tanksley & McCouch, 1997). A different application of marker-assisted selection could contribute to a genetic enrichment of breeding material. Marker-assisted selection may be used to facilitate a controlled inflow of new genetic material. 'Wild' or unadapted material often carries desired components that may be missing in cultivated material. Such components can be transferred to elite cultivated material by repeated backcrossing. However, breeders are often reluctant to apply this method because of unpredictable side-effects. These are caused by other genes, which are unintentionally transferred along with the genes that

control the target trait. It may take considerable effort and screening to get rid of the unwanted genes and return the material to an acceptable agronomic value. Markers can be used to pinpoint the genetic factors that are responsible for the desired characteristics in the unadapted material. In a backcross program, the presence of the desired QTL-alleles can be verified continuously by observing linked markers. At the same time, and with little extra effort, markers provide information on the origin of the remaining genome, allowing selection within the backcross material for genotypes that have lost the majority of unwanted donor DNA. Usually the application of this *marker-assisted back-cross* procedure will also result in a reduction of the number of backcross generations that are required, thereby speeding up the breeding program.

## Objectives and outline of the present study

At present, the conditions for application of molecular marker data and derived information in plant breeding are good. High throughput techniques have made the acquirement of marker data faster and cheaper. Several well founded algorithms and procedures for the analysis of molecular data have been implemented, while more developments in this field, especially originating from the field of animal and human genetics, may be expected in the future (e.g. Hoeschele *et al.* 1997) The continuing growth in capacity and power of modern computers has also brought computationally complex analyses within reach. The objective of this study was therefore aimed at the next step. Assuming economically affordable methods exist for obtaining marker information. Assuming furthermore that reliable methods exist that use this marker information for the determination of genetic factors, underlying important traits in cultivated crops. How can this information be used in an efficient way for the improvement of plant breeding methods?

Similar questions were raised and explored by Lande and Thompson (1990), Edwards and Page (1994), Gimelfarb and Lande (1994a,b;1995), Whittaker *et al.* (1995), Luo *et al.* (1997), Moreau *et al.* (1998) and others. These authors focussed mainly on the application of marker-assisted selection in a program aimed at a continuous improvement of populations. One of the most important and difficult questions in plant breeding relates to the selection of suitable lines or genotypes as crossing parents (Van Oeveren,

1993; Stam, 1995; Schut 1998). The present study focusses largely on marker-assisted selection of parental combinations, the final goal of the selection experiments being a single improved genotype, found among the progeny of selected parental combinations. In a practical breeding situation such an improved genotype could become a new culti var, in the case of self-fertilizing crops, or a new elite breeding line, in the case of hybrid varieties. In the following chapters different aspects related to the research question are discussed.

In chapter two the construction of a simulation model is presented. The simulation model allows selection of pairs of parents based on QTL information. For a single trait in an autogamous crop the model generates a selection of suitable crosses, aiming at the 'stacking' of desirable QTL alleles. Using simulated sets of data, the quality and performance of the model is evaluated and the influence of several genetic parameters on the selection result is investigated. These investigations are continued in chapter three. In this chapter a greenhouse experiment using *Arabidopsis thaliana* is described, aimed at verification of simulation results. Within a set of recombinant inbred lines, divergent selection for pairs of lines was performed aiming at a short time to flowering as well as a long time to flowering. Practical results obtained from marker-based selection were compared with the results of phenotypic selection.

In chapter four an extension to the model that was introduced in chapter two is presented. Options for simultaneous selection for several traits were investigated by computer simulations. These simulations explored the application of marker-assisted selection of combinations of lines as parents in a complex cross, for a wide range of popular types of populations. Simulation results obtained using MAS were compared with results achieved through phenotypic selection. Also the effect of violations of underlying assumptions and the implications for application of such a procedure in practice are discussed. In chapter five a modest verification experiment, based on the results obtained in chapter four, is described. Marker-assisted selection of parents was applied for several traits simultaneously in a set of *Arabidopsis* RILs. Experimental observations on populations that resulted from computer predicted crosses, and

observations on populations resulting from phenotypic selection are presented and discussed.

A different application of markers in plant breeding is the subject of chapter six. In this chapter a practical experiment of marker-assisted backcrossing in barley is described, aimed at the construction of near isogenic lines for QTLs that confer partial resistance to barley leaf rust. The efficiency of a range of marker-assisted backcrossing strategies was explored using computer simulations and the results are compared with the experimental data. In chapter seven the development of a computer tool is described, which assists in marker based selection and which was used to perform the selection discussed in chapter six. Finally, chapter eight presents a general discussion on the implications and expectations of marker-assisted selection, marker-assisted backcrossing and marker-assisted breeding.

# 2 Marker-assisted selection in autogamous RIL populations: a simulation study

Ralph van Berloo · Piet Stam

*Published in Theoretical and Applied Genetics 96:147-154 (1998)*

## Introduction

The advent of molecular marker techniques has had a large impact on quantitative genetics. Marker-based methods applied to segregating populations have provided us with a means to locate Quantitative Trait Loci (QTLs) to chromosomal regions and to estimate the effects of QTL allele substitution (Lander & Botstein, 1989). The ability to estimate gene effects and locations for quantitative traits can be very useful for the design and application of new, efficient, breeding strategies. A new selection strategy, marker-assisted selection (MAS), has been proposed by several authors as a way to increase gains from selection for quantitative traits (Tanksley, 1993; Lee, 1995; Knapp, 1994; Kearsey & Pooni, 1996). In backcross breeding programs, it has been shown that MAS can be effective in reducing linkage drag and optimising population sizes, by selecting against the donor genome except for the allele(s) to be introduced from the donor (e.g. Hospital *et al*. 1992). MAS can also improve selection for quantitative traits by selecting for the presence of specific marker alleles that are linked to favourable QTL alleles. This can be done for single marker loci or for an index representing several marker loci.

Breeding strategies for autogamous crops are often aimed at obtaining pure homozygous lines that show a superior phenotype. This can be done by generating genetic diversity, for instance a segregating $F_2$ population, followed by a selection of desirable individuals within the population, and repeated selfing and selection of individuals until sufficiently homozygous lines are obtained. A different strategy makes use of the genetic variation that is present in $F_2$ derived inbred lines, obtained without selection, commonly referred to as Recombinant Inbred Lines or RILs.

We consider a strategy based on intercrossing pairs of RILs. We assume that the aim of the selection is to obtain single genotypes containing as many accumulated advantageous alleles as possible. This goal is different from the aim of population improvement studied by most other authors. For example, Lande & Thompson (1990) and Gimelfarb & Lande (1994a,b) did not consider extreme genotypes within a MAS-derived segregating population, but focussed instead on improvement of the mean genotypic value of a population over several generations of selection.

In this paper, we analyse the possible benefits of MAS in autogamous crops, compared to conventional phenotypic selection. We investigate how the relative performance of MAS and conventional selection depend on the heritability of a trait, intensity of selection, genetic architecture (e.g. number and spacing of markers, number and effects of QTLs).

QTL mapping methods have improved continuously, since the earliest papers presenting and applying the approach (Soller & Brody, 1976; Lander & Botstein, 1989). In particular the use of co-factors in the analysis to account for multiple segregating QTLs can reduce the size of QTL support intervals on the genome considerably (Jansen & Stam, 1994). Nevertheless, uncertainty in estimates of QTL map locations and effects are unavoidable. We were interested to see how the performance of MAS is influenced by errors in the estimation of QTL locations and effects.

Our selection material consist of a set of RILs, obtained through single seed descent from a cross between two homozygous parents. Markers have been mapped and QTLs were supposedly mapped in the $F_2$ generation, allowing estimation of dominance effects. RILs are assumed to be completely homozygous. The problem we address is: 'which pair of RILs from this set is most promising in producing superior genotypes among their offspring?' We define superior genotypes as those genotypes that contain the favourable allele at (nearly) all detected QTLs, for the trait of interest. The performance of a pair of RILs is evaluated by considering the simulated $F_2$ offspring obtained by crossing these RILs (see below for details).

In an average sized population of RILs it is impracticable to cross and test all possible pairs of lines. Thus we wish to predict, before any RILs are crossed, which pairs are most likely to produce the most superior genotypes in the $F_2$, accumulating as many as possible advantageous alleles in a single genotype.

## Methods

In MAS, predictions for the performance of the offspring of line-pairs are used. These predictions are based on an index constructed from the genotypes of markers flanking putative QTLs in the pair of lines. In conventional selection, a line's phenotype determines if the line becomes part of a subset of selected lines. From this subset all possible pairs of lines are selected.

*Marker index construction*

The marker index value is calculated as an index for possible line combinations, based on the marker genotype of the potential $_1$ resulting from crossing two parental lines. Since the indices are connected to line pairs, a population of N lines results in ½[N*(N-1)] possible line combinations (not counting selfings and reciprocals). For each line-combination an index is calculated. This differs from the usual way combined indices are calculated (see for instance Lande & Thompson, 1990; Knapp, 1994), in the sense that this way of indexing takes genetic complementation into account. In our model, the smallest indexing unit is the marker *interval,* which consists of two markers that are located next to each other on the genetic map. If a QTL has been located within a marker interval, the interval is assigned an index number. A table is built

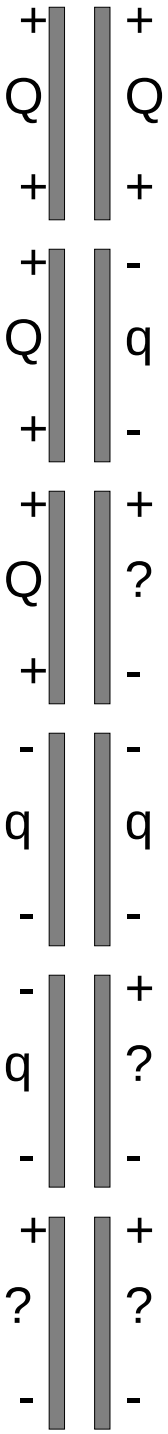



**Figure 1**: Marker interval combinations for a hypothetical $F_1$ between two RILs. The + and - indicate the alternative alleles at marker loci. The QTL alleles (Q/q) are inferred from the flanking markers. In case of uncertainty (?) the unfavourable QTL allele is assumed, and there is no contribution to the line-pair index.

connecting the index number with index values. This table contains the index values for the three possible situations (see Figure 1): (1) the favourable QTL allele is present in homozygous condition (QQ), (2) the QTL is heterozygous (Qq) or (3) the favourable QTL allele is absent (qq). The magnitude of the index values corresponds to the relative genetic effect of each allele combination; i.e. when the favourable allele is absent the index value is set to zero. It also depends on the dominant or additive character of the QTL. In all cases where the identity of the allele cannot be determined (see the '?' in Figure 1) the presence of the unfavourable allele is assumed.

$$CI = \sum_{Chrom} \sum_{Intervals} (QTL \cdot effect * Weight) \qquad \textbf{(1)}$$

Where CI is the combination index; Chrom means: all chromosomes; Intervals mean: all intervals on a chromosome; (QTL·effect * Weight) is the interval index and the Weights are defined as follows (see also Figure 1). In the case of additivity: QQ=2, Qq/Q?=1, qq/q?/??=0; in the case of dominance: QQ/Qq/Q?=2, qq/q?/??=0.

The overall index is calculated as the sum of all interval indices, according to (1). Because both parents are taken into account in the combination index, it can be seen as a predictor of the usefulness of a pair of lines.

*Phenotype*

The phenotypic value for a recombinant inbred line was calculated by adding an environmental error term, obtained from a normal distribution with mean $\mu=0$ and variance $\sigma^2=V_E$, to the line genotypic value. The line genotypic value was determined by the genotype at all QTLs, assuming additivity between QTLs. The magnitude of $V_E$ depends on the trait heritability. Genetic variance $V_G$ was calculated from the RIL genotypes, environmental variance $V_E$ was calculated according to (2), derived directly from the definition of heritability.

$$V_E = (\frac{1}{h^2} - 1)V_G \qquad \textbf{(2)}$$

Where $V_G$ is the genotypic variance, $V_E$ is the environmental (error) variance and $h^2$ is the broad sense heritability.

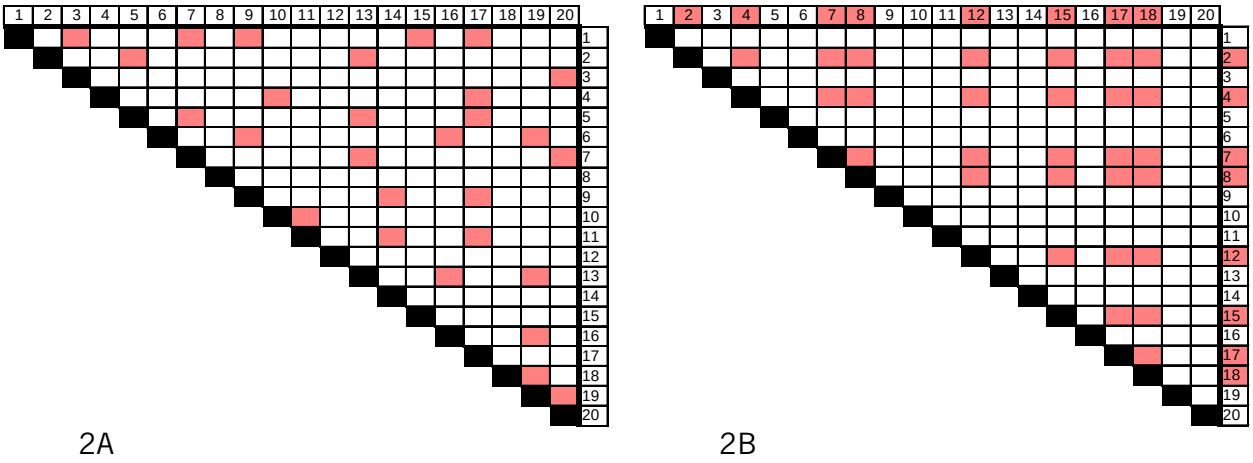



**Figure 2**: Comparison of a marker-assisted selection procedure (2A), with conventional phenotypic selection procedure (2B). With MAS, specific line combinations are selected, while with conventional selection lines are selected first and then combined with each other.

*Simulations*

Simulation consisted of the following steps.

1. Two complementary parents, defining the genetic architecture, were used to generate a set of 100 RILs. The genotype and phenotype of these RILs was calculated. Most simulation runs involved three replications, for each replication a different set of RILs was 'raised'.
2. For each set of RILs, marker indices were calculated for all RIL-pairs. Based on the combination indices a subset of all RIL pairs was selected for evaluation (MAS, Figure 2A). The size of this subset is called the 'selected fraction'.
3. Another subset of RIL pairs was selected based on the phenotype of the RILs (phenotypic selection, Figure 2B). Among the lines, RILs with the best phenotype were selected and from these RILs a set of line-pairs was derived. The number of lines that were selected was chosen in such a way that the total number of line-pairs in this second set was equal to the number in the set selected by MAS.
4. For each selected RIL-pair the $F_1$ generation was raised and subsequently selfed to obtain a segregating $F_2$ population of size 1000. For each generated $F_2$ population the average and standard deviation of the genotype was calculated. For the estimation of population extremes the $F_2$ progeny was divided into ten random groups of 100 progeny each. The most extreme genotype from each group was recorded and the

average over the ten group extremes was used as the value for the extreme genotype of the population. In this way we actually obtained an estimate of the extreme genotypic value in an $F_2$ population of size 100, which is an attainable population size in most practical situations.

5. The selection response was used to assess the success of each selected pair of RILs. The selection response was defined as the difference between the average population extreme genotypic value ($G_{ex}$) and the average genotype of all RILs ($G_{RIL}$), divided by $G_{RIL}$. This can be written as: $G_{RIL}=(\sum g_i)/N$ and: Selection Response (in %)$=100 * (G_{ex}-G_{RIL})/G_{RIL}$; where the RIL population consists of N RILs and the genotypic value of the $i^{th}$ RIL is denoted as $g_i$ . When the procedure was repeated over several sets of RIL the average selection response was used a a parameter for the success of the selection method.
6. The selection response obtained using MAS was compared to the selection response after phenotypic selection (obtained in a similar way).

We now describe the specific simulation conditions, used to investigate the influences of trait heritability, selection intensity, several aspects of genetic architecture and uncertainty in QTL locations on the performance of MAS, compared to phenotypic selection. Relevant simulation parameters are: The number of markers, the QTL positions and effects as well as the type of inheritance and linkage between QTLs, the trait heritability and the fraction of RIL-pairs that was selected. Except when stated otherwise, we assume the mapped positions of markers and QTLs are accurate, no interactions between QTLs occur and no interference is present during meiosis. The heritability is only used to estimate the magnitude of the environmental error. We assume the heritability was determined accurately in a trial of sufficient size.

*Trait heritability*

Four RIL populations were generated and used for simulation. Simulations were run for a genome containing five identical chromosomes. Nine markers were positioned at 10 centiMorgan (cM) intervals on each chromosome. Two QTLs per chromosome were located at positions 20 and 80, replacing the markers at these positions. The QTLs were linked in coupling phase. All QTLs had the same effect size, and there was no additive

interaction between QTLs. We only considered additive effects of allele substitution at each QTL. The fraction of pairs that was selected was 10%. We studied trait heritabilities ranging from $h^2=0.1$ to $h^2=0.9$.

*Selected fraction*

As stated earlier, it is ordinarily not feasible, to test all possible line combinations in a set of RILs. For this reason we assessed the amount of useful material that is lost by decreasing the number of selected RIL-pairs. Using the same configuration as for investigating heritability, we varied the fraction of selected RIL-pairs from 5% to 50% and recorded the selection response. Heritability was held constant at 0.1 and QTLs were linked in coupling phase. Only additive QTL allele effects were considered.

*Number of chromosomes, dominance, linkage phase*

We investigated the effects of different QTL configurations. For a genome consisting of 5, 10 or 20 chromosomes, we compared the selection response obtained with MAS to the selection response obtained when conventional selection was applied. nine markers were positioned at 10 cM intervals on each chromosome. Two QTLs per chromosome were located at positions 30 and 70 for the genomes consisting of 5 and 10 chromosomes, replacing the markers at these positions. One QTL per chromosome was located at position 35 for the genome consisting of 20 chromosomes. QTL alleles were linked in either coupling phase or repulsion phase. QTL allele effects were either additive or they showed complete dominance. The size of the QTL-effect was the same for all QTLs. Trait heritability was held constant at 0.1 and the selected fraction of RIL-pairs was 10%.

*Random QTL dispersion and geometric allele effects*

We also tested the genetic configuration used by Gimelfarb and Lande (1994a; their Figure 1). In this configuration 25 QTLs are dispersed randomly over 10 chromosomes of length 100. The effects of the QTL alleles constitute the 'geometric series of variance contributions' as described by Lande and Thompson (1990), which means that among the 25 QTLs there were only a few QTLs with a large effect and there were many QTLs with a small effect. It is believed that such a constitution better represents a true situation (Falconer & Mackay, 1996). We tested this setting with QTLs linked in

repulsion and coupling phase. The Gimelfarb and Lande genome has marker loci at every 10 cM, leading to a total of 110 marker loci. We also tested the effect of marker loci present every 20 cM, resulting in a map with 60 markers in total. The selected fraction of RIL-pairs was 10%. Trait heritability was held at 0.1 or 0.3.

*Errors in QTL mapping*

To study the effect of uncertainty in QTL number and positions we have run simulations for the following situations:

– **QTLs mapped to incorrect marker intervals**

  It is assumed that the mapped positions of some QTLs do not correspond with the true positions on the genome. Instead these QTLs are mapped to intervals adjacent to the true intervals, leading to selection of incorrect marker intervals by the MAS procedure. We tested a configuration with 10 chromosomes, carrying 20 QTLs with equal effects linked in coupling phase. Nine markers per chromosome were present at 10 cM intervals. Two QTLs per chromosome were present at locations 30cM and 70cM, replacing the markers at these positions. All QTL effects were additive. Trait heritability was held at 0.1 and the selected fraction of RIL-pairs was 10%. The proportion of QTLs that were not assigned to their true marker interval but to a neighbouring interval ranged from 5% to 100%.

– **Undetected QTLs (Type II errors)**

  Here we assumed that the QTL mapping procedure failed to locate one or more QTLs, causing reduced selection opportunities for MAS. The same configuration was used as described in the previous section, dealing with QTLs mapped to incorrect intervals, but a randomly chosen subset of the QTLs that were present in the simulated cross were not used by MAS. We ran simulations under the assumption that 0%, 25%, 50% or 75% of the QTLs were not included in the marker-assisted selection.

– **False positive detected QTLs (Type I errors)**

  When QTL detection is conducted, there is always the risk that the QTL mapping procedure falsely indicates the presence of one or more QTLs at positions where none

in fact exist. These 'false QTLs' were included in the index used by MAS, introducing errors in the overall combination-index. Again the same configuration as described in the previous section was used. Twenty true QTLs were present, but the number of QTLs used in marker-assisted selection ranged from 20 to 40. The 'false QTLs' were added to the genome randomly, but as a constraint, no more than four QTLs could be present per chromosome and only one QTL was allowed per marker interval.

*Software*

For the execution of these experiments a computer program, allowing simulation of crossing and selection, was created. A locus (marker or QTL) is the smallest unit that is present in the computer model. Loci are linked together in linkage groups or chromosomes and Mendelian rules apply to the simulation of recombination during meiosis. QTLs and allelic effects remain visible, but are not used for selection. Selection is based only on marker loci and intervals of marker loci. Within the model, indices are calculated for pairs of lines. Based on these index values, pairs of lines are either selected or disregarded of the selected fraction. In conventional selection, phenotypic values are used as the criterion to select RIL-pairs. The computer model was written in Borland Delphi and executed on a Pentium PC.

## Results

*Trait heritability*

The results of this experiment are summarised in Figure 3. With additive QTL effects, MAS resulted in a higher selection response at heritabilities 0.1 and 0.3, while for heritability 0.5 the advantage of MAS over phenotypic selection becomes negligible. At trait heritability approaching 1.0 we can see that the phenotypic selection response becomes larger than the selection response after MAS. This observation is probably due to the conservative way index selection is practised. If only one of two markers flanking a QTL is present, no index value is awarded, because it is uncertain which QTL allele is present. In approximately half of the cases this will be the advantageous allele, but in the other half it will be the other, undesirable allele. In this way some of the advantageous

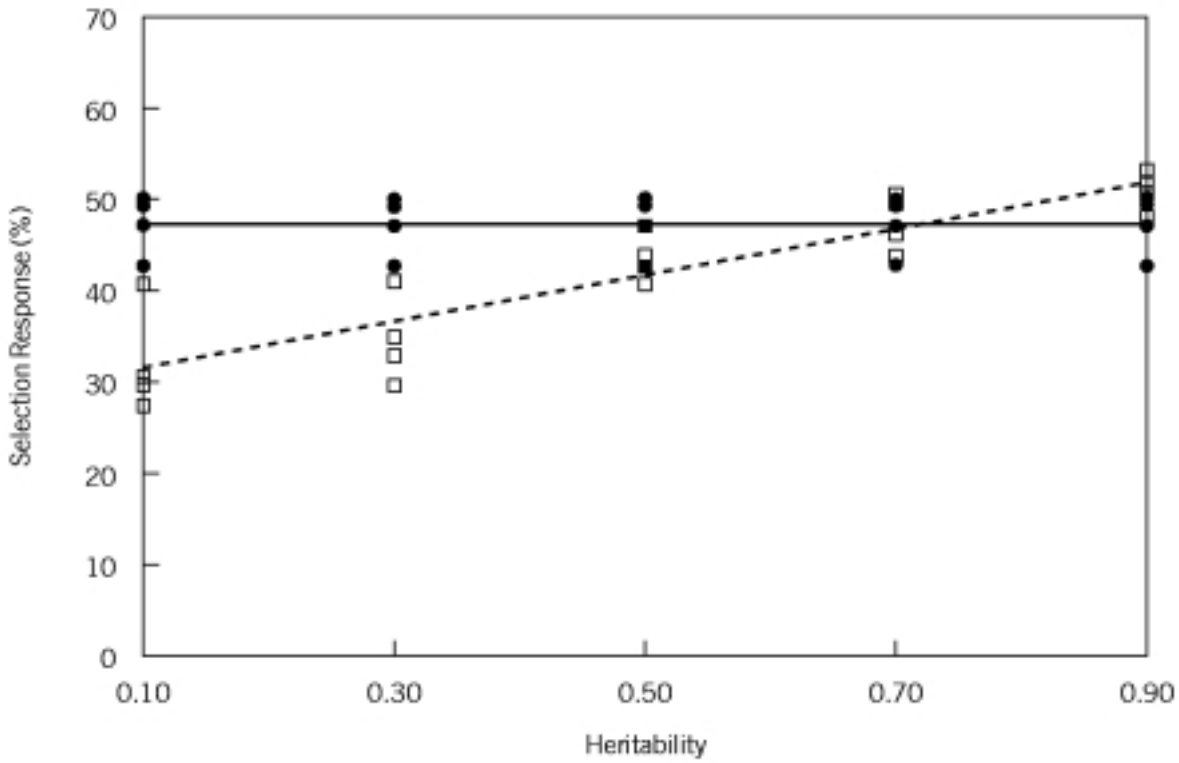



**Figure 3**: Comparison of relative selection response for different trait heritabilities. [●+ continuous line]: MAS; [□+ dashed line]: Phenotypic Selection. Lines show a linear regression through replication means. The selected fraction was kept at 10%; Simulated genome consisted of five chromosomes with each nine markers and two QTLs. Markers at 10 cM intervals and QTLs at positions 20 and 80.QTLs had additive effects and were linked in coupling phase.

alleles are missed by MAS, reducing it's power. To keep the number of tested setting practicable, we decided to set the trait heritability at 0.1 or 0.3 in the other tests, because this is where we expect the contrasts between MAS and phenotypic selection to be the largest.

*Selected fraction*

We show the selection response for a range of selected fractions of RIL-pairs in Figure 4. The superiority of MAS decreased as the fraction of selected RIL-pairs increased. The reduced selection response of MAS and conventional selection at smaller selected fractions of RIL-pairs reflects the cost of missing some of the most promising RIL-pairs when testing too few of them. The reduction in selection response for phenotypic selection was expected, because we select for extremes and a smaller subset of the population is less likely to contain the best combination of lines. When a desirable line remains unselected in phenotypic selection this will affect several RIL-pairs that would

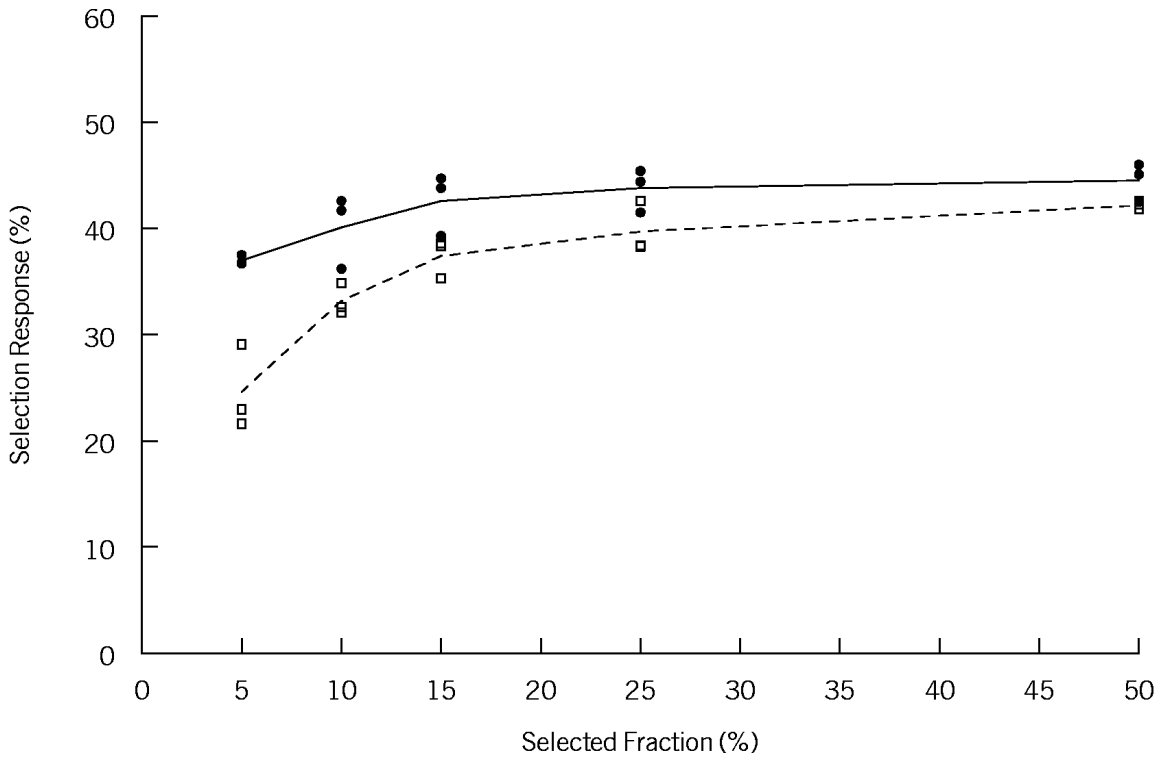



**Figure 4**: Comparison of relative selection response as a function of the selected fraction of RIL pairs. [●: Marker-Assisted Selection, solid line connects averages over replications]; [□: Phenotypic Selection, dashed line connects averages over replication]. Trait heritability was kept at 0.1, QTL alleles had additive effects. The genetic architecture was the same as described in Figure 3.

have included this line, thus lowering selection results of conventional selection as a whole. This effect is not seen for MAS because in MAS, for each RIL-pair selection is performed independently. However, marker-assisted selection still showed a drop in selection response when fewer RIL combinations were selected. This effect would not be expected if the combination index would be able to predict -without error- the usefulness for breeding of a cross. However, the conservative way in which the index value is constructed ensures that the index value of a RIL-pair never overestimates, but may underestimate, the value of a RIL-pair. This will happen when crossovers occur inside marker intervals used for indexation. This underestimation may result in missing some of the most promising RIL-pairs when the selected fraction is small. To limit the number of possible parameter settings we arbitrarily chose to select 10% of all RIL pairs in the following simulation experiments, unless indicated otherwise. For a population consisting of 100 lines this meant selection of 495 line-pairs out of a possible 4950.

**Table 1**: Relative selection responses[1] in conventional phenotypic selection (CS) and marker-assisted selection (MAS) for different genetic configurations, types of inheritance and linkage conditions. The presented data are averages over three different RIL-sets. The genome consisted of chromosomes of length 100 cM with evenly spaced markers at 10 cM intervals. The configuration with 20 chromosomes contained only one QTL per chromosome, at 45 cM. The other configurations contained 2 QTLs per chromosome located at 35 and 75 cM, linked in coupling phase or repulsion phase. QTL effects were of equal size for all QTLs. Trait heritability was fixed at 0.1 and the selected fraction of RIL pairs was 10%.

| | | **5 chrom, 10 QTLs** | | **10 chrom, 20 QTLs** | | **20 chrom,** |
|---|---|---|---|---|---|---|
| | | Coupling | Repulsion | Coupling | Repulsion | **20 QTLs** |
| Additive | CS | 32% | 34% | 27% | 23% | 34% |
| | MAS | 52% | 47% | 42% | 32% | 44% |
| Dominant | CS | 59% | 56% | 51% | 48% | 33% |
| | MAS | 84% | 72% | 68% | 58% | 45% |

[1] The selection response was calculated as: $100 * (G_{ex}-G_{RIL})/G_{RIL}$, Where $G_{ex}$ is the average of the realised extreme genotypes of the $F_2$ progenies resulting from the selected RIL pairs, and $G_{RIL}$ is the average RIL genotypic value.

*Number of chromosomes, dominance, linkage phase*

The general results of these experiments are summarized in Table 1. Selection response is presented for MAS and phenotypic selection. In all the tested configurations marker-assisted selection gave a higher selection response, compared to phenotypic selection. The effect is larger when QTL alleles are linked in coupling phase. The difference is also larger when QTL alleles exhibit dominance. This can be explained by the way the selection index is constructed. Conventional selection uses the phenotype of the RILs, while MAS uses the genotype of the $F_1$, obtained from a cross between two RILs, for selection. In this way, heterozygous $F_1$ progeny that are advantageous because of accumulated dominant genes can be selected by MAS. After selfing they can give rise to a segregating population containing more extreme genotypes. If the final objective is to obtain inbred lines for hybrid production these numbers give an indication of the progress that can be achieved. For purely autogamous crops the dominance effect will be lost in later generations of inbreeding and only the additive QTL effects remain.

**Table 2**: Relative selection responses[1] in conventional phenotypic selection (CS) and marker assisted selection (MAS) for different heritabilities and marker spacings in the case of random dispersed QTLs and geometric QTL effects. Data presented are averaged over three different RIL sets. The genome consisted of 10 chromsomes of length 100 cM with evenly spaced markers at 10cM or 20 cM intervals. The distribution of QTLs and their effects were as specified by Gimelfarb & Lande (1994a).QTL-effects were assumed additive. Linkage between QTLs on the same chromosome was either in coupling phase or in repulsion phase. Trait heritability was kept at 0.1 or 0.3. The selected fraction of RIL pairs was 10%.

| | | **Coupling** | | **Repulsion** | |
|---|---|---|---|---|---|
| | | 10cM | 20cM | 10cM | 20cM |
| $h^2$=0.10 | CS | 27% | 27% | 20% | 20% |
| | MAS | 51% | 49% | 27% | 23% |
| $h^2$=0.30 | CS | 33% | 33% | 22% | 22% |
| | MAS | 51% | 49% | 27% | 23% |

[1] The selection response was calculated as described in Table 1.

*Random QTL dispersion and geometric allele effects*

The selection response for MAS and phenotypic selection for the data set derived from the Gimelfarb and Lande (1994a) map are summarised in Table 2. Again we see that MAS results in a higher selection response compared to phenotypic selection. When the number of marker loci is reduced from 110 to 60 (the interval size is increased from 10 cM to 20 cM), the frequency of having more than one QTL within a marker interval increases. This results in a reduction of the selection response for MAS, especially when QTLs are linked in repulsion phase, because the overall effect of the marker interval will become small when neighbouring QTLs, within a marker interval, partly counterbalance each others effect.

*Errors in QTL mapping*

– **QTLs mapped to incorrect marker intervals**

The performance of MAS is affected when QTLs are not mapped at their true position. The magnitude of this effect can be seen in Figure 5. A reduction in selection response was observed, as the number of incorrectly located QTLs increased, but the effect was small. We believe this is because using a neighbouring marker interval for calculation of the index will in most cases still result in the same index. Only when recombination

has occurred within either or both of the correct and incorrect intervals will the resulting index be affected, and thus the performance of a RIL pair inaccurately predicted.

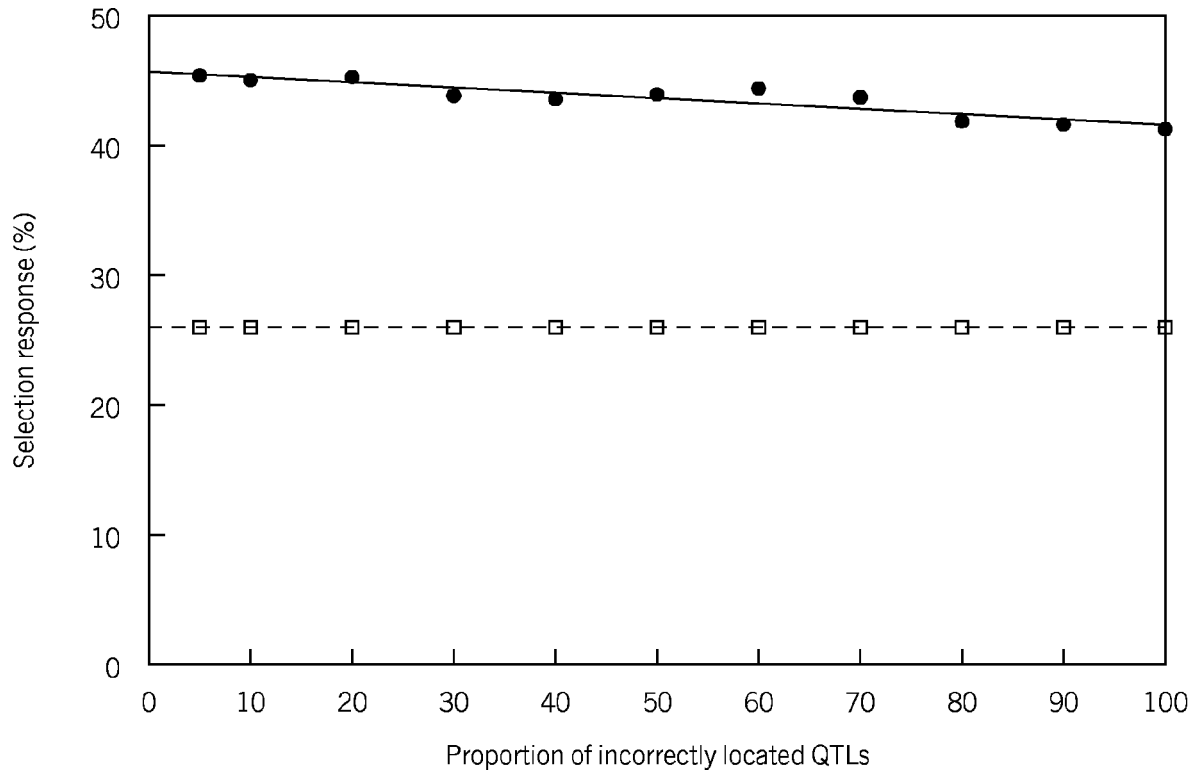



**Figure 5**: MAS relative selection response as a function of the proportion of incorrectly located QTLs. [●: Marker Assisted Selection]; [□: Phenotypic Selection]. Trait heritability was kept at 0.1. The selected fraction of RILs was 10%. The genetic architecture was the same as described in Table 1 for loci linked in coupling phase and QTL alleles with additive effects.

– **Undetected QTLs (Type II errors)**

QTLs that have an influence on the phenotype are not always detected at the mapping stage. As a result, these QTLs can not be selected by the MAS procedure. The size of the reduction in selection response caused by undetected QTLs is shown in Figure 6. A reduction in selection response was observed as the proportion of undetected QTLs increased. However, even when only 25% of the QTLs are mapped and indexed, the selection response obtained after applying marker assisted selection is still 4% larger than the response after applying phenotypic selection. This indicates that (for low

heritability traits) it is worthwhile to pursue marker-assisted selection, even if the phenotypic data did not allow detection of all QTLs.

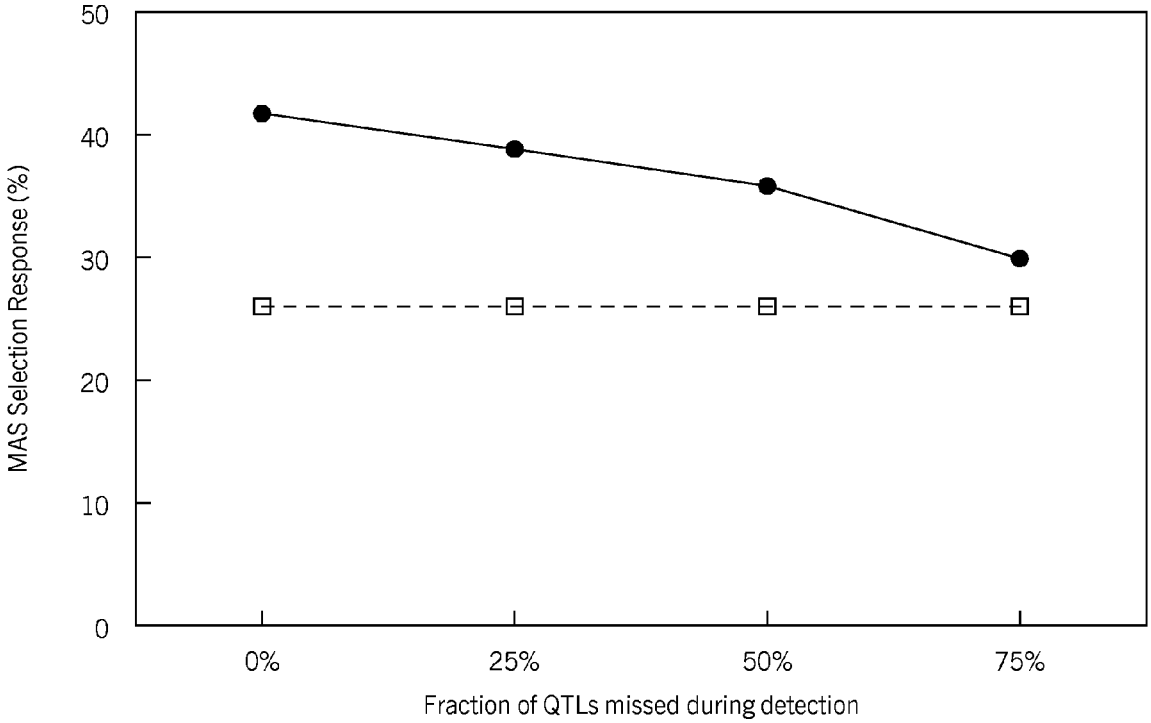



**Figure 6**: MAS relative selection response when indices are incomplete because of undetected QTLs. [●: Marker Assisted Selection]; [□: Phenotypic Selection]. Trait heritability was 0.1. The selected fraction was 10%. The genetic architecture was the same as described in Table 1 for loci linked in coupling phase and QTL alleles with additive effects.

– **False positive QTL detection (Type I errors)**

The introduction of false QTLs - QTLs that are not actually present, but were falsely identified by the QTL mapping procedure- showed no effect on the MAS selection results (data not shown). Even when the number of false QTLs equalled the number of true QTLs no significant decrease in selection response was found. Apparently the MAS procedure does not suffer much from extra information. This may be due to the configuration we tested. All QTLs were linked in coupling phase, so adding QTLs to the map will inflate the index value, but the order of index values and the line pairs that will be selected will not change dramatically.

## Discussion

We have assumed that a set of RILs obtained from a given cross, well characterised in terms of marker genotypes and QTL positions, is available as a starting point for further crossing and selection.

We have not focussed on population improvement by MAS but rather on 'breeding behaviour' of pairs of RILs. The results indicate that marker data can be a valuable extra source of information on which to base selection, especially when heritability is low. Marker information appears to add little to phenotypic information at high heritability, but at low heritability it does so. This is in agreement with results on recurrent MAS for population improvement (Lande & Thompson, 1990; Gimelfarb & Lande, 1994a,b; Gallais & Charcosset, 1994).

In all simulations we have assumed that all QTLs affect a single trait. This is, of course, a simplification but not a limitation; one can easily imagine the case where the QTLs of the model are divided into subsets, each set affecting a different trait. The 'final trait' could then be an index value, composed of a linear combination of traits. This will not change our general results, as long as the traits involved are comparable in their importance to the breeder. When many QTLs are to be accumulated the chance of getting them all with just one pair of lines is small. In this case, one may think of an extension of the procedure to three way crosses or four way crosses.

Trait heritability is the most important factor influencing the effectiveness of MAS. MAS seems to be most promising for traits with low heritability. But trait heritability is also of major importance for accuracy in the mapping of QTLs. Low heritability reduces the power of detecting QTLs, which is based on correlation between phenotype and marker genotype. This could mean that for well-mapped QTLs MAS may add little to phenotypic selection, while for traits with a very low heritability the underlying QTLs cannot be identified. It is the area in between these two extreme cases that looks most promising for application of MAS. If QTLs can be mapped for a trait having a low heritability the accuracy of the QTL position may not be very high, which is reflected in a large QTL

support interval on the genetic map (Lee, 1995). Our simulations have shown that this does affect the effectiveness of MAS, but only marginally in most cases.

To practical breeders these result may be an incentive to continue to use marker data as a source of information on which to base selection. In most cases MAS will give better selection results than phenotypic selection, for a low heritability trait. The breeder can decide if the increased selection results are worth the extra cost involved in obtaining the marker data. Index based selection opens new ways to quantify performance with regard to several traits into one index value, and use markers to select for those plants that give an optimisation of this index in the current or a future generation. This may facilitate breeding for several traits simultaneously. In future more and more marker and QTL information will be collected; also existing breeding populations will be screened for markers and QTLs. An efficient way to use this information and to predict useful crosses would require prediction and selection procedures similar to the procedure described in this paper.

**Acknowledgements**

We like to thank Dr. Scott Chasalow for reviewing this paper and for helpful comments, discussions and corrections.

# 3 Comparison between Marker-Assisted Selection and Phenotypical Selection in a set of *Arabidopsis thaliana* recombinant inbred lines

Ralph van Berloo · Piet Stam

*Published in Theoretical and Applied Genetics 98:113-118 (1999)*

## Introduction

Marker and QTL information obtained from a segregating population can be used for the design of efficient breeding strategies. In recent years major advances in marker availability and statistical methods for assessing marker-trait correlations have been achieved (e.g. Lander & Botstein 1989; Haley & Knott, 1992; Van Ooijen, 1992; Jansen & Stam 1994; Falconer & Mackay 1996). Marker-assisted selection (MAS) has been advocated as a useful tool for rapid genetic advance in the case of quantitative traits (Lande & Thompson, 1990; Knapp, 1994; Knapp, 1998). In our previous paper (Van Berloo & Stam, 1998) we describe a procedure for the application of MAS in an autogamous population of Recombinant Inbred Lines (RILs). In this paper we report experiments using *Arabidopsis* as a model species. *Arabidopsis* is well suited for model selection experiments because of it's small size and short generation cycle (Meyerowitz & Pruitt, 1985). Over the years a vast body of genetic data on *Arabidopsis* has become available. Kuittinen *et al.* (1997) described a QTL mapping experiment for flowering time in *Arabidopsis*. Five to seven QTLs affecting flowering time were found in a $BC_1$ population, derived from the Finnish Naantali genotype and the German strain Li-5. In a different population, consisting of 165 Ler x Cvi RILs, Alonso-Blanco *et al.* (1997) found four QTLs affecting flowering time. Jansen *et al.* (1995) used the *Arabidopsis* RIL set, obtained from a cross between the Columbia (Col) and Landsberg *erecta* (Ler) ecotypes (Lister & Dean, 1993), in a QTL mapping experiment involving various environments. Day length was varied and in some cases a vernalisation treatment was applied. In this experiment 12 QTLs for flowering time were detected. Eight QTLs had an effect in all environments and four QTLs showed an effect in only some of the environments.

In this paper we describe an experiment using the Col x Ler *Arabidopsis* RILs of Lister and Dean (1993). The objective was to compare a MAS breeding strategy, using molecular marker and QTL information, with conventional breeding methods, based on phenotype only. Focus lies on the selection of suitable parents for crossing. The $F_2$ offspring derived from these parents is the target generation, in which the quality of selection is evaluated. In both MAS and phenotypical selection procedures the target was the production of genotypes that contain as many as possible advantageous alleles for the QTLs that affect the trait of interest (these genotypes will be referred to as 'superior' or 'extreme' genotypes). In this case, the trait of interest is flowering time.

## Material and Methods

*Plant material*

The Col x Ler *Arabidopsis* RIL set, consisting of 99 lines, was obtained from the *Arabidopsis* stock centre in Nottingham, UK. The set of RILs was developed by Lister and Dean (1993) and was derived, through single seed descent, from an $F_2$ population that resulted from a cross between the Landsberg *erecta* and Columbia ecotypes. In our experiment we identified the lines according to the *Arabidopsis* stock centre line numbers, using RIL numbers from 1900 to 1998.

*Trait*

Our trait of interest is flowering time. Flowering time is generally regarded as a quantitative trait which may show an influence on other traits (Kuittinen *et al*, 1997). Flowering time is measured as the number of days from planting of the germinating seeds till the first petal becomes visible. Scoring of flowering time is approximated by using one-day classes.

*Marker and QTL Data*

RFLP marker data for all 99 RILs were obtained from Jansen *et al.* (1995). These data were used to construct a genetic map using the Joinmap package (Stam & Van Ooijen, 1995). This map corresponded with the integrated genetic map, which was available at

NASC (NASC, 1998). From 12 flowering time QTL estimates, obtained from Jansen *et al.* (1995, and pers. comm.), eight that had a significant effect under long day conditions without seed vernalisation, were selected for marker and QTL analysis. In our experiment we used the same set of RILs as Jansen *et al.* (1995).

**Table 1:** Presence of QTL alleles for earliness in RIL set, assessed through graphical genotype analysis.

| **Number of 'earliness' QTL alleles** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| **Frequency within set of RILs** | 0 | 1 | 9 | 26 | 27 | 21 | 12 | 2 | 0 |

*Graphical Genotypes*

The RILs were subjected to analysis using graphical genotypes (Young & Tanksley, 1989). Marker data for all RILs were displayed graphically, using a different colour for each parent of the RIL population (Col/ Ler). For analysis the computer program GGT (Van Berloo, 1999a) was used. When markers indicated that a chromosomal region at a QTL was of the same origin as the parent that contributed the favourable allele it was assumed that the RIL inherited this allele. In this way the number of favourable QTL alleles could be assessed for all RILs. The distribution of the number of favourable QTL alleles for early flowering over the RILs is listed in Table 1. Columbia contained three favourable QTLs for earliness and Landberg *erecta* five. None of the RILs contained all favourable alleles for early flowering. Furthermore, all of the RILs contained at least one favourable allele for early flowering.

*Selection*

*Arabidopsis* is a self fertilising species (Abbot & Gomez, 1989). Therefore, the selection result should be a single genotype or line that contains as many favourable QTL alleles as possible. The procedure used for obtaining this 'extreme' genotype was the same as we applied in earlier simulation studies (Van Berloo & Stam, 1998). Basically, the method identifies those pairs of RILs which, upon crossing, give rise to a high number of superior QTL-genotypes among their $F_2$ offspring. This is done by preselecting RIL-pairs on the basis of their marker-genotype and subsequently simulating their $F_2$ offspring. Selection for flowering time was aimed in two directions, for late flowering and for early flowering.

Two criteria were used to select RIL combinations for crossing: (1) the predicted breeding potential of a line-pair based on marker and QTL data, and (2) the observed line phenotype.

(1) Predicted breeding potential.

The available marker and QTL data were used by MS, the computer program for MAS, that identifies line pairs that have a high probability of accumulating favourable QTL alleles in $F_2$-offspring genotypes (Van Berloo & Stam, 1998). The program was run with marker and QTL data from the 99 RILs. This resulted in a list of preferable crosses.

(2) Observed phenotype.

RILs were ordered, according to their phenotype (calculated as an average over 24 plants). Next, a subset of RILs comprising the extreme 10% were selected. Within this subset line pairs were selected at random for crossing.

Out of a possible 4851 (½*99*98) pairs, 25 were selected using MAS and 25 pairs were selected based on their phenotype. We harvested seeds from 14 'MAS crosses' and 17 'phenotypic crosses'. A subset of 11 $F_1$'s from MAS crosses and 12 $F_1$'s from phenotypic crosses were selfed to obtain $F_2$ seeds. $F_2$ plants from four MAS crosses and four phenotype based crosses were evaluated in a greenhouse trial.

*Experimental setup*

All plants were grown in the same greenhouse under long day conditions (18 hours light, 6 hours dark). Seeds were not vernalised before sowing, but the germinating seeds were allowed 48 hours at 4°C to break dormancy. Per line 24 plants were grown in two replications. Lines were randomised within a replication. Flowering time of the RILs was observed. Selected line combinations (see selection paragraph for criteria) were crossed, and their $F_1$ seeds were harvested. Next, $_1$ seeds from 23 selected crosses were grown without replications. On average 12 plants per cross were grown. Plants were allowed to self-fertilise, and $F_2$ seeds were harvested.

**Table 2:** RIL pairs that were selected for crossing by the different selection methods and the prediction for ability to produce extreme $F_2$ offspring

| Selection type[1] | RIL pair | Allele types[2] | Prediction[3] |
|---|---|---|---|
| ME | 1991x1906 | EEEEEELE x LEEEEELE | 100 |
| | 1942x1991 | LLEEEELE x EEEEEELE | 98 |
| PE | 1926x1906 | EEEEEELL x LEEEEELE | 94 |
| | 1956x1910 | ELEEEELL x LEEELEEL | 70 |
| ML | 1962x1984 | LLLLELEE x LLLLELEL | 73 |
| | 1978x1984 | LLLLLELE x LLLLELEL | 75 |
| PL | 1916x1940 | LLELEELL x LLLLELLE | 44 |
| | 1916x1980 | LLELEELL x LLLEEEEE | 29 |

1 ME=MAS, Early flowering; PE=Phenotypic selection, Early flowering; ML=MAS, Late flowering; PL=Phenotypic selection, Late flowering.

2 Allele types indicate the QTL alleles for the 8 QTLs, listed as Parent-1 x Parent-2 = $Q_1Q_2Q_3Q_4Q_5Q_6Q_7Q_8$ x $Q_1Q_2Q_3Q_4Q_5Q_6Q_7Q_8$; E = Early allele, L =Late allele.

3 Prediction based on the average of 10 replicates of extremes found by computer simulations of 100 $F_2$ progeny; predictions, indicating RIL-pair potential for obtaining extremes, are ranging between 0-100, 100 being the highest possible value, according to the direction of selection.

For each of the four categories two crosses were selected (see Table 2). Each selected cross was represented by 200 $F_2$ plants that were grown in a greenhouse experiment. As a control 800 plants from the RIL set were grown. Four RILs were selected to represent the RIL set, one early flowering and one late flowering RIL, and two RILs of moderate flowering time. The experimental setup was a block design with 17 blocks. Plant rows were randomised within blocks and blocks were randomised over the greenhouse. For each of the 2400 plants the flowering time and the number of leaves at the time of flowering were recorded.

*Data analysis*

The observations on the 2400 plants were used to obtain estimates for population average and variance. This was done using the statistical computer package ASREML, provided by Gilmour *et al.* (1995). ASREML allows estimation of population variances

and their standard errors. A square-root transformation was applied to the discrete data in order to obtain a normal distribution of residuals. The model fitted to the data was:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \underline{h}_{jk} + \underline{e}_{ijk}$$

with $\alpha_i$ : contribution of blocks; $\beta_j$ : contribution of population mean; $\underline{h}_{jk}$ : contribution of specific plant genotype and $\underline{e}_{ijk}$ : remaining error term

Since the controls (RILs) are genetically homogeneous (within lines), the variance within these controls (averaged over RILs) was used to assess the environmental variance. The genetical component of the $F_2$ population variances was obtained by subtracting the environmental variance from the experimental variance. Heritability was estimated as the ratio of the genetic and phenotypic variance. We were interested in plants within the populations that possess 'extreme' or superior genotypes, therefore we considered the 95% percentiles of the distribution of the $F_2$ populations. From statistical theory (e.g. Levert, 1959) it is known that the 95% confidence interval for the 95% percentile of a normal distribution ($x_p$) can be found by:

$$\hat{\mu} + 1.45\hat{\sigma} < X_p < \hat{\mu} 1.88\hat{\sigma}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the estimated mean and standard deviation, respectively.

Confidence intervals for the 95% percentile of the $F_2$ phenotypic distribution were estimated for each cross.

## Results

Figure 1 shows a scatter plot of the RIL flowering time (phenotypic value) vs the number of favourable QTL alleles present in the RILs. RILs that are part of pairs that were selected by MAS or phenotypical selection are highlighted. Phenotypical selection is less successful than MAS in selecting the RILs with the highest number of favourable QTL alleles. RILs selected by MAS show a less extreme trait value. This was expected, because these RILs are selected for their ability to complement each other genetically, not because they show a high trait value themselves.

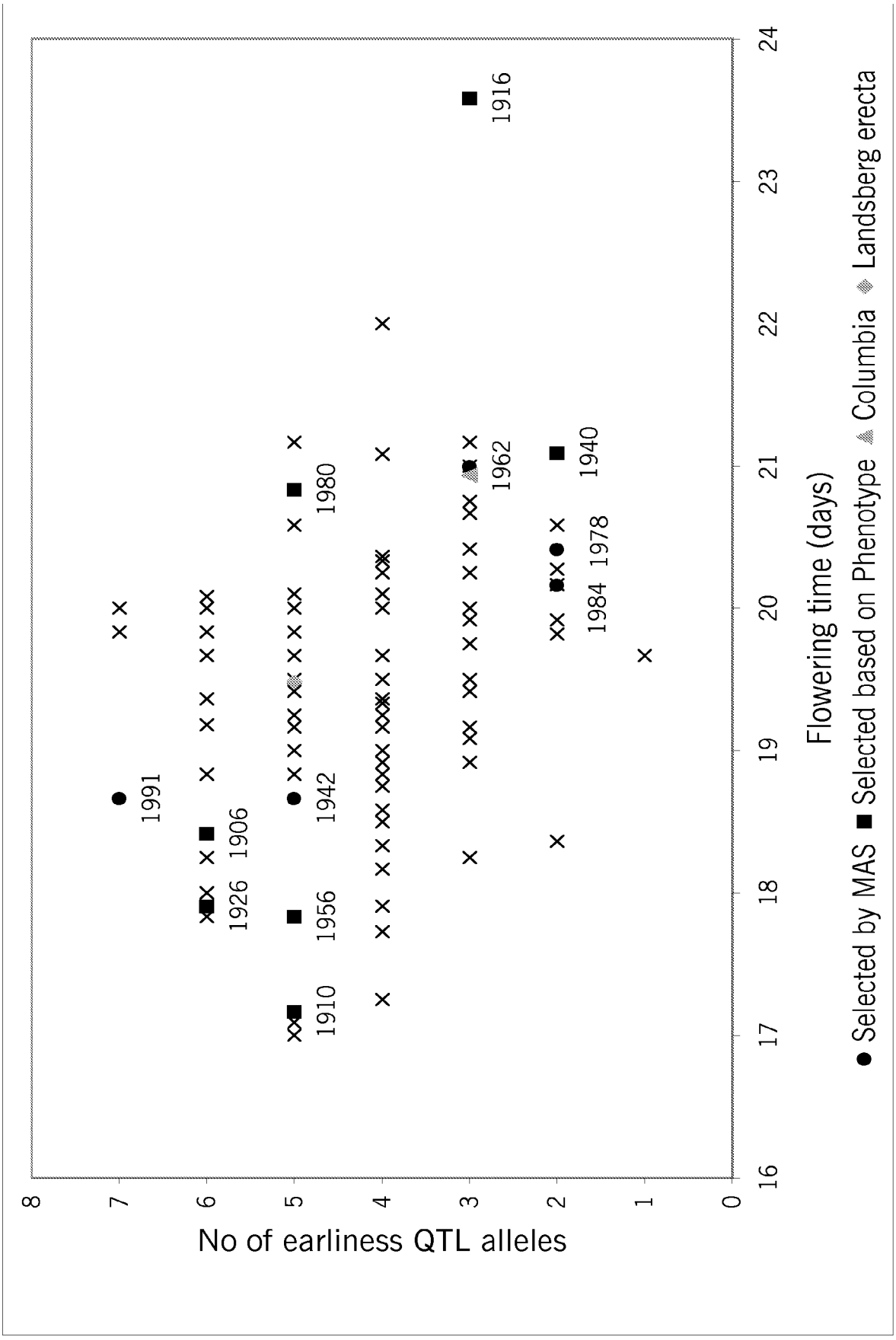



**Figure 1**: Scatter plot of the number of QTL alleles, favourable for early flowering vs the realised flowering time of the RILs. RILs that were selected by MAS or phenotypical selection are indicated separately.

For reasons of simplicity, no effect sizes of the QTL alleles have been taken into account in Figure 1. Therefore caution should be taken in making comparisons between data points. A large difference in the number of QTL alleles does not necessarily result in an equally large difference in genetic potential.

**Table 3:** flowering time means, standard deviations and heritabilities for $F_2$ populations obtained after marker-assisted selection or phenotypical selection for either late or early flowering.

| | | | √(Flowering time) | | | |
|---|---|---|---|---|---|---|
| **Population** | **Type**[1] | **S.Q.**[2] | $\hat{\mu}$ | $\hat{\sigma}_G$ | $\hat{\sigma}_E$ | $h^2$ |
| 1991x1906 | ME | 1 | 5.70 a | 0.042 | 0.15 | 0.07 |
| 1942x1991 | ME | 2 | 5.75 a | 0.085 | 0.15 | 0.24 |
| 1926x1906 | PE | 2 | 5.68 a | 0.108 | 0.15 | 0.34 |
| 1956x1910 | PE | 5 | 5.53 a | 0.060 | 0.15 | 0.14 |
| 1962x1984 | ML | 1 | 6.12 b | 0.156 | 0.15 | 0.52 |
| 1978x1984 | ML | 4 | 6.04 b | 0.159 | 0.15 | 0.53 |
| 1916x1940 | PL | 3 | 6.08 b | 0.115 | 0.15 | 0.37 |
| 1916x1980 | PL | 4 | 6.04 b | 0.143 | 0.15 | 0.48 |

[1] See legend of Table 2.

[2] S.Q.=The number of segregating QTLs, derived through graphical genotype analysis, see Table 2.

$\hat{\mu}$ :mean of $F_2$ population, a and b indicate groups that show a significant difference at $\alpha=0.05$; $\hat{\sigma}_G$: Estimated genetic standard deviation; $\hat{\sigma}_E$: Estimated environmental standard deviation; $h^2$: Observed heritability of the transformed trait ($F_2$)

The RILs showed a continuous, unimodal phenotypic flowering time distribution. Extreme flowering times were 13 and 27 days; RIL means ranged between 17 and 24 days. Table 2 shows the RIL-pairs that were selected, and the associated prediction value that resulted from the model prediction. The $F_1$ plants showed a clear distinction between the group selected for late flowering and the group selected for early flowering, as was expected (data not shown). In the $F_2$ populations we observed plant flowering times ranging from 26 to 52 days. The estimates of population means, standard deviations and heritabilities are shown in Table 3. The average heritability for flowering time over all populations was 0.34. Two distinct groups of crosses emerged: an early flowering group and a late flowering group. When the results within these groups are

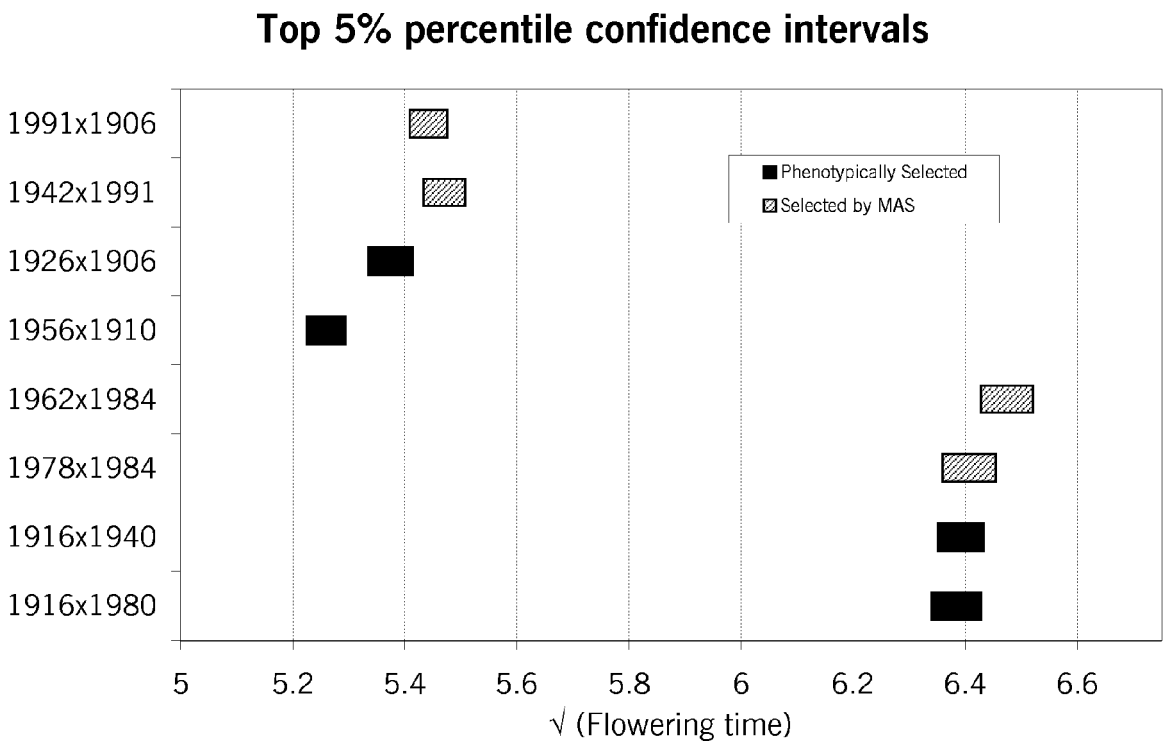



**Figure 2**: Confidence intervals (95%) for the right ('late' selections) and left ('early' selections) 95% percentile of the $F_2$ flowering time distributions.

compared, the differences are less clear, and no significant differences between phenotypically selected crosses and MAS crosses can be observed.

The 95% percentile was used as a parameter for comparison between the tails of normal distributions. Confidence intervals ($\alpha=0.05$) were calculated for the 95% percentile of each population. The right percentile was used for the 'late' crosses and the left percentile for the 'early' crosses. Confidence intervals are displayed in Figure 2. This figure again shows that selection has led to two distinct groups: a late and an early flowering group. However, within such a group no large differences between selection methods can be observed. Within the 'late' group, the MAS confidence intervals lie more in the direction the selection was aiming for than the other confidence intervals, while in the 'early' group the reverse situation is true. Most confidence intervals of the different selection methods overlap.

## Discussion

This experiment was aimed at a comparison of two different selection methods. The source of information, on which selection is based, is different for the two methods. Marker-assisted selection used only marker data and information on QTL locations, obtained from previous experiments, to predict useful crosses. Phenotypical selection used plant phenotypic data that were collected in an additional experiment. The final results do not favour one selection method over the other.

Although we expected the marker-assisted selection procedure to be more efficient in obtaining extreme phenotypes in an $F_2$ progeny resulting after crossing selected parents, the results from this experiment did not confirm this expectation. This may be due to the nature of the trait we investigated. In our experiment, we found an average heritability for $F_2$'s of 0.34 for flowering time. Assuming absence of dominance, conversion into a heritability for RILs would yield about 0.7. After all, this heritability may well be too high to take full advantage of marker-assisted selection. Benefits of the MAS procedure are to be expected only in the case when the trait heritability (calculated for RILs) lies approximately within the range of 0.1 - 0.3 (Van Berloo & Stam, 1998). When the heritability is too high, the cost involved in genotyping many plants may not outweigh the expected benefits of more direct gene selection. On the other hand, when the heritability drops below 0.1, the QTLs cannot be identified with the accuracy required to rely on flanking markers for selection.

One of the main theoretical reasons why MAS is expected to outperform phenotypic selection is that RIL-pairs selected by MAS will generate, on average, more genetic variance in the offspring because such RIL pairs will tend to be complementary with respect to QTL alleles. In our experiment, however, this advantage of MAS over phenotypic selection has, in hindsight, not been realised. From Table 3 it can be seen that there is no clear relationship between the estimated genetic variance and the number of segregating QTLs in a cross. There are possible explanations for the absence of such a relationship. First, the size of the effects may vary among QTLs; since different sets of QTLs are segregating in the crosses, this does not necessarily result in a larger genetic variance as the number of segregating QTLs increases. Second, apart from the

identified QTLs, other genes affecting flowering time may be segregating in each cross, inducing additional genetic variance. Although the true cause is unknown, it is obvious that these disturbing factors may have influenced the performance of MAS.

The RILs selected by MAS show, on average, a lower phenotypic value and a higher genotypic value than the RILs selected based on their phenotype, but the differences are small. We conclude that both methods of selection have succeeded in obtaining RIL-pairs that are roughly equal with respect to their breeding potential. In fact the prediction scores, presented in Table 2, seem to corroborate this for the early flowering selection.

This experiment showed that we were able to successfully obtain transgression in offspring populations from selected crosses. Maximum observed flowering time in the $F_2$ populations was twice the maximum value observed in the RIL population. Since these populations were not grown in the same experiment we should be cautious when comparing them. Nevertheless it is clear that the MAS procedure that we used can be applied successfully in other cases as well.

Our MAS procedure (Van Berloo & Stam, 1998) can be seen as aiming at the efficient pyramiding of favourable QTL alleles that are present in a choice of sources, i.e. the RIL set. Both in our simulation study and the experimental verification described in this paper, we have dealt with a single trait, supposedly controlled by non-epistatic QTLs. Since QTLs were mapped in a set of RILs, i.e. no dominance effects could be detected. Had we been able to detect and use dominance at QTLs this would most likely have influenced selection of RIL pairs in MAS. It is quite conceivable that, in the case of non-additivity of QTL effects, pyramiding QTLs based on the phenotype of the parents will be less efficient than pyramiding based on QTL flanking markers. In our previous paper this was demonstrated using simulated data. Although not the subject of this study, another example in which the MAS approach will outperform phenotypic selection is the accumulation of disease resistance (R) genes (e.g. Huang *et al.* 1997), when adding more genes beyond a given number of R-genes does not lead to an observable increase in phenotypic resistance. In that case pyramiding R-genes beyond a phenotypically

observable threshold may nevertheless be useful to enhance the durability of the resistance.

Although in our experiment the results of MAS are falling a little short of expectations, our experiment clearly demonstrates an important, more general, point. That is the potential usefulness of publicly available data on linkage maps and putative QTL positions for breeding purposes. Today this type of data is accumulating at a high rate. Applied plant breeders as well as the scientific community can, and should, take advantage of this information. In the present paper we have considered a single, simple trait, controlled by only a few QTLs. It needs little imagination to realize that in a more realistic setting of plant breeding, where many traits are to be considered simultaneously, knowledge about QTLs and their map positions will be of great help to design and optimise scenarios for the accumulation of favourable QTL alleles by crossing and marker-assisted selection.

**Acknowledgements**

We want to thank Dr. R.C. Jansen for sharing his molecular data and the staff of the Laboratory of Genetics for practical assistance and advise. The Nottingham Arabidopsis Stock Centre kindly provided us with seeds of the Landsberg x Columbia RILs.

# 4 Simultaneous marker-assisted selection for multiple traits in autogamous crops

Ralph van Berloo · Piet Stam

*Submitted for publication in Theoretical and Applied Genetics*

## Introduction

Since most agronomically important traits are quantitative and controlled by several genetic factors, the ability to get more control of the behaviour of these genetic factors by the introduction of linked molecular markers has been very welcome. Recently, 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 studies and improvements (e.g. Hoeschele *et al.* 1997; Henshall & Goddard, 1999; Dupuis & Siegmund, 1999), several fairly reliable procedures have been established and implemented (e.g. Lincoln *et al*. 1992; Holloway & Knapp, 1994; Van Ooijen & Maliepaard, 1996a,b). The next issue to be addressed is efficient use of information on QTLs that 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 & Lande 1994a,b; 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 & Thompson, 1990). Experimental results of applying MAS were discussed in several papers (e.g. Stuber, 1994; Van Berloo & Stam, 1999). Moreau *et al*. (1998), Van Berloo and Stam (1998) and others argued that population size and trait heritability are the key factors influencing MAS results. Tanksley and McCouch (1997) advocated a slightly different use of marker and QTL information. They 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) previously described a procedure for the simultaneous discovery and introgression of QTLs from unadapted germplasm. However, conventional plant breeding has shown continued success with the use of elite material, demon-

strating that there is still room for improvement even within the currently used genetic material. In this paper we try to evaluate how knowledge on QTL-positions can be of use to speed up and increase selection results. This is an extension of our earlier simulation and experimental work (Van Berloo & Stam, 1998; 1999).

## 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' or 'extreme' 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 marker-assisted selection of parental pairs has been implemented in a computer package. The previously implemented selection method (Van Berloo & Stam 1998) was modified and extended in three areas: 1. Selection was applied to several unrelated traits simultaneously; individual trait values were 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 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 10 chromosome pairs, each of length 100 centiMorgan (cM). This design 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. The QTLs were dispersed randomly over the genome. Also, for each locus, the parent contributing the advantageous allele was selected at random. Several of these randomly created genomes were used in each simulation experiment.

**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 |

*Trait Weighing*

Simultaneous selection for multiple quantitative traits introduces a new problem: How to compare different traits? This is in essence an economic question for which the breeder must provide a decision. We simply assigned a weight value to each trait, and created a general index, by weighed summing over traits (Formula 1)

$$GI = \sum_{Traits} (W_T * \sum_{QTLs} q_{T,I}) \quad \textbf{(1)}$$

Where GI is the General Index value, $W_T$ is the weight factor for trait T 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 index used in index selection theory. In the latter, apart from economic weights also trait heritabilities and genetic correlations are used to construct an index that predicts maximum genetic gain with selection.

*Selection Procedure*

The marker-assisted selection procedure is started in a mapping population, in which markers were scored and QTLs identified. The genetic constitution of each plant with regard to the QTLs under study is inferred from QTL-flanking markers. Because we consider several traits, many QTLs are involved. As a result, in an average sized population, the probability that 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 most promising. Here we followed a different but straightforward approach: For all possible line-pairs a test cross is simulated and from this $F_1$ a selfed progeny of sufficient size is derived. 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 line-pair potential value, indicating the potential quality of a line-pair. The line-pair potential value is compared with the most superior parental genotype present among the lines. We define the selection response as the difference between these two values, usually expressed as a percentage. The selection response obtained by applying marker-assisted selection is then compared with the selection response obtained through phenotypical selection.

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, Formula 1 was applied to obtain the 'phenotypic' value of the general index. A procedure, similar to one that could be used to improve already elite material for a single trait, was used to simulate phenotypic selection: Line combinations were assembled 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. Phenotypic selection was thus limited to 100 line pairs, mimicking a breeding program where resources for testing large numbers of progenies are limited.

*Types of populations*

In our previous paper we focussed on effects of heritability and population size in a RIL population. In this paper we discuss also simulations of MAS applied to $F_2$, $BC_1$ and Doubled Haploid (DH) populations. Parental populations contained 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, extra replications were included 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 were assigned equal weights). The second experiment was similar to the first, but now some traits were regarded more important than others. The size of the trait-weight that 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 low 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. 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, only about 15 QTLs were detected. This example illustrates the common knowledge among quantitative geneticists that any single QTL study will usually not be conclusive. 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 did occur. Beavis (1999) therefore suggested to pool available experimental results to obtain a better power of QTL detection. We studied the effect of missing QTL information due to incomplete QTL mapping by means of data removal. Selection was now based on a subset of the QTLs, but genotype and phenotype were still constructed using all 17 QTLs. Repeatedly, a random subset was removed from the list of detected QTLs, rendering selection for these QTLs through linked markers impossible. Again, the same proce dure for determining the quality of pairs of lines was applied, and results were compared with phenotypic selection. Simulations were run for cases in which 3, 5, 7 or 9 QTLs had been deleted from the list of detected QTLs.

## Results

Table 2A 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 marker-assisted selection 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 heterozygousity 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.

**Table 2A**: 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% |

In all cases MAS outperformed 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 Usually, from a breeder's point of view, some traits will be more important than others.

**Table 2B:** 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% |

Such a situation was reflected by the second simulation experiment. The results of this experiment are displayed in Table 2B, the same layout was used for displaying results.

For weighed trait-selection, we expected MAS to show an extra benefit, since lower heritability traits, that are better selectable by MAS than by phenotypic selection, are regarded more important. However, the results do not confirm this expectation. MAS results seem to drop for the more homozygous population types (DH, RIL), while only the results for the $BC_1$ population were better than the situation with equal trait-weights.

*Undetected QTLs*

Simulations that involve MAS based on incomplete QTL data were run for RIL and $BC_1$ populations. The results of these simulations are summarised in Figures 1 and 2. In Figure 1 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 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 Figure 2 the difference in selection response between MAS and phenotypic selection is plotted as a function of the QTL-fraction. This gives us

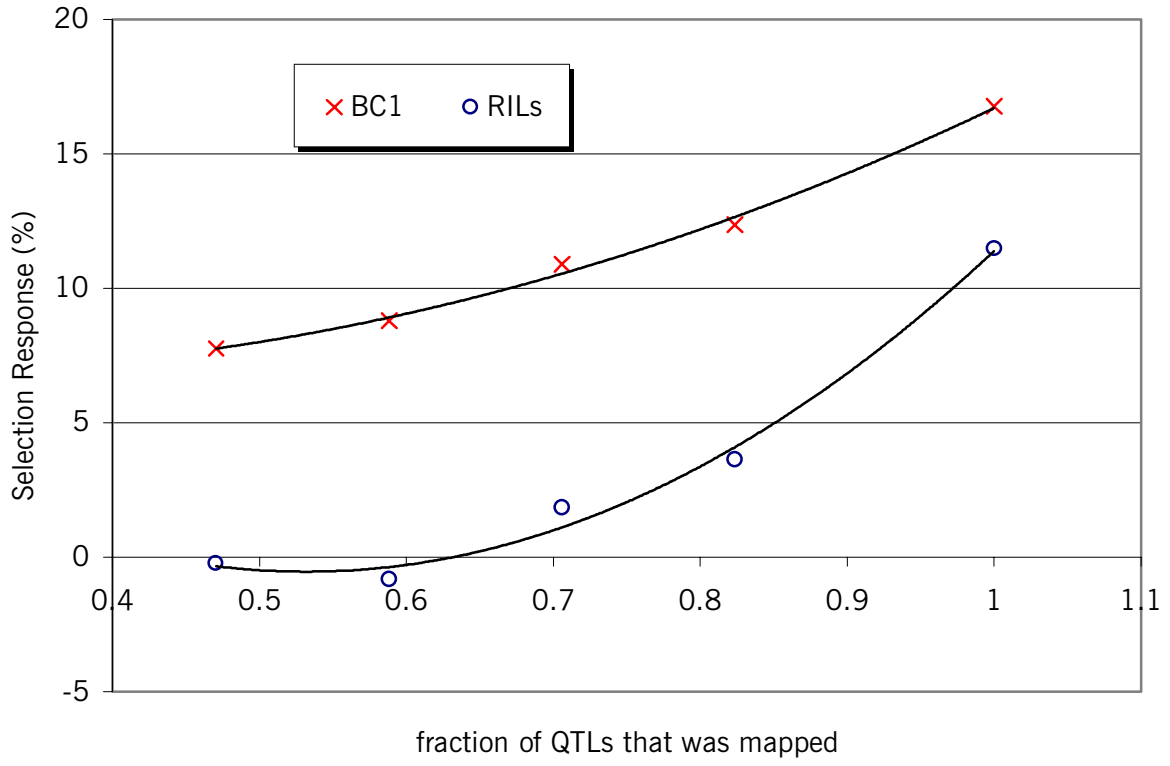



**Figure 1**: MAS Selection response as a function of the "QTL fraction" (fraction of QTLs, present in the model, that were linked to markers.

information in which cases application of MAS may yields 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.

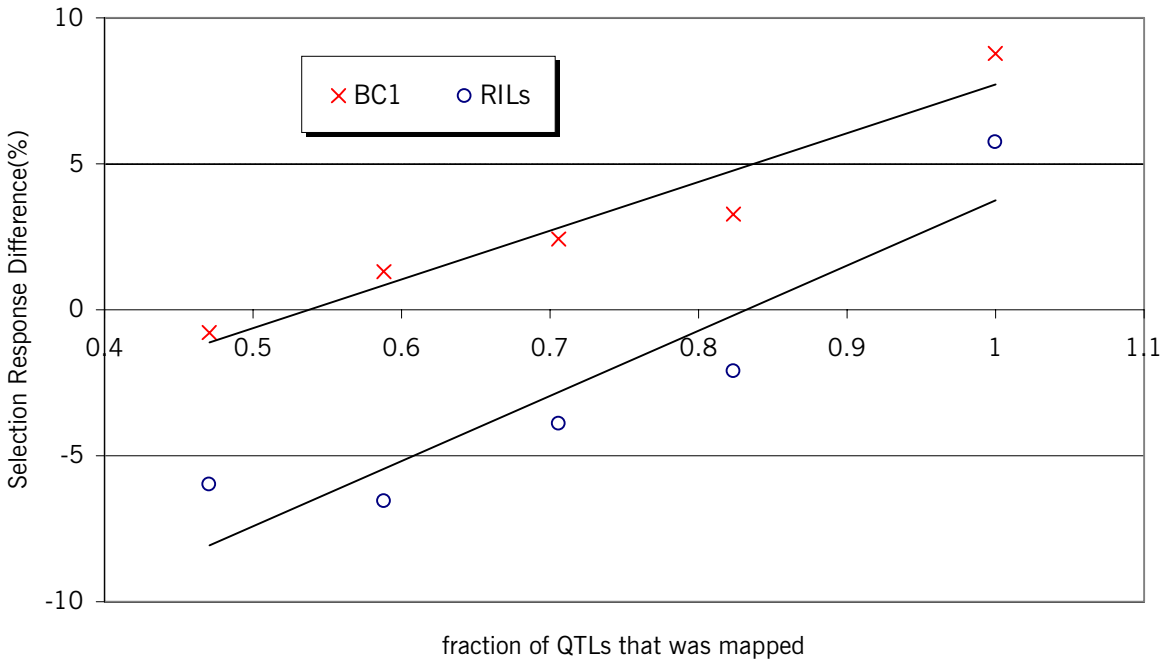



**Figure 2**: Difference in Selection response between MAS and phenotypic selection as a function of the "QTL fraction" (fraction of QTLs, present in the model, that were linked to markers.

## 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 discussed 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, 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.

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, it can be 'manipulated' by repeating and enlarging trials. 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.

The sensitivity of the proposed method to missing QTL-information reduces the possibilities for practical use. Only when extremely good molecular and field data are available and QTLs were mapped reliably, so that only a small fraction of the genetic variance remains unaccounted for, one could 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.

A possible way to compensate for QTLs or polygenes 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 & 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 subject of a future study.

Another factor that may limit 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 is to take two distinct members out of the elite gene-pool used for breeding, and use these to create a new mapping population. The superior genotypes that result from applying MAS in this population will be similar to an improved version of the original elite material, and could be used to replace this material in a conventional breeding program.

This paper only deals with selection of line-pairs, but the model has been extended to allow also selection of combinations of three or four lines. Indeed simulations that included three or four line combinations were run, but a difference with the results from selecting line-pairs could hardly be seen (data not shown). This is because 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. Still, in more complex cases, when more traits and QTLs are involved, 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 dominant QTLs. 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 observed in this study.

Another complicating factor, interaction between QTLs, is usually neglected. Most QTL-mapping software is not yet equipped to detect QTL-interactions. However, more and more information on genes that are positioned at QTLs will become available, for instance from genome sequencing projects. We expect that interaction between QTLs and also QTL x Environment interaction will become more important in the future. A method based on selection of sets of genes through linked markers may be an efficient way to make sure that interacting sets of genes are brought together and 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.

# 5 Marker-assisted selection of RIL-pairs in an *Arabidopsis thaliana* verification experiment

Ralph van Berloo · Hans van Os

## Introduction

Marker-assisted selection is a promising tool for plant breeding. The ability to manipulate genetic factors underlying quantitative traits is appealing and could be used to enhance current plant breeding methods. The implementation of the use of marker technology in commercial breeding requires a serious re-designing of breeding programmes (Stam, 1994). Analytical studies (Lande & Thompson, 1990; Knapp, 1994), simulation experiments (Gimelfarb & Lande, 1994a,b; Hospital *et al.* 1997; Van Berloo & Stam, 1998) as well as field studies (Stuber, 1994; Van Berloo & Stam, 1999; Romagosa *et al.* 1999) demonstrated the usefulness of marker-assisted selection procedures. Knapp (1998) also looked into the cost-effectiveness of marker-assisted selection (MAS). He concluded that MAS could be cost effective if the costs are less than 17 times higher than the costs of phenotypic selection. However, several assumptions favouring MAS were made in this study. Van Berloo and Stam (submitted) investigated marker-assisted selection of pairs of parents, aiming at an accumulation, in the progeny of a cross, of desirable alleles for multiple QTLs in several traits. They ran computer simulations to investigate a realistic case containing traits with a range of heritabilities, each trait being controlled by several QTLs of varying effect size. Comparison of superior genotypes found among the progeny of crosses resulting from marker-assisted selection, and progeny resulting from crosses based on phenotypic selection, showed a higher efficiency of MAS over phenotypic selection. However, some assumptions favouring MAS were made in this study, so the selection efficiency obtained using MAS may be lower in practice. In the present study the simulation results of Van Berloo and Stam were verified in an experiment using the model species *Arabidopsis thaliana*.

## Material and Methods

*Arabidopsis thaliana* has been widely accepted as a useful model species for the study of plant genetics (Koornneef, 1982; Meyerowitz, 1985). Its compact size, small genome and rapid growth have contributed to the popularity of *Arabidopsis*. Alonso-Blanco *et al.* (1998a,b) made an extensive study of a set of 163 recombinant inbred lines (RILs) derived from the ecotypes Landsberg *erecta* and CVI, an ecotype obtained from the Cape Verdi Islands. Over 50 traits were observed and a genetic map containing over 300 markers, mostly AFLPs, was constructed (Alonso-Blanco *et al.* 1998a). For a number of traits, including time to flowering, seed size and other morphological characteristics, QTL mapping studies were conducted , using a map of lesser density, which contained 99 markers (Alonso-Blanco *et al.*1998b). Based on provisional QTL information, we made a selection of traits that were regarded suitable for a marker-assisted breeding experiment. Focus lay on traits with a low heritability that, according to the provisional mapping results, seemed to be controlled by several QTLs. The traits used in the experiment were plant height, number of leaves, length of the longest rosette leaf, number of side shoots, number of branches and germination speed. Raw trait data and molecular marker observations of the 163 RILs, kindly provided by Alonso-Blanco, were used to map QTLs for these traits. QTL analysis was performed by applying the MQM module of the QTL mapping software MapQTL 3.0 (Van Ooijen & Maliepaard, 1996a,b). Based on QTL mapping results, intervals on the genetic map containing a QTL were identified. Markers bordering these intervals were used to discriminate between different intervals. Seeds of the 163 RILs were provided by the Laboratory of Genetics of the Wageningen University (Hanhart & Koornneef, pers. comm.).

*Marker-assisted selection*

Marker data were used to construct genetic models for each recombinant inbred line, similar to the models employed in simulation studies of marker-assisted selection for multiple traits (Van Berloo & Stam, submitted). A list of all possible pairs of RILs was compiled. Next, for all 13203 (½*163*162) pairs of modelled RILs a simulated $F_2$ progeny was derived. For each trait of each simulated plant the following procedure was applied to obtain a genotypic value. At each QTL interval the marker alleles of bordering markers were assessed and, depending on the origin of these markers, the QTL effect of

either Landsberg or CVI was attributed. A trait value was obtained by summing over QTL effects. An overall genotypic value for a plant was calculated by applying an index. Trait values were first normalised according to:

$$t_n = \frac{t - t_{min}}{t_{max} - t_{min}}$$

Where $t$ reflects the trait value, $t_n$ represents the normalised trait value and $t_{max}$ and $t_{min}$ are the maximum and minimum trait values observed among the RILs, respectively.

Equal importance of all traits was assumed, so the overall genotype index was simply calculated as the sum of all normalised trait values. In a population of size 100 the most superior genotype was recorded. This measure was repeated ten times and an average value over ten replications was used as a parameter indicating the potential quality of a RIL-pair. The simulation results of the 13203 RIL-pairs were arranged according to this parameter. Because we wanted to apply divergent selection, approximately 15 RIL-pairs that appeared highest and 15 RIL-pairs that appeared lowest on this list were selected for making crosses. The $F_1$'s resulting from these crosses were selfed and $F_2$ seeds were harvested. Finally $F_2$ seeds originating from two high scoring RIL-pairs and $F_2$ seeds originating from two low scoring crosses were selected for trait evaluation. Beside the potential quality of the RIL-pair, also the availability of sufficient seeds was a factor in this selection.

*Phenotypic selection*

Phenotypic selection was based on the phenotypic data previously observed by Alonso-Blanco *et al.* (1998b and pers. comm.). Trait values were normalised and for each of the 163 RILs an index value was calculated. The ten highest and lowest ranking RILs were used to create pairs of lines. At random 30 pairs of lines were selected from this set for making crosses. For each selection goal, $F_1$ derived $F_2$ plants obtained from two randomly selected crosses were evaluated in the greenhouse. Unfortunately, a serious software error was discovered after analysis of the experimental results. This error has led to the interchange of line-numbers of lines used for phenotypic selection. As a result, some wrong RIL-pairs, derived through phenotypic selection, were evaluated in the trial. This means that the grounds for making comparisons between MAS and phenotypic selection

were lost, and no conclusions could be drawn with regard to this aspect from this study. In the remainder of this study only results that apply to marker-assisted selection will be discussed.

*Selection goals*

Since *Arabidopsis* does not have any agronomic value, the choice of a selection goal was arbitrary. We defined two target phenotypes. A phenotype that we called the 'plus' type: a tall plant with a high number of leaves, long leaves, many branches, many side shoots and also strong dormancy, i.e. low germination. The other target phenotype we defined was the opposite type (referred to as the 'minus' type). A short plant with only a few, short leaves, few branches and side shoots and a good germination. Initially, both MAS and phenotypic selection were used to select crosses for each target phenotype, resulting in four categories of crosses. Each category was represented by two selected crosses (pairs of RILs). Table 1 lists the selected crosses for each category. Coincidentally, some lines were selected for crosses several times. RIL-124 features both in MAS selection for a 'minus' plant type and in phenotypic selection for the 'plus' type. This was caused by the mixup of lines, discussed in the previous paragraph.

**Table 1:** RIL-pairs selected by MAS for making crosses.

| | **MAS** |
|---|---|
| **'Plus' type** | [71x10] & [133 x 10] |
| **'Minus' type** | [124 x 36] & [124 x 125] |

*Evaluation of $F_2$ progenies*

120 seeds of $F_2$ progeny obtained from each cross were sown on moist filter paper in petri-dishes and transferred to pots after germination. To allow estimation of the environmental variance also 240 isogenic seedlings taken from the RILs were planted. Pots were arranged in rows of twelve plants, which were randomised using a complete randomized block design. Standard long-day growing conditions and plant treatment were applied (see Van der Schaar *et al.* 1997 for details). The number of leaves and the length of the longest leaf were recorded when plants started to flower. Plant height, the number

of branches and the number of side shoots were recorded when seeds were maturing. The dormancy of a sample of $F_2$ seeds was determined for each cross in a similar way as described by Van der Schaar *et al.* (1997). Observed trait values were normalised, in the same way as discussed under marker-assisted selection, and an index-trait value was derived. For each $F_2$ population an individual value for the experimental variance was obtained. These values were corrected by subtracting the environmental variance, which was assumed to represent the variance among the isogenic lines. In this way a value for the genetic variance could be obtained for each $F_2$ population. Selection results obtained aimed at different selection targets were compared through statistical analysis.

## Results

*QTL mapping*

The QTL mapping study on the six selected traits revealed 17 QTLs. Figure 1 displays an overview of the location of the detected QTLs on the genetic map. Several regions contained closely linked QTLs or pleiotropic loci. Taking this into account, eleven genomic regions remained that influence traits of interest and are separated at least ten centi-Morgan (cM) from any other region

*Evaluation of $F_2$ plants*

An overview of the phenotypic distribution of populations obtained by MAS are displayed in Table 2. Comparison of results aimed at different selection targets did not show significant differences. Since the selection was focussed on the presence of superior genotypes, the observed averages and standard deviations were translated to a measure for the extreme phenotype (see Van Berloo & Stam, 1999). The resulting confidence intervals for the 5% and 95% percentile for the four MAS populations are shown in Figure 2. The populations aimed at obtaining a 'plus'-type differ significantly at the 95p percentile from population 124x36, but not from 124x125. The 5p percentile was expected to have the lowest value for populations aimed at the 'minus'-phenotype. Confidence intervals overlap at this point, so any differences are not significant.

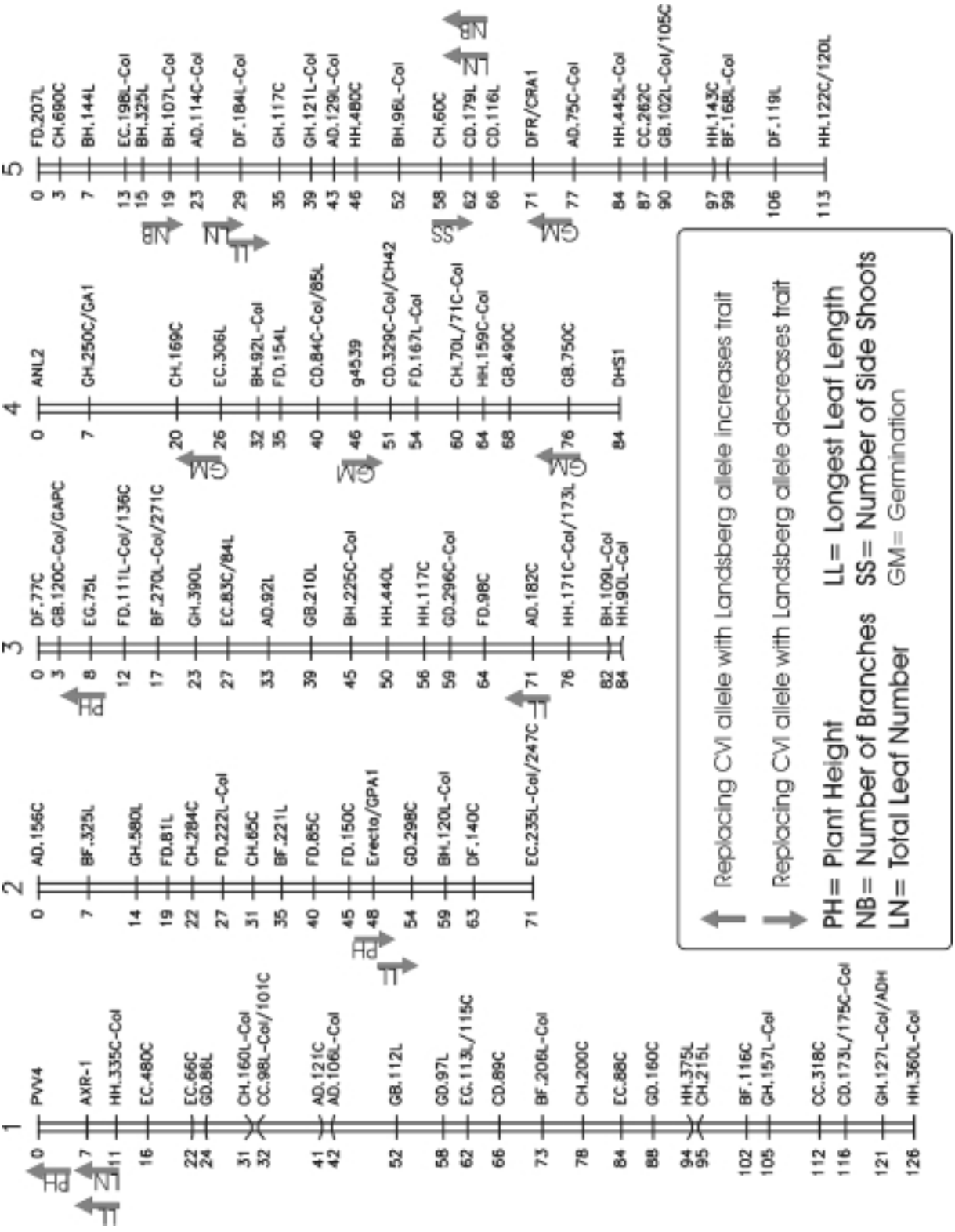



**Fig 1:** Genetic map displaying the locations of the QTLs that were detected in the set of CVI x Landsberg RILs.

**Table 2**: Averages and standard deviations of selected $F_2$ populations.

| $F_2$ Population | Plant Height[1] | Leaf Number | Leaf Length[1] | Branch Number | Side Shoots | Germi-nation[2] | Index |
|---|---|---|---|---|---|---|---|
| 71x10 [MAS plus] | 256 (63) | 7.3 (3.7) | 23.3 (6.4) | 1.7 (0.61) | 1.3 (0.61) | 107 | 1.5 (0.35) |
| 133x10 [MAS plus] | 176 (44) | 8.7 (2.9) | 24.7 (7.8) | 1.5 (0.70) | 2.0 (0.90) | 40.9 | 1.4 (0.40) |
| 124x125 [MAS minus] | 265 (72) | 8.6 (1.7) | 28.5 (7.1) | 1.9 (0.49) | 1.8 (0.71) | 39 | 1.7 (0.35) |
| 124x36 [MAS minus] | 151 (24) | 6.9 (2.1) | 16.5 (5.1) | 1.8 (0.51) | 1.6 (0.60) | 176.3 | 1.1 (0.26) |

[1] Value was measured in mm

[2] Germination data were obtained from an $F_2$ sample; this measurement did not permit estimation of standard deviation. Data were transformed as described in Alonso-Blanco *et al.* (1999b)

Again population 124x125 shows an unexpected high value for the 5p percentile. In general the expected large differences between populations that were raised aiming at different target plant types was not observed. Possible causes for these disappointing results are discussed further on. However, a more detailed inspection of some of the results may be worthwhile.

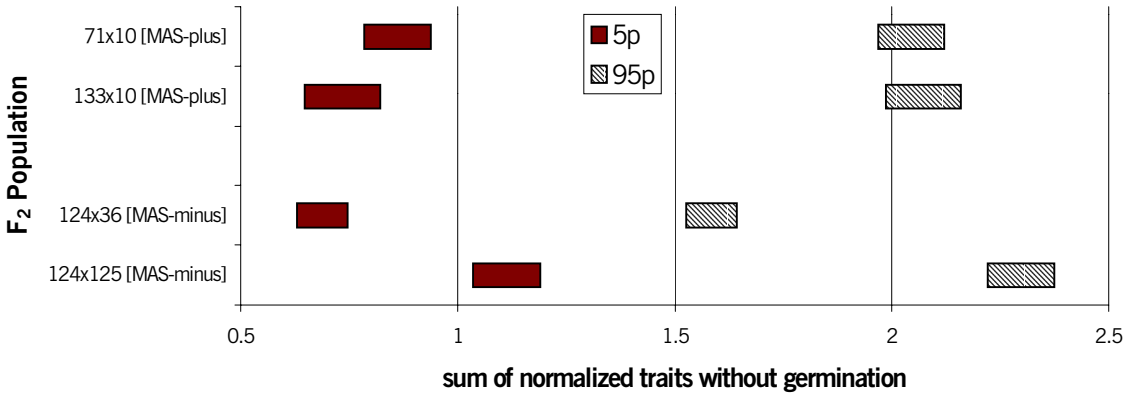



**Figure 2**: Confidence intervals for 5p and 95p percentile points of MAS derived $F_2$ populations at $\alpha=0.05$

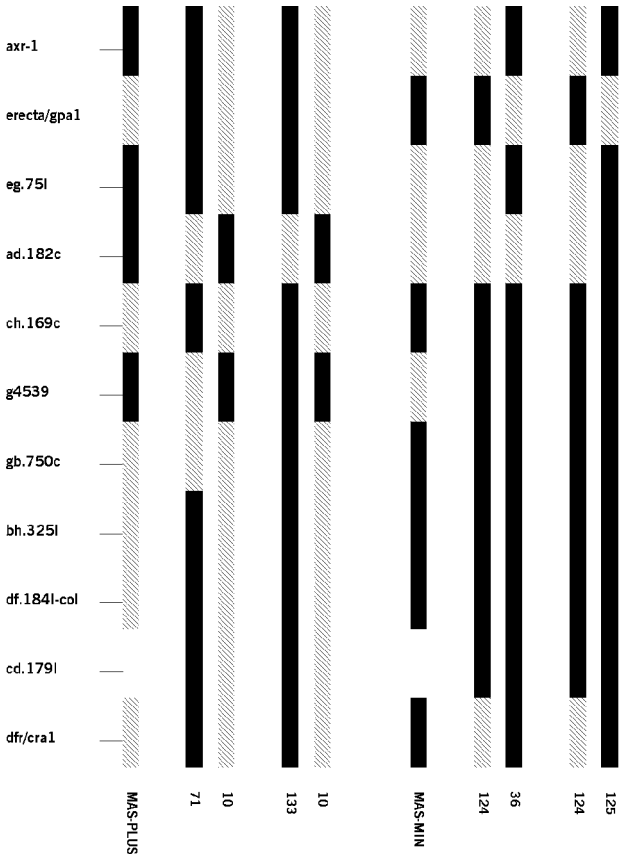



**Figure 3**: Schematic overview of the genetic composition for the 11 regions where QTLs were detected. Dark areas indicate Landsberg derived genome, hatched regions indicate CVI derived genome. MAS-PLUS and MAS-MIN represent the desired genetic composition for the 'plus' and 'minus' plant type, respectively.

*Inspection of selected parents*

Figure 3 presents a schematic overview of the genetic composition of the RILs that were used as parents to create the MAS $F_2$ populations. The eleven genomic regions (residing on different chromosomes, see Fig. 1) that were discussed earlier, were 'drawn' on top of each other using the genetic genotyping software GGT (Van Berloo, 1999a). For each of the eleven regions the composition of the genome is displayed. MAS-PLUS and MAS-MIN show the target genetic configuration for selection in the 'plus' or 'minus' direction, respectively. The desired configuration for the region around marker CD.179L could not be specified unambiguously, since several counteracting QTLs are located in this region. Figure 3 clearly shows that the selected RIL-pairs are highly complementary. For most regions the $F_1$ that is derived from two lines is either fixed for the desired origin or heterozygous, permitting fixation of the desired origin in a segregating population. An exception is the region that is associated with marker g4539. None of the RILs that were selected to obtain the 'plus'-phenotype contained the favourable QTL-allele at this locus. However, the effect of the germination-QTL that is associated with this region was the smallest of all four germination QTLs.

*Additivity of dormancy*

Dormancy, which was measured in our case as the complementary trait, germination, is a highly complex character. In another *Arabidopsis* population, Van der Schaar *et al.* (1997) investigated the genetic basis of dormancy in different environments. These

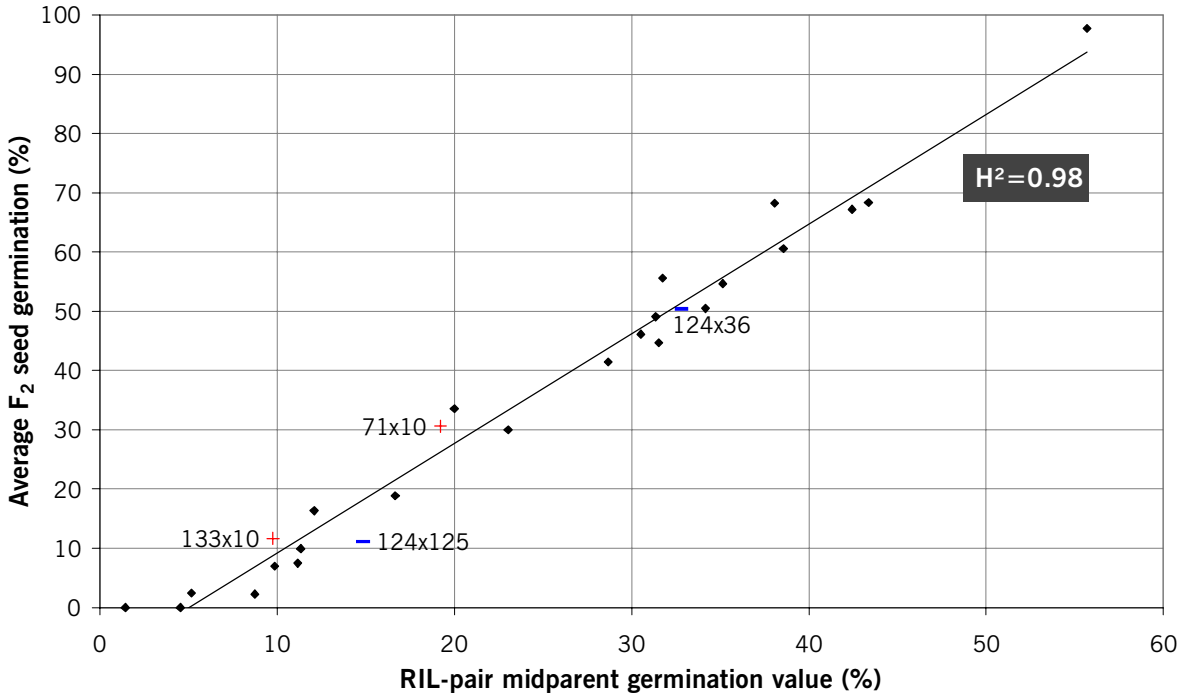



**Figure 4**: Correlation between the midparent germination and the germination observed in the offspring of 31 selected crosses. '+' and '-' symbols indicate crosses aimed at a 'plus' or 'minus' phenotype, respectively.

authors detected 14 QTLs, of which some were only expressed in a specific environment. It is conceivable that the complex background of dormancy results in a complex pattern of inheritance. However, this hypothesis was not confirmed by our results. Seeds from 31 crosses, including the eight crosses discussed earlier, were assessed for germination speed in a similar way as was done by Van der Schaar *et al.* (1997). Figure 4 presents a scatter diagram showing the average normalised germination of the $F_2$ populations plotted against the mid-parent value of the RIL-pairs that yielded the $F_2$ populations. A clear trend can be seen and a high correlation observed. Such a correlation would be expected if additive genetic effects are the most important factors in the inheritance of germination and does not suggest a more complex inheritance.

## Discussion

Because of an error in the phenotypic selection the comparisons between marker-assisted selection for multiple traits and phenotypic selection could not be realized. The experiment did not provide sufficient evidence to confirm an expected difference in

selection results of selection, aimed at different plant-types. If we regard the populations that were selected based on erroneous phenotypic information as being selected at random, and use the observed data in this way to make a comparison between 'random' selection and MAS, our expectation that MAS will result in a higher selection efficiency was not confirmed. A number of factors may have contributed to these discouraging results. The most important factors will be discussed in the following paragraphs.

*Incomplete QTL information*

The marker-assisted selection method applied in our experiment solely relies on accurate information on detected QTLs. Although the mapping population was of reasonable size, some QTLs will have remained undetected. The fact that Van der Schaar *et al.* observed 14 QTLs for dormancy, while we detected only four QTLs, supports this assumption. Studies on the effect of incomplete data on the efficiency of MAS (Van Berloo & Stam, submitted) indicate a rapid loss of the superiority of MAS over phenotypic selection, especially in RIL populations.

*Clustered loci*

As we could observe, several QTLs were very closely linked. If, in reality, a single pleiotropic gene is responsible for the observed variation in different traits, our attempts to separate two loci by means of marker selection are destined to fail. An example of this situation might be the QTLs for branching number and number of side shoots, which were found on chromosome five (near marker CD.179L), closely linked and in repulsion phase. Since the height at which a shoot stems determines our classification as a branch or side shoot, a single gene that influences the height at which shoots develop, might also explain, at least for a large part, the observed variability in the RILs.

*Small sample tested*

Evaluation of selection was performed on populations derived form eight selected crosses. This is only a very small sample out of the potential 13203 RIL-pairs. Additionally, as a result of computer prediction, more than once the same RIL was present in several selected crosses, reducing the genetic diversity even further. Although simulations have shown that, on average, the efficiency of MAS is high, quite a wide range of selection

results for crosses with an equal MAS-derived index can be observed (Van Berloo, unpublished results). By chance our sample may have contained less favourable RIL-pairs. The results found for population derived from RIL-pair 124x125 seem to illustrate this, since these results differed from what was expected in several ways. A larger experimental setup would have reduced the chance of selecting less favourable RIL-pairs, but practical limitations prevented us from performing a larger experiment.

*Unfortunate selection of traits*

Based on heritability and the expected number of QTLs a number of traits was selected. In retrospect, the choice of dormancy (germination) was somewhat unfortunate. The index, derived for the $F_2$ plants, was not as discriminative as it could have been because it was not possible to screen the dormancy of each $F_2$ plant. A solution could have been the determination of dormancy of $F_3$ seeds harvested from the $F_2$ plants. However, this would have been too labourious since in total over 1000 $F_2$ plants were involved. Another complication of the use of dormancy in our selection index was revealed when the $F_1$ seeds were grown. Many crosses yielded only a few germinating seeds, and some crosses were lost at this stage. As a result the options for selection were limited and some crosses, that were probably highly dormant, never yielded $F_2$ seeds which could be used for evaluation.

*Dominance effects*

QTL mapping was performed in a set of RILs, which are homozygous by nature. As a consequence, only additive QTL-effects could be detected. However, our evaluation experiments were based on comparisons between segregating $F_2$ populations. If dominant QTLs are present, a different phenotype than predicted by the additive model results, and selection results will differ from the expected values.

*Too ambitious all-in-one approach*

What can we conclude from the results and the above remarks? Probably our attempt to create a superior genotype in a one step approach, purely based on marker and QTL information, was a little too ambitious. When a single index value is used to represent six traits, the correlation between traits and index may weaken. In the present study a

positive correlation (0.6-0.8) between trait and index was observed for most traits, except germination, which showed no correlation, and the number of side-shoots, which showed a small, negative correlation (-0.15). Still, we think an approach like described in this paper could be applied effectively for general genotypic improvement if we take into account the following requirements. (1) QTLs are mapped reliable and the accumulated genetic variance of the detected QTLs is high, preferably close to 100%. (2) Only two or three traits are involved and these traits share a similar (economic) value, and do not show a negative correlation. (3) Phenotypic data are used to supplement marker data and to decide between alternatives of equal quality, according to the marker derived information. (4) A reasonable set of MAS-derived progenies is screened to reduce the chance of ending up with low performing crosses.

## Acknowledgements

We want to thank Dr. Alonso-Blanco for fruitful discussions and sharing of phenotypic and marker-data, and we would like to display our gratitude to Corrie Hanhart, from the Laboratory of Genetics, for providing us with seed material and practical advise on performing *Arabidopsis* experiments.

# 6 The construction of NILs in barley using marker-assisted backcrossing: experimental and simulation results

Ralph van Berloo, Hanneke Aalbers and Rients E. Niks
*Submitted for publication in Molecular Breeding*

## Introduction

Current developments in molecular and statistical genetics allow estimation of positions and effects of genome fragments responsible for variation in quantitative traits. Such quantitative trait loci (QTLs) are an important and essential source for crop improvement. This improvement can be achieved by making selected crosses within existing elite material in which QTLs have been assessed, followed by marker based selection of superior genotypes. Another option is to aim for the introgression of favourable genome fragments from unadapted material in order to obtain superior trait values, by making selected backcrosses (Tanksley and McCouch, 1997). In the case of backcrosses for introgression of QTLs a marker-assisted approach is required to verify, in each generation, the presence of the favourable allele at the QTL. The presence or absence of a favourable QTL allele cannot be determined by screening the plant phenotype. In most cases no replicated trials are possible due to the small number of plants that make up a backcross population. Also, these backcross populations may differ with respect to their genetic background, and other factors that also influence the trait may still segregate. Screening backcross populations with molecular markers with known positions on a genetic map can provide valuable information. Not only can the origin of the QTL allele be determined, also the remainder of the genome, both linked and unlinked to the QTL, can be monitored. Using a theoretical approach, Stam and Zeven (1981) considered the amount of unwanted donor genome on the same chromosome as the gene of interest in a regular backcross program, without the use of molecular markers to control unwanted linkage. They deduced that in a $BC_6$ backcross plant on average still 32% of the chromosome carrying the introgressed gene will be of donor origin. With the use of markers this figure can be substantially reduced.

The barley genome and actual markers and QTLs were taken as a basis for investigations on several backcross strategies by computer simulation. Simultaneously, a program for the creation of near-isogenic lines (NILs) in barley for three QTLs responsible for partial resistance to leaf rust was started. In this paper analyses of the simulations are compared

with experimental results. Finalisation of the construction of NILs as well as a further genetic and phenotypic characterisation of the obtained NILs will be discussed elsewhere. A short introduction on the background of the used material and the final aim of the research is presented in the next paragraphs.

*Barley leaf rust*

Barley leaf rust (*Puccinia hordei* Otth.) occurs anywhere where barley is grown (Parlevliet 1983). Symptoms of infection are pale spots on the leaves, followed by the emergence of orange brown uredosori that contain fungal spores. Barley leaf rust may cause yield losses up to 30% because of reduced plant photosynthetic capacity and metabolic competition. The disease can be controlled through the use of cultivars containing hypersensitive resistance. The mechanism underlying hypersensitive resistance is still subject of discussion (Kilary & Barna, 1985; Dang *et al.* 1996), but a clear association between abortion of the fungal infection and plant cell necrosis is often observed. Unfortunately, most hypersensitive resistance genes have been rendered ineffective due to rapid adaptation of the pathogen. Another type of resistance is partial resistance. Partial resistance is, in contrast to hypersensitive resistance, controlled by polygenes, and is not associated with plant cell necrosis. The polygenic nature makes breaking of the resistance, due to adaptation of the pathogen, more difficult. Hence this type of resistance is claimed to be more durable (Alemayehu & Parlevliet, 1996; Qi, 1998). Plants that are partially resistant show a reduced rate of infection, compared to susceptible plants, caused by a lower rate of colonisation by the fungus.

*QTLs*

Qi (1998) has reported thirteen QTLs responsible for partial resistance in several populations and stages of plant development. The mechanism behind partial resistance has been studied for some lines with a high level of this resistance, but it has not been feasible to study the effects of the various minor genes. Individual QTLs can only be studied in lines that differ for a single QTL and are identical for the remainder of the genome. Development of such NILs requires a controlled program of backcrossing, ensuring that for only one QTL the resistance allele is present and all other known QTLs carry the allele for susceptibility. Screening plants with molecular markers allows such a controlled program of backcrossing and also allows a more efficient selection against donor genome. This study describes the most important steps in the development of three near-isogenic barley lines each carrying a different QTL for partial resistance to *Puccinia hordei*, by applying marker-assisted backcrossing. The scope of this paper is to describe a

fast procedure for obtaining a specific target genotype and to compare the applied procedure with other backcrossing strategies.

## Materials and Methods

*Resistance QTLs*

Previously, Qi *et al.* (1998a) constructed a high-density molecular linkage map from a set of L94 x Vada derived recombinant inbred lines (RILs). L94 originates from an Ethiopian landrace and is highly susceptible to *Puccinia hordei*. The Dutch cultivar Vada (grown commercially in the early 1960's) shows a high level of partial resistance to *Puccinia hordei.* From the dense map, a skeleton map was derived that was used to map QTLs in the set of RILs.

**Table 1**: Details on location, effect-size and size of estimated QTL support interval of the three QTLs used in this study.

| QTL name | linkage map location | explained phenotypic variance | length of QTL support interval |
|---|---|---|---|
| Rphq2 | chrom. 2, at 185 cM | 4% | ~10 cM (1%) |
| Rphq3 | chrom. 6, at 58 cM | 11% | ~7 cM (0.7%) |
| Rphq4 | chrom. 7, at 6 cM | 45% | ~5 cM (0.5%) |

Data taken from Qi, 1998; explained phenotypic variance refers to the disease score (area under disease progress curve, AUDPC) in adult plants; length of QTL support interval was estimated from the published QTL map.

In total six QTLs for resistance to barley leaf rust were detected in this population (Qi *et al.* 1998b). These QTLs were named Rphq1-6. Figure 1 show the positions of the six QTLs on the genetic map of barley. Our research focussed on three QTLs that showed the largest and most consistent effect in adult plant stage: Rphq2, Rphq3 and Rphq4. The individual properties of these QTLs are listed in Table 1. Together, the three QTLs explained 60% of the observed phenotypic variance for disease severity in adult plant stage.

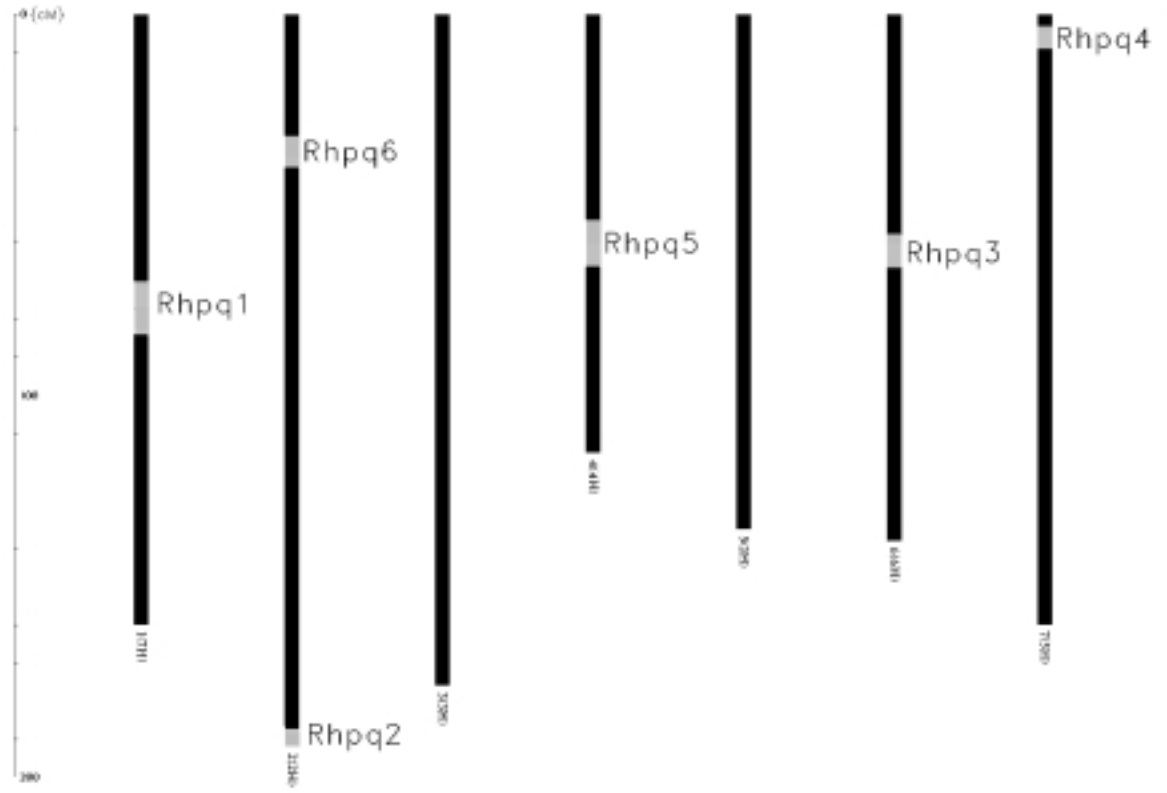



**Figure 1:** locations of the QTLs for partial resistance to *Puccinia hordei* on the Barley genetic map.

*Generation of backcross populations*

A population of 114 backcross plants was obtained from a cross between L94 x (L94 x Vada). DNA samples were extracted from plant leaf material. An adapted protocol was used to obtain AFLP markers in barley (Vos *et al.* 1995; Qi and Lindhout 1997). Due to the nature of a backcross population only AFLP markers that correspond to a Vada-derived amplified fragment are informative, since AFLP is a dominant type of marker and all plants in the backcross population carry at least one allele derived from the recurrent parent, i.e. L94. Ten primer combinations were selected based on the high-density map of Qi *et al.* (1998a). These primer combinations resulted in 107 polymorphic AFLP markers. These markers showed a good overall coverage of the genome, while some markers were located close to known QTL positions. The marker data were arranged in a format commonly used for genetic mapping. Marker loci carrying the L94-allele in homozygous condition were labelled 'A', heterozygous loci were labelled 'H'. Next, we used GGT, a computer-program for the display of Graphical Genotypes (Van Berloo,

1999a), for further analysis of the $BC_1$ plants. Selection of a subset of the plants, suitable for continued backcrossing was performed based on three criteria:

1. The fragment carrying the target QTL had remained heterozygous (i.e. the Vada derived allele was still present).
2. The heterozygous fragment around the QTL was as short as possible.
3. The remainder of the genome showed as much as possible absence of Vada markers.

GGT allowed a quick selection of lines that complied with the first criterion; next, this subset was studied in detail and a further refinement of selection was achieved. The remaining subset contained fifteen plants, which were backcrossed with the recurrent parent (L94). Not all backcrosses were successful. From the obtained $BC_2$ populations the best six populations were selected and up to 30 seeds per population were planted. The $BC_2$ plants were genotyped for a few key markers to determine the presence or absence of the QTL-carrying fragments. About 50% of the plants had lost the Vada derived markers on the QTL-carrying fragment. These plants were discarded. The remaining $BC_2$ plants were genotyped in more detail, using six AFLP primer combinations. An estimate of the genetic constitution of the complete genome was obtained. A subset of the population was selected based on the graphical genotypes obtained through GGT, in a similar way as was done in the $BC_1$ stage. The selected $BC_2$ plants were then backcrossed with L94. $BC_3$ seeds from ten selected $BC_2$ plants were harvested and planted.

At this stage further backcrossing to L94 was no longer required, since the expected amount of remaining Vada genome had decreased substantially. $BC_3$ plants were allowed to self-fertilise and $BC_3S_1$ seeds were harvested. To determine which $BC_3S_1$ populations should be screened in detail, DNA samples were obtained from 92 $BC_3$ plants. seven AFLP primer combinations yielded 56 polymorphic markers. Marker data on six markers from $BC_2$ and $BC_1$ progenitors were used to supplement marker data in areas that were sparsely covered, making use of the fact that markers that were fixed for the L94 allele in an earlier generation could only have transmitted this L94 allele.

*Backcross simulations*

Many procedures alternative to the ad hoc method described above, with a more or less intensive use of markers can be thought of. Hospital *et al.* (1992) and Hospital & Charcosset (1997) published detailed studies on the efficiency of introgression of unlinked QTLs. These authors described analytical and simulation results for optimisation of population sizes, in the case of constant population sizes over generations. Here we describe specific simulations using a more general approach, allowing selection requirements and population sizes to vary between generations. These simulations were set up as follows.

An exact copy of the final map (62 markers) that was obtained in the $BC_3$ of our practical experiment, showing a good coverage of the seven chromosomes, was used as a starting point for genetic simulations. This enabled a proper comparison between experimental and simulated data. Mendelian rules of inheritance and crossover frequency were applied, assuming absence of interference. A cross between two homozygous parental genotypes was simulated, resulting in a hybrid genotype. Next, the hybrid was back-crossed for three generations with the recurrent parent. New plants were added to the backcross population until an individual was found that complied with the given selection demand. However, if such an individual was not found in 5000 plants, it was decided that the experiment was unsuccessful, and no NIL genotype could be obtained. The creation of a cascade of BC populations was replicated 10000 times for each different set of selection demands. In this way we obtained reliable estimates of population statistics and empirical distributions of populations sizes that were required to satisfy the selection demands. Also statistics on the proportion of donor genome were collected. The proportion of donor genome was calculated as the summed map length of donor containing fragments, divided by the total map length. We hereby assumed that all crossovers were located exactly midway of marker locations.

*Selection demands*

The demands used for marker-assisted selection were classified into the following five categories that show an increasing stringency of selection.

A: Selection on the heterozygous condition of the marker(s) that lie within the QTL support interval.

B: as A, but in addition one of the markers flanking the QTL support interval must be homozygous for the recurrent parent. allele

C: as A, but in addition markers flanking the QTL support interval on both sides must be homozygous for the recurrent parent. allele

D: as C, but in addition background selection; for all chromosomes at least 2 markers (positioned at approximately 1/3 and 2/3 of the chromosome) must be homozygous for the recurrent parent allele.

E: Strong background selection; all markers, except the QTL support interval of the target QTL, must derive from the recurrent parent. Markers lying within the QTL support interval region must be heterozygous.

Categories A, B and C can be called 'foreground' selection (emphasizing the desired constitution at and around QTL positions) while D and E include both foreground and background selection (also considering the genomic background). In the case of telomeric QTLs B and C are equivalent.

Apart from demands on the origin of specific markers, overall demands were set in some of the simulations, specifying an upper limit of the genome proportion derived from the donor.

*Backcross strategies*

The procedure that we described above was applied both in the case of a QTL interval positioned at a telomeric location (similar to Rhpq2 and Rphq4) and in the case of a QTL interval located roughly in the middle of the chromosome (similar to Rphq3). In the latter case selection is expected to be less effective since two independent recombination events are needed to free the QTL from linked donor genome. Our simulations allowed a single plant from generation $BC_x$ to transmit any number (including none at all) of offspring to the next generation. Therefore numerous backcross strategies, especially with regard to the intensity of marker-based selection at individual generations, could be screened.

**Table 2**: Specification of the demands used for the different selection strategies; refer to the text for an explanation of the symbols used.

| QTL | I | II | III |
|---|---|---|---|
| Rphq2/4 (Telom) | Type of demands/ %Donor allowed | Type of demands/ %Donor allowed | Type of demands/ %Donor allowed |
| BC1 | C / - | A / 20-40 | - / - |
| BC2 | D / - | A / 3-15 | - / - |
| BC3 | E / - | E / - | -D / 5%;3% |

| QTL | Ia, Ib, Ic | | | II | III |
|---|---|---|---|---|---|
| Rphq3 (Central) | Type of demands/ %Donor allowed | | | Type of demands/ %Donor allowed | Type of demands/ %Donor allowed |
| BC1 | C / - | B / - | B / - | A / 25-35 | - / - |
| BC2 | D / - | D / - | D / - | A / 7,10 | - / - |
| BC3 | E / - | E / - | D/2 | D / 2 | -D / 5;3 |

We limited ourselves to an analysis of only a few possible backcross selection strategies. The details of the demands set in each generation are displayed in Table 2. The strategies can be divided into the following categories.

I foreground selection in the $BC_1$; foreground selection + weak background selection in the $BC_2$; selection for true NIL in $BC_3$

II in $BC_1$ and $BC_2$: The presence of the target QTL allele is required and the amount of donor genome must be reduced (a range of allowed donor fractions was used); selection for true NIL in $BC_3$

III no selection in $BC_1$ and $BC_2$, maintaining a population derived through SSD of 300 plants; selection for the QTL interval in $BC_3$ allowing 3% or 5% of remaining donor-genome.

In the generation obtained after selfing the $BC_3$ ($BC_3S_1$) 25% of the progeny should show the true NIL genotype, with regard to the target QTL. Therefore, a situation like defined under 'E' in the section dealing with selection demands is required. But, for some cases where this goal could not be reached, a situation as defined under 'D', supplemented with demands on the allowed proportion of donor genome, was regarded satisfactory.

## Results

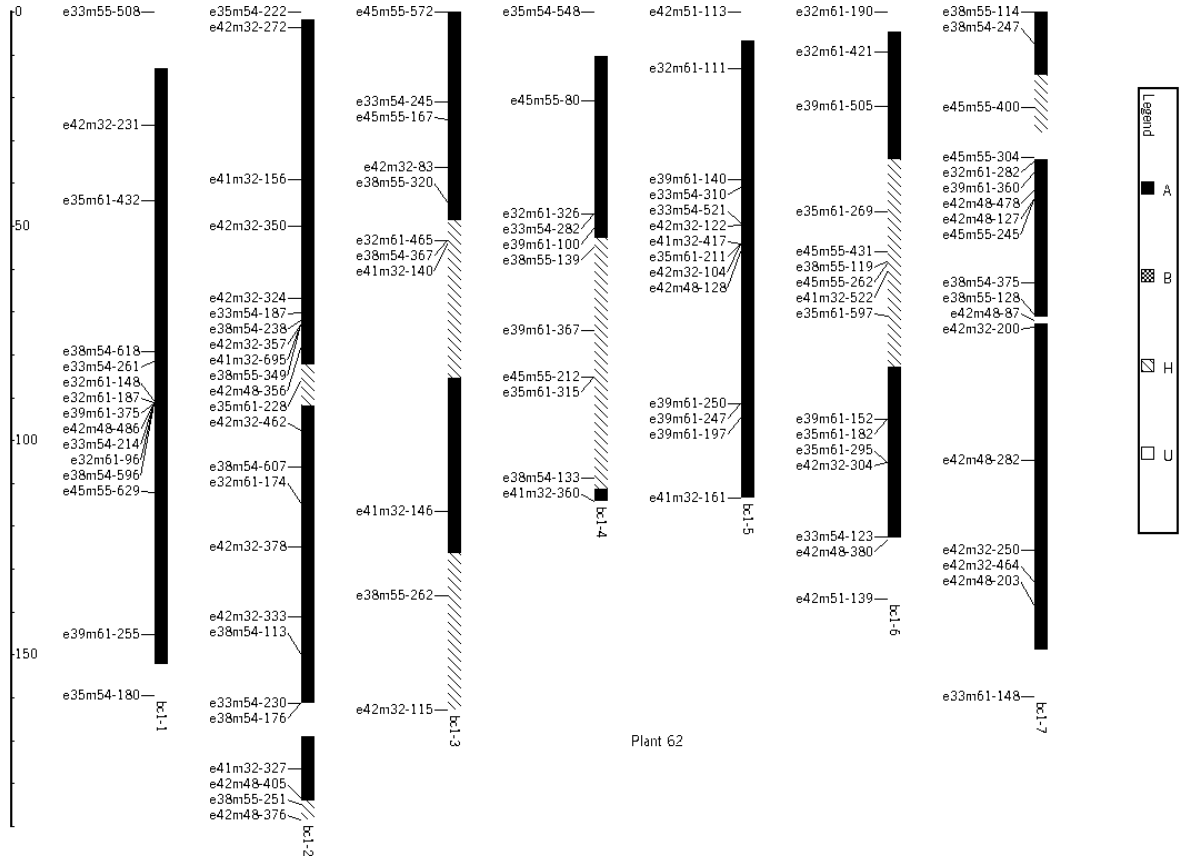
*Experimental Results*

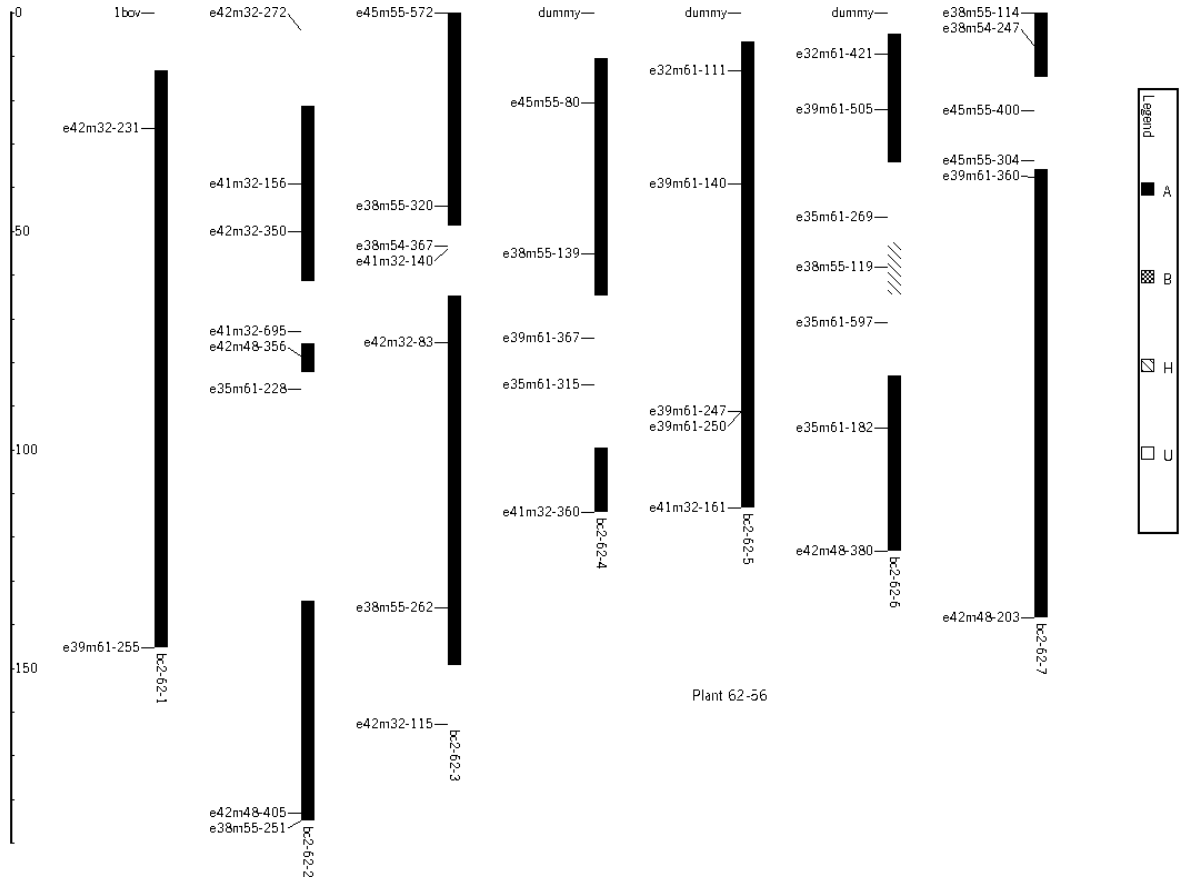
Table 3 lists overall statistics on the experimental results of the marker-assisted backcross procedure. The percentage of remaining donor genome and the number of unwanted donor fragments are listed.

**Table 3:** Proportion of recipient genome and number of retained donor fragments observed in the pedigree of selected $BC_1$, $BC_2$ and $BC_3$ plants.

| **$BC_1$** | **$BC_2$** | **$BC_3$** | Target |
|---|---|---|---|
| Overall [48%] | Overall [81.8%] | Overall [92.4%] | QTL |
| 13 [62% - 9] | 13-8 [88% - 4] | <none selected> | |
| | 13-30 [88% - 4] | 13-30-4 [91.3% - 2] | Rphq4 |
| | | 13-30-5 [92.8% - 3] | Rphq4 |
| 62 [76% - 5] | 62-46 [92% - 3] | 62-46-3 [96.5% - 1] | Rphq2 |
| | | 62-46-8 [98.7% - 0] | Rphq2 |
| | 62-50 [92% - 2] | <none selected> | |
| | 62-56 [93% - 1] | 62-56-3 [97.0% - 0] | Rphq3 |
| | | 62-56-8 [97.2% - 0] | Rphq3 |
| | | 62-56-9 [97.2% - 0] | Rphq3 |
| 63 [51% - 6] | 63-63 [87% - 6] | 63-63-6 [93.8% - 1] | Rphq3 |
| 67 [54% - 8] | 67-100 [84% - 8] | 67-100-1 [85.1% - 3] | Rphq4 |
| 72 [65% - 10] | 72-108 [83% - 2] | <none selected> | |
| | 72-116 [83% - 2] | 72-116-4 [93.1% - 1] | Rphq3 |
| 89 [40% - 7] | 89-141 [82% - 3] | 89-141-3 [84.9% - 2] | Rphq4 |

Numbers: Plant numbers of selected plants; Within square brackets: proportion of genome derived from the recipient – number of remaining unwanted donor fragments.


0
50
100
150
e33m55-508
e42m32-231
e35m61-432
e38m54-618
e33m54-261
e32m61-148
e32m61-187
e39m61-375
e42m48-486
e33m54-214
e32m61-96
e38m54-596
e45m55-629
e39m61-255
e35m54-180
bc1-1
e35m54-222
e42m32-272
e41m32-156
e42m32-350
e42m32-324
e33m54-187
e38m54-238
e42m32-357
e41m32-695
e38m55-349
e42m48-356
e35m61-228
e42m32-462
e38m54-607
e32m61-174
e42m32-378
e42m32-333
e38m54-113
e33m54-230
e38m54-176
e41m32-327
e42m48-405
e38m55-251
e42m48-376
bc1-2
e45m55-572
e33m54-245
e45m55-167
e42m32-83
e38m55-320
e32m61-465
e38m54-367
e41m32-140
e41m32-146
e38m55-262
e42m32-115
bc1-3
e35m54-548
e45m55-80
e32m61-326
e33m54-282
e39m61-100
e38m55-139
e39m61-367
e45m55-212
e35m61-315
e38m54-133
e41m32-360
bc1-4
e42m51-113
e32m61-111
e39m61-140
e33m54-310
e33m54-521
e42m32-122
e41m32-417
e35m61-211
e42m32-104
e42m48-128
e39m61-250
e39m61-247
e39m61-197
e41m32-161
bc1-5
e32m61-190
e32m61-421
e39m61-505
e35m61-269
e45m55-431
e38m55-119
e45m55-262
e41m32-522
e35m61-597
e39m61-152
e35m61-182
e35m61-295
e42m32-304
e33m54-123
e42m48-380
e42m51-139
bc1-6
e38m55-114
e38m54-247
e45m55-400
e45m55-304
e32m61-282
e39m61-360
e42m48-478
e42m48-127
e45m55-245
e38m54-375
e38m55-128
e42m48-87
e42m32-200
e42m48-282
e42m32-250
e42m32-464
e42m48-203
e33m61-148
bc1-7
Legend
A
B
H
U
Plant 62


0
50
100
150
1bov
e42m32-231
e39m61-255
bc2-62-1
e42m32-272
e41m32-156
e42m32-350
e41m32-695
e42m48-356
e35m61-228
e42m48-405
e38m55-251
bc2-62-2
e45m55-572
e38m55-320
e38m54-367
e41m32-140
e42m32-83
e38m55-262
e42m32-115
bc2-62-3
dummy
e45m55-80
e38m55-139
e39m61-367
e35m61-315
e41m32-360
bc2-62-4
dummy
e32m61-111
e39m61-140
e39m61-247
e39m61-250
e41m32-161
bc2-62-5
dummy
e32m61-421
e39m61-505
e35m61-269
e38m55-119
e35m61-597
e35m61-182
e42m48-380
bc2-62-6
e38m55-114
e38m54-247
e45m55-400
e45m55-304
e39m61-360
e42m48-203
bc2-62-7
Legend
A
B
H
U
Plant 62-56

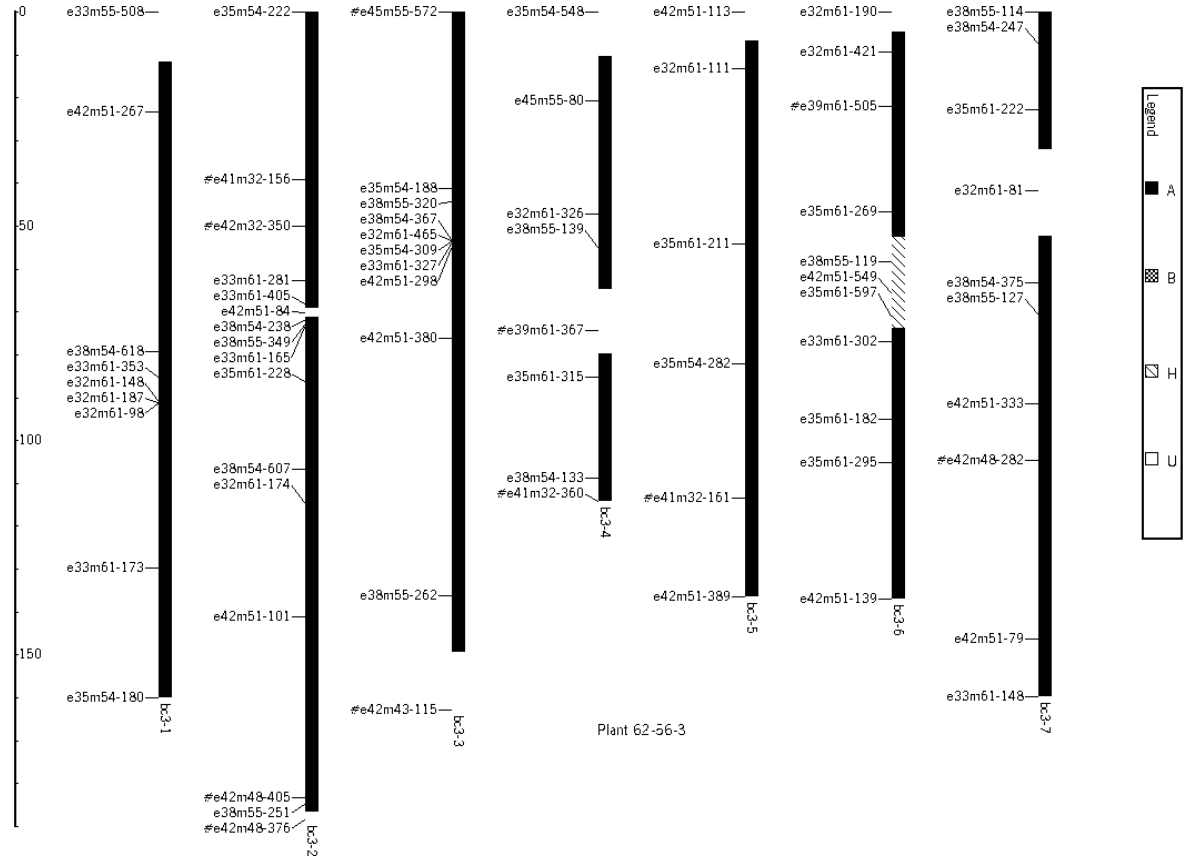



**Figure 2 (A-C):** Graphical genotypes for selected plants in $BC_1$ (2A; plant 62), $BC_2$ (2B; plant 62-56) and $BC_3$ (2C; plant 62-56-3) for introgression of Rphq3, located on chromosome 6. Explanation of legend symbols: A= homozygous L94, B=homozygous Vada, H=heterozygous, U=unknown

Bearing in mind that the expected proportion of recipient genome, when no selection is applied, is 50% in the $BC_1$; 75% in the $BC_2$ and 87.5% in the $BC_3$, the effect of MAS on the selection result is clear. For Rphq2 and Rphq3, genotypes with an acceptable amount of remaining donor genome (< 3%) were present in the $BC_3$ generation. For Rphq4 we were less successful. It will be necessary to select among a larger number of progeny obtained from selfed $BC_3$ plants to obtain an acceptable NIL genotype. It is expected that 25% of the progeny that is obtained from selfing selected $BC_3$ plants will contain the Vada allele at the target QTL in homozygous form, in a L94 genetic background. Preliminary results from disease tests on a sample of the $BC_3S_1$ plants confirm these

expectations. In the case of plants BC3-13-30-5, BC3-63-63-6 and BC3-89-141-3, on two other chromosomes than the chromosome with the target QTL Vada fragments are still present. Statistically it is expected for each of these fragments that only 25% will become homozygously L94. This means that only $(¼)^3$ (1 out of 64) of the progeny will qualify as NIL. Still, when a large enough number of seedlings is screened it is expected that a suitable individual can be found. Plant 67-100-1 is a special case. Although this plant still contains a high amount of donor genome, a recombination within the QTL interval makes it interesting for future QTL fine-mapping studies. Figures 2A-C display the graphical genotypes of plants BC1-62, BC2-62-56 and BC3-62-56-3. These images illustrate the steps that led to one of the selected $BC_3$ plants and show the introgression of the Vada allele for QTL Rphq3.

*Simulation Results - Telomeric QTL*

We start with a discussion of the results in the case of a telomeric target QTL. Strategy I (see Table 2 for selection strategy details) was successful in 99.8% of the cases. We used the median of the observed values for comparison because the number of plants, that was required before selection criteria were fulfilled, showed a truncated Poisson type of distribution. Median population sizes were 11 in the $BC_1$, 78 in the $BC_2$ and 15 in the $BC_3$ generation. These numbers are within the range that can be handled practically, indicating that strategy I, which is the most straightforward approach, also has practical value. The median for the total number of plants that were required over the three generations (104) is very reasonable. However, we expect a more flexible strategy can do better.

Strategy II was translated into a series of simulations with a range of allowed proportion of donor genome both in $BC_1$ and $BC_2$. In all cases the demands in the $BC_3$ were a heterozygous target QTL in a completely homozygous recipient background. A low overall population size, summed over the three generations, and a stable population size in each generation were preferred. Table 4 lists detailed results obtained when applying strategy II. These results demonstrate that a more or less constant population size is not automatically achieved.

**Table 4:** Tabulated summary of Strategy I & II simulation results.

| | | | $BC_1$ | | $BC_2$ | | $BC_3$ | |
|---|---|---|---|---|---|---|---|---|
| Selection[1] | Success[2] | ∑(med)[5] | ltv[3] | med[4] | ltv | med | ltv | med |
| QT-C-D-E | 99.8 | 104 | 1-46 | 11 | 4-910 | 78 | 1-231 | 15 |
| QT-A:20%-A:3%-E | 100 | 401 | 12-1905 | 343 | 2-697 | 56 | 1-21 | 2 |
| QT-A:25%-A:3%-E | 100 | 223 | 4-548 | 101 | 3-1326 | 120 | 1-21 | 2 |
| QT-A:30%-A:3%-E | 99.6 | 257 | 2-211 | 38 | 4-2495 | 217 | 1-20 | 2 |
| QT-A:35%-A:3%-E | 97.7 | 426 | 1-88 | 16 | 10-3667 | 408 | 1-20 | 2 |
| QT-A:40%-A:3%-E | 91.8 | 620 | 1-46 | 9 | 13-4206 | 609 | 1-19 | 2 |
| QT-A:20%-A:5%-E | 100 | 366 | 13-1876 | 348 | 1-98 | 12 | 1-71 | 6 |
| QT-A:25%-A:5%-E | 100 | 134 | 4-549 | 105 | 1-200 | 23 | 1-75 | 6 |
| QT-A:30%-A:5%-E | 100 | 87 | 2-195 | 38 | 2-390 | 43 | 1-72 | 6 |
| QT-A:35%-A:5%-E | 100 | 100 | 1-88 | 17 | 2-724 | 77 | 1-72 | 6 |
| QT-A:40%-A:5%-E | 99.6 | 146 | 1-44 | 9 | 4-1342 | 131 | 1-72 | 6 |
| QT-A:20%-A:7%-E | 100 | 372 | 12-1882 | 357 | 1-39 | 6 | 1-119 | 9 |
| QT-A:25%-A:7%-E | 100 | 121 | 4-561 | 102 | 1-78 | 10 | 1-130 | 9 |
| QT-A:30%-A:7%-E | 100 | 67 | 2-201 | 39 | 1-140 | 18 | 1-132 | 10 |
| QT-A:35%-A:7%-E | 100 | 57 | 1-93 | 17 | 1-267 | 30 | 1-135 | 10 |
| QT-A:40%-A:7%-E | 100 | 67 | 1-42 | 9 | 2-445 | 48 | 1-136 | 10 |
| QT-A:20%-A:10%-E | 100 | 367 | 13-1853 | 350 | 1-15 | 3 | 1-237 | 14 |
| QT-A:25%-A:10%-E | 100 | 125 | 4-555 | 104 | 1-25 | 4 | 1-237 | 17 |
| QT-A:30%-A:10%-E | 100 | 62 | 2-205 | 37 | 1-43 | 7 | 1-240 | 18 |
| QT-A:35%-A:10%-E | 100 | 45 | 1-86 | 17 | 1-76 | 11 | 1-233 | 17 |
| QT-A:40%-A:10%-E | 100 | 46 | 1-44 | 9 | 1-124 | 17 | 1-268 | 20 |
| QT-A:20%-A:15%-E | 100 | 367 | 13-1853 | 350 | 1-15 | 3 | 1-237 | 14 |
| QT-A:25%-A:15%-E | 100 | 135 | 4-566 | 104 | 1-10 | 2 | 1-496 | 29 |
| QT-A:30%-A:15%-E | 100 | 75 | 2-202 | 38 | 1-14 | 3 | 1-540 | 34 |
| QT-A:35%-A:15%-E | 100 | 62 | 1-85 | 17 | 1-21 | 4 | 1-666 | 41 |
| QT-A:40%-A:15%-E | 100 | 53 | 1-43 | 9 | 1-32 | 5 | 1-618 | 39 |
| QC-C-D-E | 60 | 169 | 4-465 | 88 | 2-1058 | 62 | 1-596 | 19 |
| QC-B-D-E | 52 | 448 | 1-49 | 9 | 6-4019 | 420 | 1-682 | 19 |
| QC-B-D-D:2% | 96.0 | 318 | 1-70 | 14 | 5-3362 | 280 | 1-922 | 24 |
| QC-A:25%-A:7%-D:2% | 100 | 194 | 6-804 | 148 | 1-174 | 18 | 1-621 | 28 |
| QC-A:30%-A:7%-D:2% | 100 | 113 | 2-265 | 49 | 1-357 | 36 | 1-681 | 28 |
| QC-A:35%-A:7%-D:2% | 100 | 115 | 1-104 | 20 | 2-683 | 66 | 1-651 | 29 |
| QC-A:40%-A:7%-D:2% | 100 | 152 | 1-50 | 10 | 3-1233 | 115 | 1-657 | 27 |
| QC-A:25%-A:10%-D:2% | 100 | 214 | 6-797 | 150 | 1-42 | 6 | 1-1192 | 58 |
| QC-A:30%-A:10%-D:2% | 100 | 122 | 2-255 | 50 | 1-76 | 10 | 2-1197 | 62 |
| QC-A:35%-A:10%-D:2% | 100 | 101 | 1-109 | 22 | 1-145 | 17 | 2-1299 | 62 |
| QC-A:40%-A:10%-D:2% | 100 | 102 | 1-50 | 10 | 1-244 | 28 | 1-1286 | 64 |

1 Selection specification is arranged as: QTL – $BC_1$ sel type :allowed donor% – $BC_2$ sel type :allowed donor% – $BC_3$ sel type:allowed donor%; QT= telomeric QTL, QC = central QTL. See text for explanation on selection type coding.

2 Success: percentage of replicated simulations that resulted in progeny meeting the selection demands in a progeny < 5000 plants

3 ltv: observed 95% confidence interval for the number of plants that was required before selection was satisfied.

4 med: median of the required number of plants.

5 ∑(med): sum of median of required number of plants in $BC_1$, $BC_2$ and $BC_3$.

In many cases a clear trade-off between selection criteria in different generations can be seen. A less-stringent selection in one generation, that can be fulfilled using only a few plants, will be followed by a more stringent selection in the next generation, which requires many more plants before selection criteria are met. It is therefore advisable to find a balance that requires not only a few plants, accumulated over generations, but also approximately the same number of plants in each generation.

Figure 3 displays the sum of the median required population sizes plotted against the percentage of donor genome that was allowed in $BC_1$ and $BC_2$. We clearly see that less plants are needed to obtain a NIL when less stringent demands are set for the remaining donor genome. An optimal situation, with a minimal amount of plants, is reached when 35% donor genome is still allowed in the $BC_1$ and 10% in the $BC_2$. But neighbouring sets of selection settings resulted in a similar performance and may be preferred if they show a more constant population size over generations.

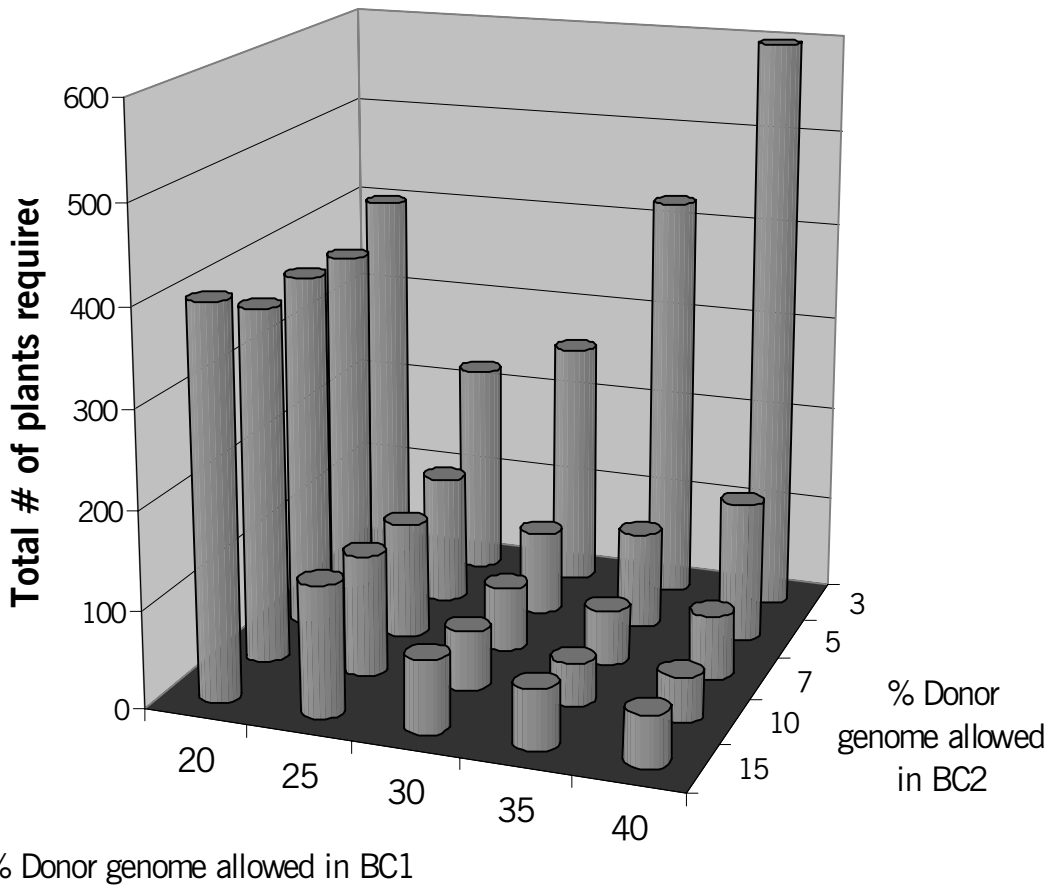



**Figure 3:** sum of population sizes in $BC_1$, $BC_2$ and $BC_3$ as a function of the variable demands of strategy II.

Strategy III was different from the previous strategies because a fixed number of plants (300) was used in all generations. Each plant contributed only one new plant to the next generation (single seed descent). No selection was applied in $BC_1$ and $BC_2$ generations, and foreground as well as background marker selection (D) was applied in the $BC_3$. Also in the $BC_3$ the proportion of donor genome was allowed to be at most 5% or 3%, respectively. Strategy III is clearly less efficient than Strategies I and II. Although larger populations were used and less stringent final genotypes defined, the success rate for this strategy was only 70% (when 5% donor genome was allowed) and 48% (when 3% donor genome was allowed). This means there is a fair chance that the target genotype will not be found. The biggest advantages of strategy III are the minimal requirements with regard to the amount of genotyping that has to be performed. It is only in the final generation that plants are genotyped. But, the larger population size results in more $BC_3$ samples that need to be genotyped. Also, the reduced labour needed for genotyping may be counterbalanced by an increase in the amount of labour required for backcrossing. Still the simplicity of the method may be appealing, especially for species where backcrossing is less labour-intensive than in barley, and if final demands on the target genotype are not very extreme.

*Simulation Results - Central QTL*

Very similar simulation experiments were performed in the case of a QTL located close to the centre of the map (for details see Table 2). For this configuration a lower success rate is expected, since it takes two independent recombination events to separate the donor QTL-allele from the surrounding genome. This is clearly confirmed by the simulation results. Strategy III performed worst. Only 37% (when 5% donor genome was allowed) or 19% (when 3% donor genome was allowed) of the simulations were successful. These success rates are unacceptably low. Strategy Ia and Ib gave a success rate of only 60% and 52% respectively (Table 4). This was mainly due to the demands set in the $BC_3$ generation. We therefore decided to relax the selection criterion in the $BC_3$ by applying a combined foreground and background selection (D), allowing at most 2% donor genome to remain. Strategy Ic reflects this situation. Strategy Ic is still quite a stringent selection, if we take into account that the target QTL region itself represents 0.7% of the donor genome. Strategy Ic was successful in 96% of the cases and, although

selection was less intense, we also find a higher average recipient genome content than observed in strategies Ia and Ib. Strategy II resulted in similar results as discussed for the telomeric QTL. Detailed results are listed in Table 4. When demands were set to a maximum of 35% donor genome in the $BC_1$, 10% in the $BC_2$ and 2% in the $BC_3$ with foreground and background selection, the sum of plants required in $BC_1$, $BC_2$ and $BC_3$ reached a minimum of 101.

## Discussion

In the case of partial, polygenic resistance, marker-assisted selection provides a valuable tool to identify and manipulate underlying genetic factors. One of the possibilities for the use of marker-assisted selection is the pyramiding of resistance genes (e.g. Huang *et al.* 1997). The pyramiding of resistance genes could result in more durable resistant genotypes. One could even think of using marker-assisted selection to enhance the resistance-*durability* by adding resistance genes to a genotype that already shows a resistant phenotype, since it is expected that extra resistance genes will make breaking of resistance by adaptation of the pathogen more difficult.

The experimental procedure we used to develop near isogenic lines is similar to the recommended procedure by Hospital *et al.* (1992), giving focus during the first stages of backcrossing to proximal recombination events. However, the number of markers we used exceeded the recommendations. This is due to the nature of AFLP markers. A single primer-combination, which was selected because it yielded markers proximal to the desired QTL interval, also yielded marker loci elsewhere. Moreover, even in the $BC_3$ generation we found that applying a generous set of markers enabled us to be more selective and also to identify interesting individuals that showed a recombination event within the QTL-supporting interval. Such individuals could be interesting for QTL fine mapping and gene cloning. When comparing the selection intensity that we applied in the barley experiment with our simulation results, in retrospect, selecting a larger set of plants in the $BC_1$ generation might have improved selection results, and reduced the number of required plants in later generations.

In this paper a stepwise method is discussed for the creation of near isogenic lines for QTLs. Clear benefits of this method are a high level of control with regard to the backcross process. Also the number of labour intensive backcrosses could remain limited. However, quite some effort for the determination of molecular marker data is required. There is a clear trade-off between an increased investment in laboratory work and an investment in making backcrosses and increasing population sizes. The balance between these factors will depend on the relative ease with which backcrosses can be made and markers can be scored, and will differ between crops and populations. We wanted to verify if the proposed method is also efficient, and if other methods could yield similar results with less effort. Our computer simulations of a number of alternative strategies showed that, when focus lies on a single gene, more efficient procedures than practised in our experiment are possible. The population sizes we used were larger than necessary, according to the simulations. The main cause for this was a time limitation. DNA isolation and marker analyses needed to be completed before plants were flowering and back-crossed. In some cases full marker information was not yet available at the time the backcrosses had to be made. This resulted in decisions that were made on the basis of incomplete data. In such cases extra plants and crosses were included to be on the safe side. Simultaneous selection for several QTLs was not taken into account in the decisions made during computer simulations. However, in the barley experiment a plant (BC1-62) was selected because it could serve as a NIL progenitor for two QTLs in a later generation (Table 3).

We found that the creation of a NIL in three generations of backcrossing poses no problems, when using MAS. When the selection intensity is not too strong during the first generations of backcrossing, sufficient variation remains to allow selection of a NIL-genotype using strict criteria in the final generation, without the need for excessive population sizes. This situation was found to be optimal in this genetic background.

For practical reasons our experimental and simulation studies were performed simulta-neously. A more optimal situation would be when simulation studies are followed by practical experiments. An investment of some time and resources to explore the possible

options through simulation may well increase the efficiency and enhance the results of practical experiments.

# 7 The development of software for the graphical representation and filtering of molecular marker data: graphical genotypes (GGT)

*An abbreviated version of this paper was published in the Journal of Heredity 90:328-329 (1999).*

## Introduction

In the early days of genetics, the options for obtaining relevant genetic information were scarce. Morphological markers and isozymes were a welcome tool to aid geneticist in unravelling the genetic background of field observations. Nowadays, molecular markers have introduced many new possibilities to increase our understanding of the genetic constitution of plants and animals (Tanksley, 1993). High throughput marker systems have become standard equipment in many laboratories, and a huge amount of genetic data is produced every day. As the retrieval of molecular marker data is no longer limiting, data interpretation becomes more significant. Efficient use of DNA markers for genomic research and crop improvement will depend as much on computational tools as on laboratory technology (Nelson, 1996). Computer tools that assist in the analysis of molecular data have become important, since the analysis 'by hand' is too labourious for the large numbers of data involved. Visualization of molecular data can help geneticists to improve their understanding and to apply selection more efficiently. Young and Tanksley (1989) described an application for visualization of molecular marker-data, introducing the concept of graphical genotypes. The depiction of marker-genotypes in a graphical way was also included in the genomic software package QGene (Nelson, 1997). Recently, the services of commercial biotechnology companies also include depiction of molecular data (Keygene, 1999). The arrival of a large molecular data set at our laboratory introduced the need for a visualization tool. The development of a simple tool was started, and gradually this tool was extended to a versatile piece of software that was named GGT (short for graphical genotypes). The options and features of GGT are described in the next paragraphs.

## Methods and Features

*Visualization*

The main function of GGT is to visualize molecular marker data. It is required that the position on the genetic map of all markers is known; other genetic software like Mapmaker (Lander *et al.* 1987) or Joinmap (Stam, 1993) can be used to construct such a map. Visualization is done by drawing the genetic map. Regions of the map are drawn in different colors or hatch patterns, depending on the allelic compositions of the markers that are located in the region. A change of allelic composition is reflected by a gradual change of colors. However, sometimes a clearer image is obtained when sharp boundaries are drawn. This is optional, and results in an sudden color change midway of two markers of different origin.

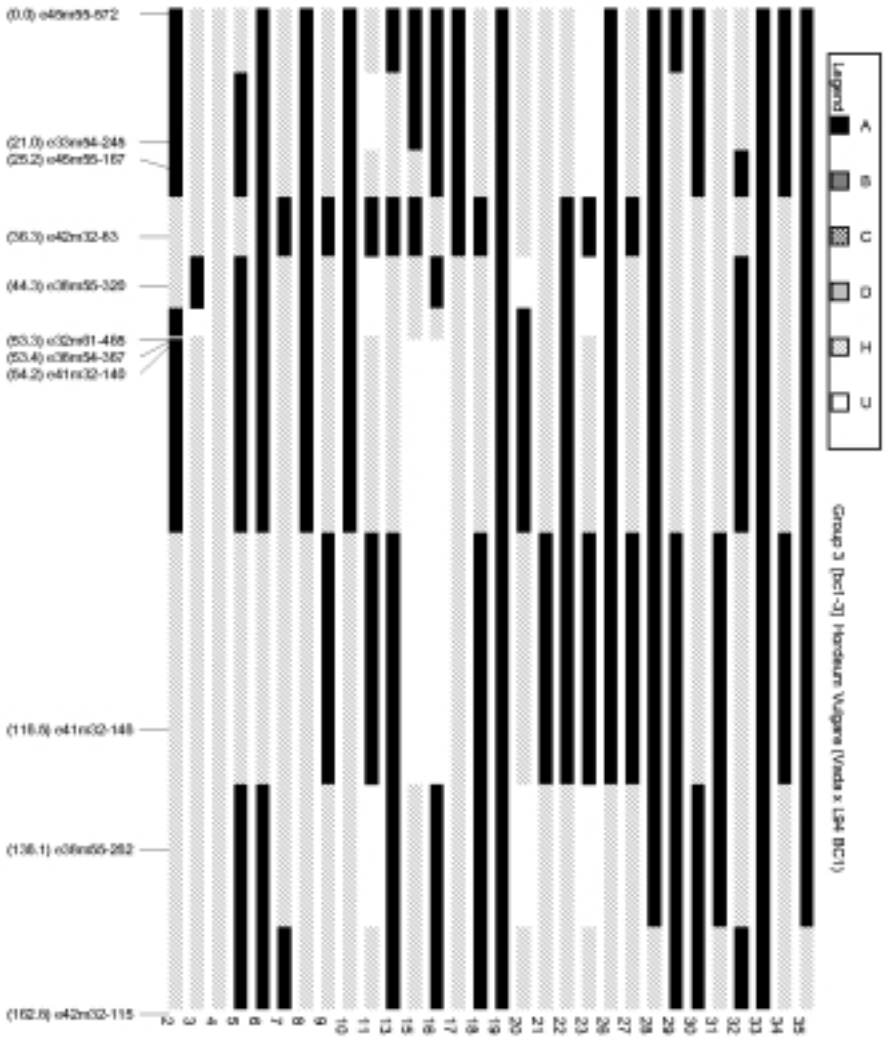



**Figure 1:** Example of a GGT drawing in 'Linkage Group' mode. Chromosome 3 is depicted for 30 barley backcross plants.

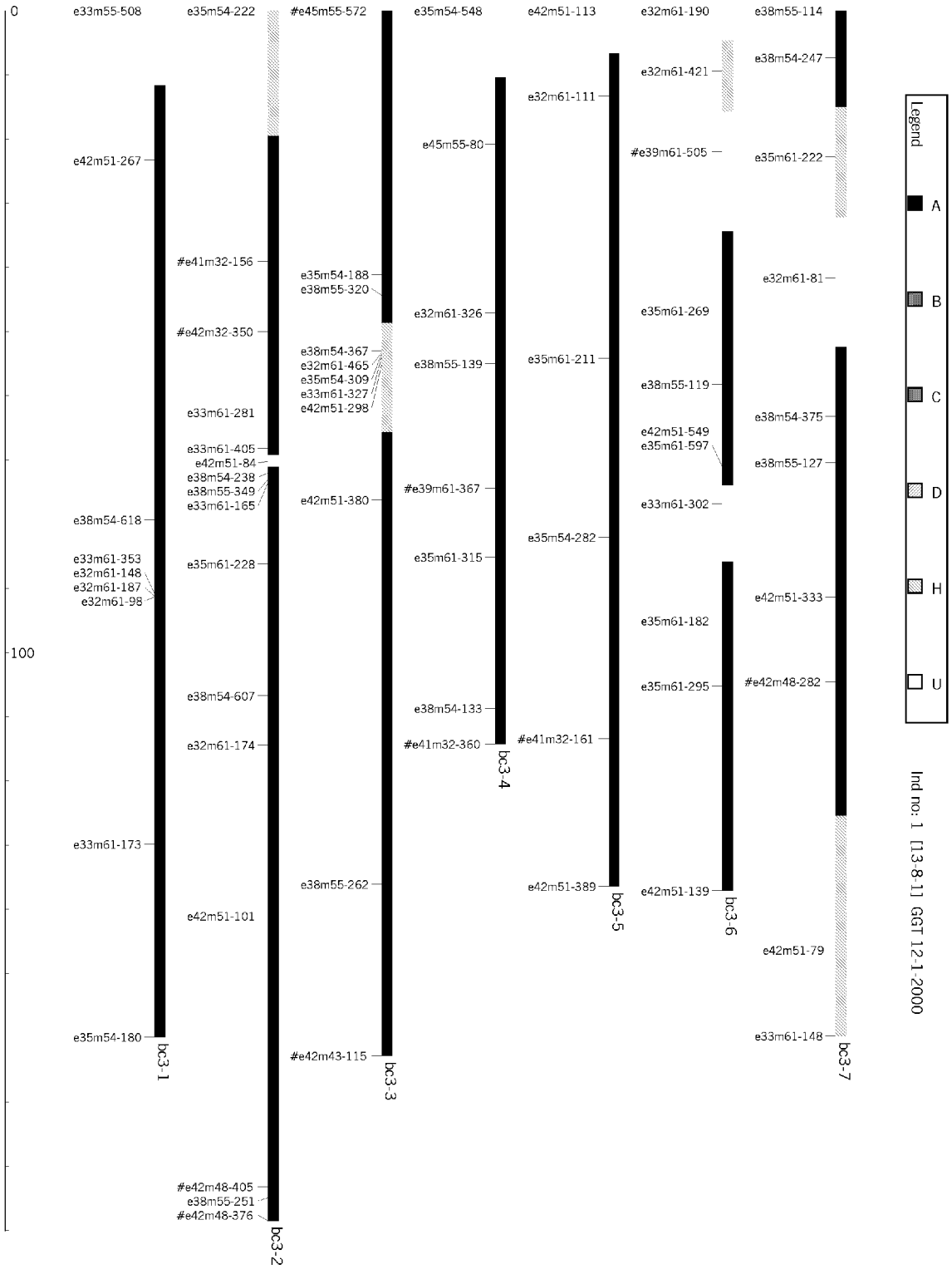



**Figure 2:** Example of a GGT drawing in 'individual view'l Graphical marker genotype image of the genome of a barley backcross plant

Molecular marker information is usually gathered and arranged per marker, as a list of alleles of individual plants. When markers are arranged in the correct map order this arrangement can be visualized, resulting in a display of the genomic arrangement of a single chromosome for all individuals. Such a drawing allows a quick inspection of all individuals, for instance to select those that have a favourable composition for the chromosome displayed. In GGT such a drawing is obtained when the option 'Organize by linkage group' is selected. Figure 1 shows an example of such a drawing.

The normal arrangement of marker data generally does not make it easy to get a complete picture of the whole genome of a single individual (plant or animal). GGT can provide schematic representations of the genome composition of all chromosomes for a single individual in a composite drawing. The option 'Organize by individual' results in such a drawing. This arrangement of the data is useful to browse through potentially valuable individuals, and verify the genome composition of individuals for regions other than the region of primary interest. An example of an image depicting the genomic composition of a barley inbred line is shown in Figure 2.

*Filtering & Selection*

When the location on the genetic map of genes or QTLs for traits of interest is known, it is possible to devise an ideal genotype which has, for all markers, the desired allele. Such an 'ideal genotype' or ideotype (e.g. Kearsey & Pooni, 1996) can be sought for by selecting within the population, using markers, for individuals that comply with the desired allelic composition. However, it may be difficult to obtain a true ideotype. A gradual

**Specified Selection Criteria:**
**[Selection performed in :]**
**listerset <Lister1.ggt>**
**listerset <Lister2.ggt>**
**listerset <Lister3.ggt>**
**listerset <Lister4.ggt>**
**listerset <Lister5.ggt>**

**[Criteria:]**

| | | | |
|---|---|---|---|
| **<GROUP 1>** | **[13.3]** | **w113** | **= A** |
| **<GROUP 1>** | **[16.6]** | **w203** | **= A** |
| **<GROUP 2>** | **[6.9]** | **g3843** | **= A** |
| **<GROUP 2>** | **[7.6]** | **w301** | **= A** |
| **<GROUP 3>** | **[42.5]** | **w148** | **= A** |
| **<GROUP 3>** | **[43.4]** | **w139** | **= A** |
| **<GROUP 3>** | **[45.7]** | **m216** | **= A** |
| **<GROUP 4>** | **[17.6]** | **g4564-b** | **= B** |
| **<GROUP 4>** | **[21.1]** | **m249** | **= B** |
| **<GROUP 5>** | **[34.7]** | **w83** | **= A** |
| **<GROUP 5>** | **[36.1]** | **w194** | **= A** |

**Selection Results:**
**2 individuals selected:**
**Nr. 7 [RIL-7]**
**Nr. 81 [RIL-81]**

**Figure 3**: example of GGT output when selection is applied. Here selection for an arbitrary set of marker alleles in the Lister & Dean set of *Arabidopsis* (ColxLer) RILs is practised.

reduction of the number of selected individuals, through a stepwise addition of selection terms, could lead to an acceptable, near-ideal genotype. The 'Marker Selection' option of GGT permits the specification of selection demands. For each single marker the desired allele can be indicated. In this way a stepwise increase of the number of selection criteria is possible. The selection demands, which can affect several chromosomes simultaneously, are verified, and the genotypes that comply with the specified criteria are gathered in a list and their genomic composition is displayed. Figure 3 gives an example of a selection that was performed in a set of 99 *Arabidopsis* RILs (ColxLer; Lister & Dean, 1993), and the selection results that were obtained.

*Statistics*

The combination of genetic map information and information on marker alleles permits an estimation of the genome composition. GGT provides figures for the proportions of the genome that are homozygously derived from one parent, homozygously derived from the other parent or genome that is heterozygous. Also the number of recombinations (i.e. change of colour in the drawing) and the number of heterozygous fragments (useful for evaluation of backcross progenies) are presented. These statistics are calculated on an individual basis, when a single genotype has focus, but overall statistics, arranged per individual or per marker, are also available. The data that is calculated can be printed, saved or exported to a Microsoft Excel spreadsheet for further analysis.

*Data input and output*

Locus data (using the Joinmap style of coding) arranged in plain text files serves as input for GGT, together with map-data, which can be derived from the locus data. GGT contains a module to 'merge' these data into the GGT data format. Graphical genotype images can be printed in high resolution, moved to the windows clipboard or saved to disk as graphical file.

*Extending the functionality of GGT*

GGT is not static software. As the field of genetics continues to evolve, so do the demands for software used in exploration of genetic data. Based on user feedback some targets for improvement of the functionality of GGT were identified. Future versions of

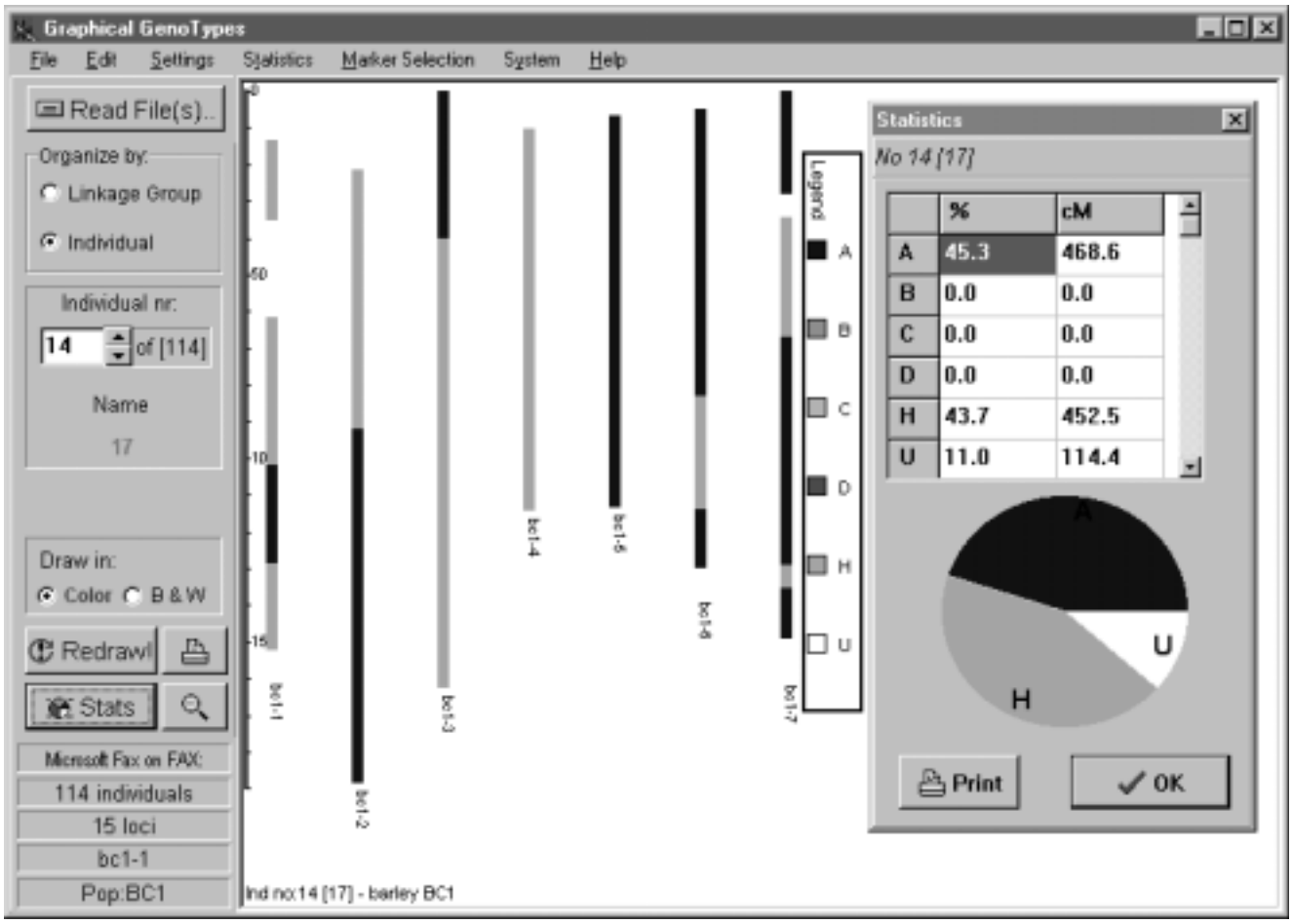



**Figure 4**: A screenshot of GGT featuring the pop-up statistics window. The genome of a barley $BC_1$ plant is drawn and the statistics on the genomic composition of this plant are summarised.

GGT will contain extended options for graphical data representation (similar to options available in QGene and Keygene software) and include the ability to deal with 'scenario study-like' questions. For instance, a user who is interested to see what would happen if markers at positions X and Y are swapped, could try this out and see. In this way errors in the genetic map, that result in an unexpected high number of singletons, could be more easily detected. Other improvements will be aimed at a better handling of cross pollinated data, containing more than two alleles per locus, and a more versatile selection option including boolean operators (AND, OR).

## Availability

GGT was developed as public domain software and a package, containing the program executable, a manual and sample data files, is available for download on the Internet, at the Laboratory of Plant Breeding. [URL: *http://www.spg.wau.nl/pv/PUB/ggt/*]

# 8 General Discussion

The possibilities that accompany indirect selection were already recognized in the 1920's by Sax (1923). Still, until recently, only incidentally reports of the use of indirect selection were published (discussed in Tanksley, 1993). Indirect selection via markers was proposed by Thoday (1961) as a valuable new method, but the lack of suitable markers hampered the application of this idea in practice for a long time. This has changed with the advent of molecular markers. Isozymes first provided tools for indirect selection, but were later replaced by DNA-based markers such as RFLPs, RAPDs, microsatellites and AFLP markers. Currently, the prospects for application of marker-derived information in plant breeding are good. Markers are now applied routinely to replace time-consuming or expensive tests. As technology continues to evolve, the availability of markers at low costs will become a reality for many crops. This opens up a range of new options to exploit information that is obtained from linkage between markers and genes or QTLs. It is essential that plant breeders make use of all tools at hand to provide the world with improved varieties (Visser, 1999).In this chapter the prospects and limitations that accompany some of these new options and tools will be discussed.

*Marker-assisted introgression and backcrossing.*

Current cultivated crop species are the result of a process of domestication and selection by man that started about 10.000 years ago (Zeven & De Wet, 1982). It is likely that many of the current cultivated crops, during one or more time periods, were represented by only a limited number of plants. Such 'genetic bottlenecks' in the past could still limit the genetic diversity of current cultivated species (Tanksley & McCouch, 1997). In contrast, undomesticated 'wild' species and landraces often harbour a large genetic diversity. However, a large proportion of this material is unadapted and is therefore unattractive for use in a breeding program. Yet desirable genes are often still present in these 'exotic' gene-pools, and markers could be a useful tool to detect such genes and to facilitate a controlled introduction of these genes, using conventional breeding methods, into the current cultivated material. Furthermore, history has shown that the aims and

focus of plant breeding are not always constant, and new priorities are set continuously. Valuable genes affecting traits that were neglected in the past may have been lost over generations of breeding. These valuable genes could still be present in unadapted material or in old landraces that are being maintained in gene banks. Marker-assisted backcrossing of genes responsible for such traits (e.g. flavour in fruits, scent in flowers, disease tolerance) is an efficient way to re-introduce these desired characteristics. Furthermore, since only conventional breeding methods are used, breeders don't have to fear lack of acceptance of 'enriched' varieties, as would be the case when gene cloning methods would be applied. Until recently, exotic genetic resources were mainly exploited to introduce monogenic traits into elite breeding material. Classic examples are monogenic resistances to airborne fungal diseases in bread wheat (*Triticum aestivum*) that derive from wild relatives or progenitors, and a variety of resistance genes in cultivated tomato, introgressed from a number of wild *Lycopersicon* species. The reason why breeders have been reluctant to resort to exotic germplasm for crop improvement with respect to complex, polygenic traits is the labourious and time-consuming process that is needed to achieve genetic improvement by phenotypic selection during a repeated backcross programme. For quantitative traits this would require the evaluation of large, segregating backcross generations in field trials. When dealing with crops, like small grains, where the performance of individual plants with respect to a quantitative trait is virtually impossible under normal growing conditions, introgression of desired quantitative traits would be even more labourious. The advent of molecular markers and the statistical tools for detecting linkage between 'quantitative' genes and markers has drastically changed this situation.

The results of simulated and experimental marker-assisted backcrossing, which were discussed in chapter six, confirm the findings of other authors (Tanksley & Nelson, 1996; Hospital & Charcosset, 1997; Bernacchi *et al.* 1998a,b; Hill, 1998), advocating the use of markers in a repeated backcross program for a fast reduction of the proportion of unwanted donor genome. Generally, two or three generations of controlled backcrossing should be sufficient to obtain a desired genotype for a single gene or QTL of interest. Intercrossing several genotypes that contain single introgressed genes could then result in a suitable genotype that is enriched for one or more specific traits.

*Marker-assisted selection and breeding*

Population improvement through the use of marker-assisted selection has been the subject of many analytical and simulation-based studies (Lande & Thompson, 1990; Zhang & Smith, 1992+1993; Gimelfarb & Lande 1994a,b; Whittaker *et al.* 1995; Luo *et al.* 1997; Moreau *et al.* 1998). Marker-assisted selection could also be a relevant tool for the selection of parents, used in crosses. In many breeding programs genetic variation is created by crossing genetically divergent parents (Schut, 1998). Especially parents that are complementary to each other at the genetic level are expected to yield a large variability among their offspring. Although a large diversity at the genetic level is not always clearly visible at the phenotypic (field) level, it can be revealed through molecular marker analysis. The problems that are involved in parent selection are not new. The selection of parents, used for crossing, is often based on the expectation of the quality and variability of the offspring. In practice, the quality of parents can be assessed by evaluating small scale test crosses (e.g. Van Oeveren, 1993). A different method to seek for complementary genotypes is to consider the genetic distance between potential parents. Parameters that give an insight in the genetic distance between lines can be obtained from pedigree information, from morphological observations, and from genetic markers (Schut, 1998). A cross between genotypes that are genetically separated by a large genetic distance is expected to display a highly diverse offspring, yielding valuable material for selection by the breeder. At the level of alleles of genes, genetically complementary parents can be sought for when information on the location and effect of genes is available. This was the subject of our studies on the relative efficiency of marker-assisted selection of parents with regard to the performance of their offspring. These studies, which were presented in chapters two, three and four, confirmed also for this type of selection the potential superiority of MAS. Trait heritability was identified as one of the most important factors affecting MAS efficiency. We found similar figures for the optimal heritability (ranging between 0.1 and 0.3) as were found by Lande & Thompson (1990) and Moreau *et al.* (1998) for marker-assisted improvement of populations. In the case QTLs are discovered that encode for quantitative, partial resistance, application of MAS opens up another possibility. Marker-assisted selection could be used for the pyramiding of resistance QTLs, even when the addition of another resistant allele to an already resistant genotype does not add much to the level of resistance. It is expected

that adding extra resistance QTLs does add to the durability of the resistance, since it becomes increasingly difficult for the pathogen to adapt to a genotype containing a range of resistance genes.

*Practical application of marker-assisted selection.*

Most simulation studies show good results for application of marker-assisted selection but, as indicated in the discussion section of chapters two and four and also recognized in other studies (Whittaker *et al.* 1995; Knapp, 1998), simplifications and assumptions favouring MAS are often made. The effects of some of these assumptions were explored in our studies and, in general, relaxation of these assumptions only results in a small decrease of the efficiency of MAS. However, this does not implicate that marker-assisted selection is always to be preferred over phenotypic selection. Many factors play a role in the decision which selection strategy to apply. For instance, for traits with a high heritability MAS may still outperform phenotypic selection, but the high costs of obtaining genetic fingerprints, necessary for performing MAS, may render the procedure cost-ineffective. When MAS is applied in a case with incomplete QTL information the efficiency may actually be worse than phenotypic selection, since some undetected factors remain unselected by MAS. Furthermore, long term objectives should be considered. In population improvement, the high efficiency that is observed when MAS is applied is seen mostly during the first generations of selection. It has been reported (Gimelfarb & Lande, 1994a; Hospital *et al.* 1997; Dekkers, 1999) that continued marker-assisted selection may yield lower selection efficiency in the long term, compared with conventional selection procedures. This is mainly seen when stringent MAS is applied in earl generations and 'minor QTLs', which remain undetected until all 'major QTLs' have become fixed in later generations, are lost. Another important parameter is population size (Gimelfarb & Lande, 1994a; Moreau *et al.* 1998; Chapter 2). When larger populations are used it may be expected that MAS will be able to extract, in a more efficient way than phenotypic selection, the superior genotypes or parents from this population. Also, large mapping populations allow a more reliable detection of QTLs. However, in most cases practical and economic considerations limit the population sizes that can be used. In a situation where budgets are fixed and the costs of genotyping plants in order to be able to perform MAS come at the expense of fewer plants that can be grown (i.e. a smaller

population size), it remains to be seen if MAS will end up as the superior selection strategy.

*Successes in practical application of marker-assisted selection*

Nevertheless, already many success stories on the use of MAS in practical plant breeding have been reported. In a number of papers Stuber described successes in the application of MAS in corn breeding (Stuber & Sisco, 1992; Stuber, 1994). Tanksley and others were successful in the identification and transfer of valuable genes, derived from wild relatives, into cultivated tomato (Tanksley & Nelson, 1996; Bernacchi *et al.* 1998a,b). Huang described the pyramiding of resistance QTLs in rice lines (Huang *et al.* 1997). The experiments described in this thesis showed mixed results. For a simple case, but studying a relative high heritability trait, we demonstrated that selection that was purely based on marker information was just as effective as phenotypic selection (chapter 3). In a more complex case however, selection results were unable to confirm the expected superiority of MAS (chapter 5). Most simulation studies did not consider economic cost-efficiency, although Knapp (1998) argued that application of MAS may well be an economically sensible exercise in many cases. The difficulty in assessing matters related with costs are the time dependencies of many factors. New equipment, protocols etc., which can reduce the cost of obtaining marker data dramatically, are emerging at a high rate. Studies like the one discussed in this thesis can merely provide a rough estimate on the amount of expected gain in selection efficiency and on the amount of required 'data-points'. The breeder remains the key person to decide if it is worthwhile to pursue such an exercise, for his own crop and conditions, and with the current available information on the 'price per marker data-point'.

*Tools for analysis*

The increasing supply of large molecular data sets demand the availability of a robust set of tools for analysis. The capacities of modern computers (the hardware) seem to keep up with the growing supply of data; therefore, the real demand is for intelligent software that is able to use the available data to provide answers to scientific and applied questions. The theoretical work on the principles of QTL mapping has now achieved a solid background (Lander & Botstein, 1989; Haley & Knott,1992; Van Ooijen, 1992; Jansen &

Stam, 1994; Jansen, 1995; Doerge & Rebaï, 1996; and others) and is still subject of further study and improvement. Implementations of the developed methods, in a diversity of 'flavours' are now available (Lander *et al.* 1987; Basten *et al.* 1994; Holloway & Knapp, 1994; Tinker & Mather, 1995; Van Ooijen & Maliepaard, 1996a,b; Nelson, 1997). Many of the currently available software packages were created by enthusiastic scientists, sometimes on an ad-hoc basis directed at solving an emerging problem. The available software packages each incorporate different (sets of) solutions to tackle genetic problems (see Li, 1999 for an extensive overview of available genetic software). Most scientific programmers have given emphasis to a sound methodology and functionality of their software, but paid less attention to the user-friendliness and standardisation of data used for in- and output. Furthermore, since many of these software packages were created 'pro-deo' and are freely available, user support is rarely provided and maintenance is irregular or absent. This diversity has not made the practical use of QTL mapping very accessible to the community of scientists, working in related fields, and plant breeders.

The high speed at which developments in marker and computer technology continue to advance induce a need for standardisation and a more automated processing and analysis of molecular marker data. The large size and multidimensional character of marker data sets invite novel approaches to data visualization (Nelson, 1997). User friendly 'smart' software packages are therefore a prerequisite for practical use of marker derived information on a large scale. Although there are efforts to provide users with software that is easier to use (e.g. Korol *et al.* 1999; Van Berloo, 1999b), it would be a good idea if professional software developers were to be involved in the development and introduction of standardised, robust and user-friendly software. Some efforts in this direction can currently be seen (Van Ooijen, pers.comm.). A wide acceptation of such a suite of programs would not only keep the software affordable, but would also permit easier transfer, sharing and combining of data, which could help to increase experimental resolution (Beavis, 1999).

*Interaction, Correlation, non-mapping populations*

Studies on genetic improvement mainly focus on main effects. Interaction factors such as genotype by environment (GxE and QTLxE) interaction or interacting genes (epistasis) are difficult to handle and unpredictable. Improved algorithms that enable detection of epistasis provide new options to steer selection in these cases. Selection decisions could take advantage of knowledge on interacting genes and, through the use of marker assisted selection, favourable sets of genes could be assembled or undesired combinations of alleles prevented. Recently, studies on simultaneous detection of QTLs for multiple traits were described (Hackett, *in prep*.; Korol *et al.* 1998; Ronin *et al*. 1999). Such an approach may result in an increased power of QTL-detection. Multiple-trait QTL-detection might become a natural partner of multiple-trait marker-assisted selection procedures, of which an example was described in chapter four of this thesis. Of course, new problems arise when these kinds of procedures are to be applied. Not all traits can be measured with the same accuracy, hence data of unequal quality is used for QTL detection. The reliability of the data should in some way be reflected in the QTL-analysis. When data from a variety of sources is used, the inclusion of the experimental design into QTL detection methods could become important.

Several authors reported conservation of QTL locations over populations that were derived from different progenitors. This indicates that the usefulness of a detected QTL may transcend the population in which it was found. More general QTL detection methods, which are not limited to the use of mapping populations and assumptions on the normality of trait distributions, could enhance the applicability of QTL-based selection strategies. Such detection methods are already being developed for analysis of human and animal populations, and could also be employed fruitfully in plant populations that are derived from a diversity of sources, as is common practice in plant breeding.

In the introduction of this thesis another major advance in genetics was briefly mentioned; the enormous efforts directed at the retrieval of the complete DNA sequence of important plant and animal species. Sequences that seem to consist of functional genes, but which function is still unknown, can be mapped onto a genetic linkage map.

Comparison of the map position of functional sequences with detected QTLs could provide information on the metabolic function of genes located at QTLs, which would increase the understanding of genetics and genetic regulation and provide new options for selection and controlled genetic improvement.

The search for genes that interact with other genes, the search for genes that interact with the environment and the detection of QTLs through the analysis of correlated traits all require extensive calculations, due to the large number of combinations that need to be evaluated. Nevertheless, I expect that these options, together with the introduction of desired alleles from related gene-pools will receive the most attention in the time that lies ahead.

# References

**Abbott RJ, Gomes MF** (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* 62:411-418

**Agrevo**, 1996. Global Food Forecast. (www.agrevo.com)

**Alonso-Blanco C, El-Assal S, Peeters AJM, Koornneef M** (1997) Genetic analysis of developmental traits based on the Ler/Cvi natural variation. 8th International Conference on *Arabidopsis* research, Madison, Wisconsin 12-1

**Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J and Kuiper MTR** (1998a) Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotpes and construction of a L*er*/Cvi recombinant inbred line population. *The plant journal* 14:259-271

**Alonso-Blanco C, El-Din El-Assal S, Coupland G and Koornneef M** (1998b) Analysis of natural allelic variation at flowering time in the landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* 149:749-764

**Alemayehu F and Parlevliet JE** (1996) Variation for resistance to Puccinia hordei in Ethiopian barley landraces. *Euphytica* 90: 365-370

**Basten CJ, Weir BC and Zeng ZB** (1994) Zmap-a QTL cartographer. *In* Smith, Gavora, Chesnais, Fairfull, Gibson, Kennedy, and Burnside (Eds.),*Proc 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software*, Vol 22, Guelph, Ontario, Canada, pp. 65-66.

**Beavis WD** (1999) Integrating the central dogma with plant breeding. *Abstracts Plant & Animal Genome Conference VII*: Scherago Intl. Inc. New York p.28

**Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D and Tanksley S** (1998a) Advanced backcross QTL analysis in tomato I. Identification of QTLs for traits of agronomic importance from *Lycopericum hirsutum*. *Theor Appl Genet* 97:381-397

**Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Samaya H, Uhlig J, Zamir D and 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 *Lycopericum hirsutum* and *L. Pimpinellifolium*. *Theor Appl Genet* 97:170-180

**Bindraban PS** (1997) Bridging the gap between plant physiology and breeding; Identifying traits to increase wheat yield potential using systems approaches. *Thesis Wageningen Agricultural University* (145pp). Wageningen, The Netherlands.

**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

**Dang JL, Dietrich RA and Richberg MH** (1996) Death don't have mercy: cell death programs in plant-microbe interactions. *Plant Cell* 8:1793-1807

**Dekkers JM** (1999) Optimizing strategies for marker-assisted selection. *Abstracts Plant & Animal Genome Conference VII*. Scherago Intl. Inc. New York. p.29

**Doerge RW and Rebaï A** (1996) Significance thresholds for QTL interval mapping tests. *Heredity* 76:459-464

**Dupuis J and Siegmund D** (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 151:373-386

**Edwards MD and Page NJ** (1994) Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* 88:376-382

**Falconer DS and Mackay TFC** (1996) Introduction to quantitative genetics. 4th ed. 464 pp. Longman Group, Essex.

**FAO** (1996): Report of the World Food Summit, Rome, November 1996; FAO Technical Documents. FAO Rome. (www.fao.org)

**Gallais A and Charcosset A** (1994) Efficiency of marker-assisted selection. *In* Biometrics in plant breeding: Applications of molecular markers. *Proc ninth meeting Eucarpia section biometrics:* 91-98. Jansen J & Van Ooijen JW ed. CPRO-DLO, Wageningen, The Netherlands.

**Gilmour AR, Thompson R, Cullis BR** (1995) Average Information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51:1440-1450

**Gimelfarb A and Lande R** (1994a) Simulation of marker assisted selection in hybrid populations. *Genet Res Cambr* 63:39-47

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

**Gimelfarb A and Lande R** (1995) Marker-assisted selection and marker-QTL associations in hybrid populations. *Theor Appl Genet* 91:522-528.

**Haley CS and Knott SA** (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315-324

**Henshall JM and Goddard ME** (1999) Multiple-Trait Mapping of Quantitative Trait loci after selective genotyping using logistic regression. *Genetics* 151:885-894

**Hill WG** (1998) Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. *Genetics* 148:1341-1352

**Hoeschele I, Uimari P, Grignola FE, Zhang Q and Gage KM** (1997) Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics 147:1445-1457**.**

**Holloway JL and Knapp SJ** (1994) G-Mendel Users guide. Oregon State University; Corvallis, Oregon, USA

**Hospital F, Chevalet C and Mulsant P** (1992) Using Markers in gene introgression breeding programs. *Genetics* 132:1199-1210

**Hospital F and Charcosset A** (1997) Marker-Assisted Introgression of Quantitative Trait Loci. *Genetics* 147:1469-1485

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

**Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennet J and Khush GS** (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor Appl Genet* 95:313-320

**Jansen RC** (1995) Genetic mapping of quantitative trait loci in plants - a novel statistical approach. *Thesis Wageningen Agricultural University* (109pp). Wageningen, The Netherlands.

**Jansen RC and Stam P** (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447-1455

**Jansen RC, Ooijen JW van, Stam P, Lister C, Dean C** (1995) Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. *Theor Appl Genet* 91:33-37

**Kearsey MJ and Pooni HS** (1996) The genetical analysis of quantitative traits. Chapman & Hall, London. 381 pp.

**Keygene** (1999) Molecular marker services - plant breeding applications (http://www.keygene-mms.com/html/plant_breeding.htm)

**Kiraly Z and Barna B** (1985) Hypersensitivity revisited. *Acta Phytopathologica Academiae Scientarum Hungaricae* 20:1-6

**Knapp SJ** (1994) Selection using molecular marker indexes. *In Proc second plant breeding symposium of the Crop Sci.* Soc. Amer. And Amer. Soc. Hort. Sci., Corvallis, Oregon. pp 1-11. Amer. Soc. Hort. Sci. Alexandria, Virginia, USA.

**Knapp SJ** (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Sci* 38:1164-1174

**Koornneef M** (1982) The genetics of some planthormones and photoreceptors in *Arabidopsis thaliana* (L) Heynh. *Thesis Landbouwhogeschool Wageningen* (157pp). Pudoc, Wageningen, The Netherlands.

**Korol AB, Ronin YI, Nevo E and Hayes PM** (1998) Multi-interval mapping of correlated trait complexes. *Heredity* 80:273-284

**Kuittinen H, Sillanpää MJ, Savolainen O** (1997) Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. *Theor Appl Genet* 95:573-583

**Lande R and Thompson R** (1990) Efficiency of Marker Assisted Selection in the improvement of quantitative traits. *Genetics* 124:743-756

**Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE and Newburg L** (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181

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

**Lee M** (1995) DNA markers in plant breeding programs. *Adv Agron* 55: 265-344

**Levert C** (1959) Een nomogram voor betrouwbaarheidsgrenzen van quantielen en overschrijdingskansen bij normaal verdeelde grootheden. *Statistica Neerlandica* 13:3-14

**Li W** (1999) An alphabetic list of genetic analysis software. (http://linkage.rockefeller.edu /soft/list.html)

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

**Lister C and Dean C** (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *The Plant Journal* 4:745-750

**Luo ZW, Thompson R and Woolliams J** (1997) A population genetic model of marker-assisted selection. *Genetics* 146:1173-1183

**Malthus TR** (1798) An essay on the principle of population. St Pauls church yard, London.

**Meyerowitz EM and Pruitt RE** (1985) *Arabidopsis thalian* and plant molecular genetics. *Science* 229:1214-1218

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

**Mullis KB and Faloona FA** (1987) Specific synthesis of DNA in vitro via the polymerase catalysed reaction. *Meth Enzymol* 255:335-350

**NASC** (1998) Nottingham Arabidopsis stock centre (http://nasc.life.nott.ac.uk)

**Nelson JC** (1997) QGENE: software for marker-based genomic analysis and breeding. *Molecular Breeding* 3:239-245

**Parlevliet JE** (1983) Race specific resistance and cultivar specific virulence in the barley leaf rust pathosystem and their consequences for the breeding of leaf rust resistant barley. *Euphytica* 32: 367-375

**Qi X** (1998) Identification and mapping of genes for partial resistance to Puccinia hordei Otth in barley. *Thesis Wageningen Agricultural University* (165pp). Wageningen, The Netherlands.

**Qi X and Lindhout P** (1997) Development of AFLP markers in barley. *Mol Gen Genet* 254:330-336

**Qi X, Stam P and Lindhout P** (1998a) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. *Theor Appl Genet* 96:376-384

**Qi X, Niks R, Stam P and Lindhout P** (1998b) Identification of QTLs for partial resistance to leaf rust (Puccinia hordei) in barley. *Theor Appl Genet* 96:1205-1215

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

**Ronin YI, Korol AB and Nevo E** (1999) Single- and Multiple-trait mapping analysis of linked quantitative trait loci: some asymptotic analytical approximations. *Genetics* 151:387-396

**Sax K** (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 75:709-726

**Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson H, Bouchez D and Dean C** (1995) Physical map and organisation of *Arabidopsis thaliana* chromosome 4. *Science* 270:480-483

**Schut JW** (1998) Prediction of cross performance in barley. *Thesis Wageningen Agricultural University*. Wageningen, The Netherlands.

**Smith HO and Wilcox KW** (1970) A restriction enzyme from *Hemophilus influenzae*. 1. Purification and general properties. *J.Mol. Biol.* 51:379-392

**Soller M and Brody T** (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor Appl Genet* 47:35-39

**Stam P** (1993) Construction of integrated genetic linkage maps by means of a new computer package: Joinmap. *The Plant Journal* 3:739-744

**Stam P** (1994) Marker-assisted breeding. *In Biometrics in plant breeding: Applications of molecular markers. Proc ninth meeting Eucarpia section biometrics*: 32-44. Jansen J & Van Ooijen JW ed. CPRO-DLO, Wageningen, The Netherlands.

**Stam P** (1995) Weloverwogen keuzes. Inaugural speech Wageningen Agricultural University (Dutch) 22pp.

**Stam P and Van Ooijen JW** (1995) Joinmap (tm) version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen, The Netherlands.

**Stam P and Zeven AC** (1981) The theoretical proportion of the donor genome in near isogenic lines of self-fertilisers bred by backcrossing. *Euphytica* 30 :227-238

**Stuber CW** (1994) Successes in the use of molecular markers for yield enhancement in corn. *Proc 49th annual corn and sorghum industry research conference* : 232-238. Am seed trade association, Washington DC.

**Stuber CW and Sisco P** (1992) Marker facilitated transfer of QTL alleles between elite inbred lines and responses in hybrids. *Proc 46$^{th}$ annual corn and sorghum research conference*: 104-113. Am seed trade association, Washington

**Tanksley SD** (1993) Mapping polygenes. *Annu Rev Genet* 27:205-233

**Tanksley SD and McCouch SR** (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:1063-1066

**Tanksley SD and 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

**Thoday JM** (1961) Location of polygenes. *Nature* 191:368-370

**Tinker NA and Mather DE** (1995) MQTL: software for simplified composite interval mapping of QTL in multiple environments. *Journal of Agricultural Genomic* Vol1:2 (http://www.ncgr.org/ag/jag/papers95/paper295/jqtl16r2.html)

**Van Berloo R** (1999a) GGT: Software for the Display of Graphical Genotypes. The *Journal of Heredity* 90: 328-329

**Van Berloo R** (1999b) MapQTL Assistant: A user-friendly shell round MapQTL®. *Journal of Agricultural Genomics* Vol4:2 (www.ncgr.org/ag/jag/papers99/paper199/indexp199.html)

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

**Van Berloo R and Stam P** (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 Berloo R and Stam P**. Simultaneous marker-assisted selection for multiple traits in autogamous crops. *Submitted for publication* (Chapter 4)

**Van Berloo R, Aalbers H and Niks RE.** The construction of NILs in barley using marker-assisted backcrossing: experimental and simulation results. *Submitted for publication* (Chapter 6)

**Van der Schaar W, Alonso-Blanco C, Léon-Kloosterziel KM, Jansen RC, Van Ooijen JW and Koornneef M** (1997) QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. *Heredity* 79:190-200

**Van Oeveren AJ** 1993 Efficiency of single seed descent and early selection in the breeding of self-fertilizing crops. *Thesis Wageningen Agricultural University*. Wageningen, The Netherlands.

**Van Ooijen JW** (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theor Appl Genet* 84:803-811

**Van Ooijen JW and Maliepaard C** (1996a) MapQTL ™ version 3.0: Software for the calculation of QTL positions on genetic maps. *Plant Genome IV Conference abstracts* p.316

**Van Ooijen JW and Maliepaard C** (1996b) MapQTL ™ version 3.0: Software for the calculation of QTL positions on genetic maps. CPRO-DLO, Wageningen.

**Visser, RGF** (1999) Plantenveredeling: Kiezen, verbeteren en weggooien. Inaugural speech Wageningen University (in dutch) 30pp.

**Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M** (1995) AFLP: a new technique for DNA fingerprinting. *Nucl Acids Res* 23:4407-4414

**Whittaker JC, Curnow RN, Haley CS and Thompson R** (1995) Using marker-maps in marker-assisted selection. *Genet Res Camb* 66:255-265

**Welsh J, McClelland M** (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucl Acids Res* 18:7213-7218

**WRR** (1995) Sustained risks: a lasting phenomenon. Netherlands Scientific Council for Government Policy. Reports to the Government 44. SDU Publishers, The Hague. 205pp.

**Young ND and Tanksley SD** (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor Appl Genet* 77:95-101

**Zeven AC and De Wet JMJ** (1982) Dictionary of cultivated plants and their regions of diversity. Pudoc Wageningen, the Netherlands

**Zhang W and Smith C** (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor Appl Genet* 83:813-820

**Zhang W and Smith C** (1993) Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. *Theor Appl Genet* 86:492-296

**Zuurbier PJP** (1994) Diagnosing the seed industry. *Prophyta* 4:10-27

# Summary

Molecular markers provide plant breeding with an important and valuable new source of information. Linkage between molecular markers can be translated to genetic linkage maps, which have become an important tool in plant and animal genetics. Linkage between (quantitative) trait-data and occurrences of marker alleles allow identification of important genetic factors, underlying observable traits. Knowledge that results from such analyses, i.e. the location on the genome of important genetic factors (quantitative trait loci or QTLs), can and should be applied when making selection and breeding decisions.

Selection of parents is an important issue in plant breeding. Basing selection on QTL information, i.e. applying marker-assisted selection, can result in an increased selection efficiency. This is especially true for quantitative traits with a low heritability. For efficient application of marker-assisted selection reliable and fairly complete QTL-mapping results are required. When QTLs were mapped for several traits a multiple trait-selection can be devised, through the use of a suitable index. In this case an ideal target genotype, containing favourable alleles for QTLs that affect the traits of interest, can be constructed and crosses can be made between selected parents in such a way that the probability of obtaining the target genotype is maximised. Although this approach looks promising, and simulation results show an improved selection performance, several problems remain which are limiting application in practise. A more reliable and complete mapping of QTLs, including mapping of interaction between QTLs, mapping of QTLs with a higher reliability, for instance resulting from a combined mapping of several traits, and mapping of QTLs in more diverse non-mapping types of populations could greatly contribute to an increased application of marker-assisted selection, and hence a more efficient selection in plant breeding.

Although it is common practise to resort to unadapted material when searching for new genetic variation, the undesired characteristics that accompany the genes coding for the target trait of interest, limit the applicability of introducing 'foreign' genes. With the help of marker and QTL-analysis the genome region that harbours genes which are responsible for the desired characteristics can be identified more precisely and thus the size of the

fragment that needs to be introgressed can remain restricted. Marker-assisted backcrossing allows a much more controlled method of gene introgression, limiting the amount of 'linkage-drag' and requiring less generations of backcrossing than conventional backcrossing for yielding suitable genotypes.

Developments that favour application of marker-assisted selection are still progressing at a high rate. New technical enhancements in the field of molecular biology, new protocols and methods for identification of genetic factors, new versatile software for data analysis and visualisation all contribute to new ways of selection and breeding that take advantage of this newly acquired knowledge and information. These novel methods should be used to continue to create genetic improvement, in a faster or more efficient way than before, and to introduce quality enhancing genetic factors into cultivated crops.

# Samenvatting

Moleculaire merkers leveren de plantenveredeling een belangrijke en waardevolle nieuwe bron van informatie. Koppeling tussen moleculaire merkers kan worden vertaald in genetische koppelingskaarten, die een belangrijk hulpmiddel vormen in de planten- en dieren genetica. Associaties tussen (kwantitatieve) data van eigenschappen en het al dan niet voorkomen van merker-allelen maken het mogelijk belangrijke genetische factoren te herkennen die ten grondslag liggen aan waarneembare eigenschappen. De kennis die uit deze analyse voortvloeit, de positie die belangrijke genetische factoren (quantitative trait loci afgekort QTLs) op het genoom innemen kan en moet gebruikt worden bij het nemen van veredelings en selectie beslissingen.

De keuze van geschikte ouders is een belangrijk onderwerp in de plantenveredeling. Het baseren van deze selectie op informatie omtrent QTLs, d.w.z. het toepassen van merker gestuurde veredeling kan resulteren in een verhoogd selectieresultaat. Dit geldt vooral voor kwantitatieve eigenschappen die vererven met een lage erfelijkheidsgraad. Efficiënte toepassing van merker gestuurde selectie vereist betrouwbare en complete QTL-detectie resultaten. Als er QTLs gedetecteerd zijn voor meerdere eigenschappen kan, door gebruik te maken van een geschikte index, gelijktijdig worden geselecteerd voor meerdere eigenschappen. In dit geval kan vooraf een ideaal genotype worden geconstrueerd dat voor alle QTLs die de eigenschap beïnvloeden de gewenste allelen bevat. Er kunnen kruisigen worden gemaakt waarbij ervppr wordt gezorgd dat de kans op het verkrijgen van het ideale genotype maximaal is. Hoewel deze aanpak veelbelovend lijkt, en simulatie resultaten wijzen op een verhoogd selectie resultaat, zijn er nog verschillende problemen die de praktische toepasbaarheid van deze methode belemmeren. Een betrouwbaardere en meer complete QTL detectie, met inbegrip van de detectie van QTL-interactie zou kunnen bijdragen aan een grotere toepasbaarheid van merker gestuurde selectie. Ook het simultaan karteren van QTLs voor meerdere eigenschappen en het karteren van QTLs in andere dan standaard splitsende populaties kan hieraan bijdragen en op die manier een efficientere selectie in de plantenveredeling mogelijk maken.

Hoewel het gangbaar is om gebruik te maken van onaangepast, wild materiaal, wanneer men op zoek is naar nieuwe genetische variatie. Helaas zorgen de ongewenste eigenschappen die veelal samengaan met genen die coderen voor de doeleigenschap voor een beperking in de toepasbaarheid van de introductie van 'vreemde' genen. Merker- e QTL-analyse kunnen gebruikt worden om de regio's op het genoom die genen dragen die verantwoordelijk zijn voor de gewenste eigenschappen nauwkeuriger te identificeren. Hierdoor kan de grootte van het in te brengen genoomfragment beperkt blijven. Merker gestuurde terugkruising biedt de mogelijkheid voor een gecontroleerde methode van gen introgressie, waarbij de hoeveelheid 'linkage-drag' kan worden beperkt en minder generaties met terugkruisen nodig zijn voordat een acceptabel genotype bereikt is dan bij een conventionele terugkruisings procedure.

De progressie van ontwikkelingen die bijdragen aan de toepassing van merker gestuurde selectie is nog steeds groot. Nieuwe technische verbeteringen in de moleculaire biologie, nieuwe protocollen en methoden om genetische factoren te herkennen, nieuwe bree toepasbare software voor data analyse en visualisatie dragen allemaal bij aan nieuwe methoden voor selectie en veredeling die van deze nieuwe ontwikkelingen gebruik maken. Deze nieuwe methoden moeten worden aangewend om genetische verbetering te blijven boeken, sneller en efficiënter dan voorheen, en om gewassen te voorzien van kwaliteit verbeterende factoren.

# Nawoord

Het nawoord is bij uitstek de plaats om de lezer een kijkje in de keuken te geven van het ontstaan en verloop van een AIO onderzoeksproject. Tevens is het een goede gelegenheid om de indruk weg te nemen, die wellicht na lezing van het voorgaande is ontstaan, dat zulk onderzoek vooral een solitaire bezigheid is.
Al tijdens het hoorcollege *kwantitatieve genetica*, bij de behandeling van het onderwerp RFLP merkers, krabbelde ik in de kantlijn "afstudeervak?". Dat is er toen niet van gekomen, maar na mijn afstuderen kwam de kans alsnog, in de vorm van een AIO-positie bij de Vakgroep Plantenverdeling. Het is altijd geweldig mooi als je van je hobby je werk kunt maken en ik heb het gevoel dat me dat in de afgelopen periode, die bij elkaar een kleine 5 jaar heeft geduurd, aardig is gelukt. Een periode van onderzoek, waarin een mix van moleculaire technieken, selectiemethoden, statistiek, schrijven van computer-programma's, dataverwerking en het uitvoeren van kas-experimenten de voornaamste ingrediënten waren, heeft uiteindelijk geleid tot dit boekwerkje. Het vergaren van de gegevens die uiteindelijk hier op papier terecht gekomen zijn was niet gelukt zonder de medewerking van velen, die ik daarvoor graag wil bedanken.

Allereerst wil ik mijn begeleider en promotor, Piet Stam, bedanken. Piet, bedankt dat je mij goed genoeg vond om jouw "bruidsschat AIO" te worden. Bedankt ook voor de vrijheid waarmee je me met de onderzoeksmaterie hebt laten stoeien, je prikkelende opmerkingen en suggesties die af en toe inspireerden tot het schrijven van bruikbare computerprogramma's, en de goede coaching, vooral op het einde van het onderzoeks-traject. Ik hoop onze samenwerking nog even te kunnen voortzetten en zal vanuit mijn nieuwe functie zeker nog regelmatig een beroep doen op je expertise in het karteren van genen.
Vervolgens wil ik graag de overige leden van de werkgroep selectiemethoden bedanken voor hun bijdragen aan het denkproces rondom mijn onderzoek, ik wil daarbij tevens betrekken de leden van het Centrum voor Biometrie van het CPRO, waarmee we gedurende enkele jaren zinvolle en inspirerende discussies hebben gevoerd. Ies, bedankt voor de pannenkoeken en je opmerkzaamheid op het sluipend foutief gebruik van genetische termen. Johan Dourleijn, bedankt voor je advies op het gebied van een

verantwoorde proefopzet en voor de openhartige gesprekken die we recentelijk nog voerden. Johan Schut, bedankt voor de hulp bij het vinden van de weg binnen de onderzoeksschool PE en voor je creatieve ideeën die hebben bijgedragen aan het vormgeven van mijn onderzoek. Scott, behalve op het werk was je ook in de klimhal vaak een onmisbare steun. Arnold, ik vind het bijzonder leuk dat je mijn paranimf wilt zijn, en ik hoop dat we onze regelmatige gesprekken over genetische zaken, het verbeteren van 'onze' internetsite en vele, vele andere zaken nog even kunnen voortzetten. Johan van Ooijen, bedankt voor je programmeeradviezen en voor je bereidheid bij mijn project betrokken te blijven, Ritsert, bedankt voor de *Arabidopsis*-data. Hans Jansen, Jaap, Fred, Tytti, Leo, Chris, Hans van Os, Baboucarr, and all others that contributed to the selection methods meetings, thanks for your contribution.

Een deel van het onderzoek betrof experimenten met *Arabidopsis*, ofwel de zandraket. Zonder de ruimhartige hulp en inzet van Corrie Hanhart, van het Laboratorium voor Erfelijkheidsleer was het een stuk moeilijker geweest om deze experimenten uit te voeren. Ik wil daarom Corrie van harte bedanken voor haar inzet, tevens wil ik in dit verband Hetty en Maarten bedanken voor hun waardevolle hulp en advies. Quería agradecer a Carlos Alonso-Blanco por haber compartido sus datos Arabidopsis y su experiencia. ¡Muchas Gracias!

Veel werk is verzet door studenten die in het kader van een afstudeervak een deel van het onderzoek hebben uitgevoerd. Hans, Hanneke en Beyene, ik bewaar goede herinneringen aan onze samenwerking, bedankt voor jullie inzet.
De verzorging van de kas-experimenten was in de goede handen van UNIFARM, speciale dank aan Gerrit en Anton voor de goede verzorging van mijn *Arabidopsis*- en gerst-planten.
Voor computerproblemen was ik de afgelopen tijd een veel geraadpleegde vraagbaak, toch leverden hard- en software voor mij ook wel eens problemen op. Gelukkig waren er dan altijd de mensen van TUPEA systeembeheer om mij uit de brand te helpen. Hans, Jo, Evert en Stephan bedankt voor de soepele samenwerking die regelmatig de gebruikelijke service ver te buiten ging. Mijn zwager, Arjan Lamaker, heeft mij geïnfecteerd met het "Delphi virus". Bedankt daarvoor en voor de vele programmeeradviezen.

Op het laatst moest het er dan toch van komen, voor het gerstonderzoek moest ik zelf het lab in om AFLP data te verzamelen. Omdat mijn laatste ervaring met laboratoriumwerk toch al weer enkele jaren geleden was vergde dit wel wat uitleg en begeleiding. Petra, Fien en iedereen op het marker-lab, bedankt voor jullie geduld. Binnenkort zal ik weer een beroep op jullie doen!
A special word of gratitude for my friend and colleague Xiaoquan Qi, thanks for your patient help in the barley experiments, for sharing your data and plant material and for co-supervising some of my graduate students. Silvan wil ik bedanken voor de gezellige en frequente computerbabbeltjes en experimentjes.
Jair, zoals je laatst zelf al opmerkte, we hebben het ruim 4 jaar op dezelfde kamer uitgehouden. Je was steeds een gezellige vriend en kamergenoot met een openhartigheid die ik altijd zeer heb gewaardeerd. Ons wederzijds vertrouwen blijkt wel uit het feit dat je de finale handelingen rond je proefschrift aan mij toevertrouwde. Bedankt voor dit vertrouwen en voor je keuze om mij als paranimf te vragen. Ik vind het bijzonder leuk dat onze samenwerking, middels het schrijven van een gezamenlijke publicatie, nog een vervolg heeft gekregen.

Op zijn tijd was er ook de nodige ruimte voor verbreding, bijvoorbeeld tijdens workshops, congressen en cursussen, maar ook de onderzoeksschool Productie Ecologie (PE) bood hiertoe ruime mogelijkheden middels het studentenplatform, de onderwijscommisie en het methodologie discussiegroepje. Ik wil graag alle deelnemers bedanken voor hun geduld met het onvermijdelijke genetische jargon.
De “systems approach” die binnen PE gehanteerd wordt, en die voor mij persoonlijk een element is met een duidelijke meerwaarde, was voor mij aanleiding om in het inleidende hoofdstuk de nadruk te leggen op positie van mijn onderzoek in het grotere landbouw-vraagstuk.

Voor een aangename onderbreking van de werkzaamheden zorgde frequent een klaverjas-middagpauze, en ook buiten werktijd waren de “labuitjes”, wadloop en ski-weekends, bowlingavonden en klaverjasdrives altijd een leuke en gezellige tijdsbesteding. Een ieder die hieraan heeft bijgedragen, binnen en buiten de personeelsvereniging, bedankt!

Mijn ouders wil ik bedanken voor hun aanmoedigingen tijdens mijn studie en voor de keren dat ze als klankbord wilden fungeren. Dat geldt tevens voor mijn goede vrienden, Giliam en Mila. Het belangrijk om zo nu en dan de dagelijkse beslommeringen te relativeren en in een ander licht te plaatsen, onze gesprekken tijdens gezellige etentjes en vakanties boden mij daarvoor een goede gelegenheid. Giliam, ook bedankt dat je ondanks je drukke bezigheden op allerlei vlak mijn paranimf wil zijn.

Tenslotte, last but not least, Kitty, Tim en Max. De afgelopen jaren waren er perioden waarin het AIO-schap ook buiten werktijd de nodige aandacht vroeg. Dank voor jullie begrip en geduld. Kitty, speciale dank voor je niet aflatende steun de afgelopen jaren, juist in een periode die door afstuderen, sollicitaties en de geboorte van onze kinderen zeker niet rustig genoemd kan worden. Je hebt nog heel wat kook en afwasbeurten van me tegoed! Zowel ikzelf als onze computer zijn nu weer wat vaker voor jou en de kinderen beschikbaar.

# Curriculum Vitae

Ralph van Berloo werd onder de naam Remco geboren in Alkmaar, op 15 juli 1969. Al na twee jaar verhuisde hij met zijn ouders naar Aalsmeer. Na het doorlopen van het VWO aan het Haarlemmermeerlyceum te Hoofddorp begon hij in 1987 een studie Elektrotechniek aan de TU Delft. Na 9 maanden stapte hij in Delft over op de studie Technische Informatica, waarvoor hij in 1989 het propedeutisch diploma behaalde. Vanwege een knagende biologische belangstelling begon hij in datzelfde jaar aan een studie Plantenveredeling aan de Landbouwuniversiteit Wageningen. Tijdens deze studie liep hij stages bij het veredelingsbedrijf Royal Sluis en bij de Universiteit van Lerida in Spanje, terwijl afstudeervakken werden voltooid op de vakgroepen Erfelijkheidsleer en Plantenveredeling. In 1994 werd deze studie met succes afgerond waarna hij, na een korte periode werkzaam te zijn geweest als toegevoegd onderzoeker op het CPRO-DLO, in 1995 werd aangesteld als assistent in opleiding (AIO) bij de vakgroep Plantenveredeling. Gedurende deze periode was Ralph actief binnen de personeelsvereniging en binnen de onderzoeksschool PE. Het AIO project en het schrijven van dit proefschrift werden voltooid in 1999, en vanaf januari 2000 is Ralph als Postdoc werkzaam bij het Departement Plantwetenschappen van Wageningen Universiteit en houdt hij zich bezig met een onderzoeksproject dat zich richt op het verkrijgen van resistentie in de cultuurtomaat tegen de schimmelziekte *Botrytis cinerea*.

120 pag

4312 regels

32300 woorden

langste zin: 61 woorden

gemiddelde lengte zin: 12 woorden