

Implementation of molecular markers for quantitative traits in breeding programs - challenges and opportunities

James B. Holland¹

¹ USDA-ARS Plant Science Research Unit and Department of Crop Science, North Carolina State University, Raleigh, NC 27695-7620 USA <http://www4.ncsu.edu/~jholland/homepage.htm>. Email james_holland@ncsu.edu.

Abstract

Molecular marker assisted selection (MAS) has been touted as a means to improve the efficiency and speed of plant selection programs. The most widespread use of MAS to date is to assist backcrossing of major genes into already proven, elite cultivars. Markers can aid selection for target alleles that are not easily assayed in individual plants, minimize linkage drag around the target gene, and reduce the number of generations required to recover a very high percentage of the recurrent parent genetic background. The use of MAS in forward-selection programs is less common and is mainly restricted to marker loci linked to major-effect genes with consistent linkage phase across numerous populations. MAS has not been widely used for the improvement of polygenic traits because quantitative trait loci (QTL) mapping techniques remain insufficiently precise and because QTL information cannot be easily extrapolated from mapping populations to other breeding populations. The accuracy of QTL mapping can be improved by increasing population sizes and the number of testing environments, but these same techniques also improve conventional phenotypic selection. Therefore, MAS for polygenic traits is mainly restricted to situations where phenotypic selection cannot be easily implemented (such as for traits that are difficult to phenotype on individual plants or in offseason nurseries). MAS will remain a specialized breeding tool until QTL mapping can be extended to estimation of breeding values across many diverse breeding crosses and subpopulations such as those that compose typical plant breeding programs.

Media summary

DNA fingerprinting techniques can be applied to plant breeding, mainly to rapidly transfer single genes between varieties, but conventional selection programs are still necessary.

Key Words

Selection, breeding, quantitative trait loci (QTL), linkage disequilibrium (LD)

Introduction

At present, DNA marker-assisted selection (MAS) is routinely used in plant cultivar development programs only for selecting alleles with large effects on traits with relatively simple inheritance. Despite recent suggestions that breeders should be able to “control all allelic variation for all genes of agronomic importance through a combination of precise genetic mapping, high-resolution chromosome haplotyping and extensive phenotyping” (Peleman and van der Voort 2003), it is not likely that markers will soon be generally useful for manipulating complex traits like yield. Beavis (1998) observed that, “it seems fair to say that molecular markers have had little impact on crop improvement despite hundreds of published QTL experiments during the last 10 years.” One obvious reason for this is that DNA marker technologies are relatively expensive and cannot be implemented without a substantial investment in technical infrastructure. Even as marker costs continue to decrease, however, breeders will still face some fundamental reasons that markers have not had a greater impact on breeding crops, such as limited precision of QTL mapping experiments done on an affordable scale and lack of integration between mapping experiments and already efficient conventional breeding methods. MAS has a high relative efficiency compared to conventional selection in some situations, but these are limited relative to overall breeding efforts. The objective of this review is to identify those situations in which MAS should be more efficient than conventional selection, so that existing DNA marker infrastructure can be used most wisely

and effectively. In addition, by investigating the reasons that MAS is not being routinely used for complex traits, we can better define critical research questions that should be addressed to widen the scope and utility of MAS in the future. Unless plant breeders develop robust methods to accurately define QTL positions and effects and integrate this information across multiple populations, MAS will remain a useful but limited breeding tool.

Efficient uses of marker-assisted selection

Markers have been most widely used in practical plant breeding to assist backcrossing of major genes into elite cultivars that were previously developed through conventional plant breeding and to select alleles with major effects on high-value traits when marker information is reliable across multiple populations.

Marker-assisted backcrossing

Markers are effective aids to selection in backcrossing in three ways. First, markers can aid selection on target alleles whose effects are difficult to observe phenotypically. Examples include recessive genes, multiple disease resistance gene pyramids combined in one genotype (where they can epistatically mask each other's effects), alleles that are not expressed in the selection environments (e.g., genes conferring resistance to a disease that is not regularly present in environments), or genes whose phenotypic assays are more expensive than marker assays (e.g., some end-use or grain quality traits).

Second, markers can be used to select for rare progeny in which recombinations near the target gene have produced chromosomes that contain the target allele and as little possible surrounding DNA from the donor parent. Young and Tanksley (1989) demonstrated that large amounts of donor parent chromosomal material can remain around a target gene even after many generations of conventional backcrossing. Since this surrounding material may contribute to "linkage drag," especially if the donor parent is a wild relative or exotic germplasm source, minimizing the size of the introgressed segment from the donor parent is often critical to the successful backcross breeding of a new cultivar.

Third, markers that are unlinked to a target allele can also be useful in marker-assisted backcrossing, by permitting selection for those progeny with higher proportions of the recurrent parent genetic background. After two backcrosses, for example, progeny are expected to have recurrent parent alleles at *on average* 87.5% of loci unlinked to the target gene. However, variation around this average proportion exists among the progeny, and "background markers" can identify those progeny that are most similar to the recurrent parent. This can reduce the number of generations needed to obtain a genotype with 98 or 99% genetic similarity to the recurrent parent for a fixed sample size (Frisch et al. 1998).

A fine example of marker-assisted backcrossing that used markers for all three of these objectives was performed by Chen et al. (2000). They backcrossed the *Xa21* gene, which confers resistance to a wide spectrum of bacterial blight races, into the most popular rice line in China. During three backcross generations they selected for donor alleles at two markers tightly linked to *Xa21* and for recurrent parent alleles at flanking markers outside of the gene region to reduce linkage drag. In the third backcross generation, they used background selection on 128 RFLP loci to recover a line essentially identical to the recurrent parent cultivar, but possessing the *Xa21* allele.

Since marker-assisted backcrossing is a relatively simple procedure, probabilities of program success can be calculated and more efficient breeding strategies can be identified using simulations. For each backcross progeny selected, the probability of losing the target allele by recombination when selection is performed on a linked marker locus is simply equal to r , the recombination frequency between marker and target loci. If this is continued for t generations of backcrossing, the probability of losing the target allele by recombination is $1 - (1 - r)^t$. For example, if the marker locus exhibits 10% recombination with the target gene, there is a 10% chance of losing the target allele each generation, and a 27% chance of losing the target allele after three generations of backcrossing. Tightly linked markers, of course, do much better: three generations of backcrossing and selection on a marker locus with 1% recombination with the target gene has only a 3% chance of losing the target allele. Identifying very tightly linked markers can be

a laborious process, however, so selection on a pair of markers flanking the target locus can also effectively identify genotypes with the target allele, even if the flanking markers are not very tightly linked to the target locus. If marker loci *A* and *B* flank the target locus, one would select backcross progeny that have both *A* and *B* alleles from the donor parent. The probability of losing the target allele with flanking marker selection is equal to the probability of selecting a double recombinant progeny from among the doubly heterozygous backcross progeny. If the flanking loci have recombination frequencies r_A and r_B , respectively, with the target locus, the probability of losing the target allele due to double crossovers

within the selected region (ignoring crossover interference) is: $\frac{r_A r_B}{1 - r_A - r_B + 2r_A r_B}$. This probability can be much lower than the probability of losing the target allele based on selection for a single marker locus. For example, if the flanking markers each have 10% recombination frequency with the target locus, there is only a 1.2% chance of losing the target allele after a single generation. In any case, with tighter linkage, the chance of losing the target allele and the amount of linkage drag are reduced.

Codominant markers are most useful for marker-assisted backcrossing because selection among backcross progeny involves selection for heterozygous progeny. If a dominant marker, such as an AFLP band, is used for selection, it will be informative during backcross generations if the dominant allele (conferring band presence) is linked to the donor parent allele. If the recessive allele (conferring band absence) is linked to the donor parent allele, then all backcross progeny will either be heterozygous or homozygous for the dominant allele that produces the marker band, so the marker will be useless for selection among backcross progeny. Progeny testing of each individual in each backcross generation would be required to identify those segregating for the recessive marker allele, thereby doubling the number of generations required to perform the backcrossing procedure.

Frisch et al (1998) used simulations to compare several different backcrossing strategies in terms of how quickly they recovered a large proportion of the recurrent parent genotype. They found that recovery of 97% or more of the recurrent parent genotype was accomplished in one or two generations less with MAS than traditional selection. They also found that screening more than 20 individuals per generation or more marker loci than about one per 20 cM had little effect on rate of recovery of the recurrent parent genotype. They recommended a four-step selection process to quickly recover the recurrent parent genotype as follows: (1) select individuals carrying the target allele, (2) select individuals homozygous for recurrent parent genotype at loci flanking the target locus, (3) select individuals homozygous for recurrent parent genotype at remaining loci on the same chromosome as the target allele, and (4) select one individual that is homozygous for recurrent parent genotype at most loci (across whole genome) among those that remain.

While MAS may provide greater relative selection efficiency, MAS programs also require greater economic efficiency than traditional backcrossing to justify their implementation. Morris et al. (2003) compared costs of marker-assisted and traditional backcrossing of a single major gene into an elite line and found, as expected, that MAS was faster but cost more than traditional selection. They concluded that, "the cost-effectiveness of DNA markers depends critically on four parameters: (1) the relative cost of phenotypic vs. genotypic screening, (2) the time savings achieved using MAS, (3) the size and temporal distribution of benefits associated with accelerated release of improved germplasm, and (4) the availability to the breeding program of operating capital. All four of these parameters can vary significantly between breeding projects..." (Morris et al. 2003). Their last point indicates why private industry has adopted MAS more rapidly than public sector programs, and why breeding programs in developed nations and internationally funded programs have adopted MAS more rapidly than public programs funded by developing nations.

A potential drawback of marker-assisted backcrossing of major disease resistance genes is the widespread deployment of a few resistance genes, resulting in faster evolution of pathogen virulence to these genes. For example, the bacterial blight resistance gene, *Xa21*, has been backcrossed into rice cultivars that are widely grown in China and India, and also into IRRI's elite new plant type lines that are widely adapted (Chen et al. 2000; Sanchez et al. 2000; Singh et al. 2001). *Xa21* is a highly desirable

allele because it confers resistance to most known races of the pathogen (Chen et al. 2000; Singh et al. 2001), but the global scale of the deployment of this resistance gene increases the selection pressure for virulence in pathogen populations, which could lead to rapid spread of pathogen types able to overcome the resistance gene (Holland 2001a). One way breeders are trying to protect this valuable gene is by pyramiding it with other resistance genes, such as *xa5* and *xa13*, also with MAS (Huang et al. 1997; Sanchez et al. 2000; Singh et al. 2001).

More fundamentally, marker-assisted backcrossing is a limited use of marker technology because backcrossing is the most conservative of breeding methods (Lee 1995). Backcrossing improves current cultivars only at a few genes at a time. Backcrossing alone will not produce new combinations of alleles that are needed to generate improvements in multiple quantitative traits. Instead, “forward crossing” techniques that attempt to recombine alleles throughout the genome are needed for progress to occur for multiple traits.

Marker-assisted selection in “forward crossing” programs

Eagles (2001) reviewed the use of DNA markers in wheat breeding programs in Australia and observed that they regularly use 19 marker assays in cultivar development programs. The markers are linked to or diagnostic for disease resistance, abiotic stress resistance, or grain quality traits. A good example is MAS for two *Cre* genes, conferring resistance to cereal cyst nematode. “Although expensive, implementation was considered cost-effective because of the importance of the disease to wheat production in southern Australia and the high cost and unreliability of the bioassay,” (Eagles et al. 2001). Koebner and Summer (2003) suggested that the most appropriate targets for MAS in wheat breeding programs currently include *Fusarium* head blight (which is difficult to evaluate phenotypically), rust resistance, and virus resistance. Quality traits may be future targets, but MAS in forward crossing is expected to be most useful for a relatively small number of genes, primarily those affecting traits that are difficult to phenotype.

Another good candidate for widespread use of MAS in forward crossing programs is soybean cyst nematode resistance. MAS for this trait seems to be better than conventional selection because the phenotype is very difficult to score, the trait causes economically important yield losses, the genetics of resistance are relatively simple (one major gene, *rhg1*, is needed for resistance, although additional minor genes may also be necessary), and a tightly linked (1 – 2 cM) marker for this gene was found to be highly predictive across three populations (Young 1999).

Examples where MAS has been, or is expected to soon be, an important part of mainstream forward crossing breeding programs have in common two important factors. First, the markers are tightly linked to a small number of loci with relatively large effects on traits that are difficult or costly to accurately phenotype. Second, specific marker alleles are associated with desired alleles at target loci consistently across multiple breeding populations. This second point is key because it eliminates the need to establish the linkage phase between markers and their target alleles in every population. Markers must be consistently diagnostic for target alleles to implement MAS in forward crossing programs where many crosses are made annually between constantly changing sets of breeding parents. Unfortunately, these two situations are not expected to be generally applicable for most traits and most populations (Luby and Shaw 2001).

The soybean cyst nematode example illustrates a typical situation where a marker allele for a favorable trait is expected to be consistent across breeding populations. The resistance allele *rhg1* was introduced to the relatively narrow North American soybean breeding gene pool from a small number of exotic germplasm lines (Cregan et al. 1999). Therefore, marker alleles tightly linked to the resistance allele in the original mapping populations will also be linked to the resistance allele in other populations derived from the same exotic lines. Resistant progeny lines developed from such crosses will also tend to have the same marker allele as their resistant parent, so the marker will be useful in most future populations made by crossing adapted resistant lines to adapted susceptible lines. It is possible for resistant and susceptible elite lines to have an identical in state marker allele, in which case the marker would be non-informative in crosses between the pair of lines. The ability to discriminate resistant from susceptible genotypes can be increased by using marker loci with many alleles (such as simple sequence repeats,

SSRs) and by identifying multiple marker loci closely linked to the target gene, permitting discrimination on the basis of the haplotype of several markers rather than just the genotype at one marker. For example, Cregan et al. (1999) developed two SSR markers tightly linked to the *rhg1* gene. Neither marker alone could distinguish all resistant from all susceptible genotypes, because of identical in state alleles shared by some resistant and susceptible lines, but the two markers together could discriminate almost all resistant and susceptible lines. One resistant cultivar carried the susceptible allele at both loci, presumably due to recombination between marker and resistance loci during line development. Thus, recombination can change the linkage phase between markers, but if MAS is used first to select putatively resistant lines, followed by phenotypic evaluation of resistance, the linkage phase will remain intact in all selected progeny. Therefore, MAS can be self-reinforcing, ensuring that the same set of markers will be effective in future crosses.

The relative success of markers for cereal cyst nematode resistance and for boron toxicity tolerance in the Australian wheat breeding programs also illustrates the importance of gene pool-wide linkage disequilibrium (LD) between markers and trait loci for the success of MAS. LD exists when alleles at different loci occur together in genotypes more frequently than expected by chance (Flint-Garcia et al. 2003). Markers for two cereal cyst nematode resistance genes derived from a landrace and a wild relative of wheat (Ogbonnaya et al. 2001) are being employed successfully to pyramid the two resistance genes in the same cultivar (Eagles et al. 2001). In contrast, markers identified for major boron toxicity tolerance genes within crosses of cultivated wheat (Jefferies et al. 2000) have not been as useful for cultivar development (Eagles et al. 2001). Although MAS should be efficient for both traits because they are difficult to score phenotypically, the nematode resistance markers are more useful because they are tightly linked to the target genes and derived from germplasm sources outside of the elite Australian wheat gene pool (Ogbonnaya et al. 2001), increasing their level of LD between with the resistance genes across all crosses. In contrast, the boron tolerance genes were mapped within a cultivated wheat cross. Since the alleles mapped at the boron tolerance and marker loci have been in the same gene pool for a longer time, there has been greater opportunity for recombination between them. As a result, there is a greater likelihood that the LD between markers and tolerance loci observed within the mapping population will not be maintained across populations. Indeed, the marker alleles associated with boron toxicity resistance in the mapping population have been observed in many susceptible lines in Australian wheat breeding programs (Eagles et al. 2001).

The probability of consistent LD between marker loci and major resistance genes across populations in the examples given was increased in several ways. First, the markers are tightly linked to the target gene. In some cases where MAS is being employed successfully, e.g., the wheat cereal cyst nematode resistance genes (Ogbonnaya et al. 2001), the *Xa21* locus in rice (Chen et al. 2000), and grain storage protein subunits in wheat (Eagles et al. 2001), the markers identify the alleles directly. In these cases, since the markers are part of or products of the target genes themselves, LD between marker and target gene is complete. Second, the original sources of resistance are distinct from the elite germplasm pools that breeders were using for forward crossing, making the resistance-associated markers consistently useful. Third, the markers can be verified as being diagnostic for the resistance alleles by comparison of the marker genotypes to the previously known phenotypes of cultivars and breeding lines. Unfortunately, this will not be possible for many polygenic traits, such as yield.

Once consistently useful marker loci are identified, MAS can be highly effective in these instances because the resistance phenotypes are not only affected by genes with major effects, but also by the environment. Thus, MAS can be implemented on a single-plant basis (such as in F_2 populations) where conventional selection would not even be attempted because of the extremely low heritability of the trait on a single-plant basis. Bonnet et al. (2004) demonstrated that selection for F_2 plants that are either homozygous or heterozygous for the desired alleles at several marker loci (" F_2 enrichment") is an efficient method to reduce the number of lines that require extensive phenotypic evaluation in later generations.

Although currently MAS is used less in forward crossing than in backcrossing programs, forward MAS should become more common after novel alleles have been introgressed from exotic germplasm or wild relatives into elite genetic backgrounds. Marker-assisted backcrossing may be an important first step for implementing MAS in forward crossing programs for some traits. The introgression lines resulting from

backcrossing programs can be used as parents in crosses with other elite germplasm to improve the genetic background for multiple traits. At the same time, markers can be used in these populations to select for the linked major genes that were introgressed into one of the parents by marker-assisted backcrossing.

Challenges of implementing marker-assisted selection for polygenic traits

Many important agricultural traits, including yield, are under polygenic control with considerable environmental influence and genotype-by-environment interaction on trait expression. Such traits are the most difficult to breed for, typically requiring large-scale, multi-environment testing in order to make progress from selection. DNA markers could have a great impact on plant breeding if they could be used to aid selection for quantitative traits. There have been some successful uses of MAS for polygenic traits in plants (Johnson and Mumm 1996; Schneider et al. 1997; Stuber 1998; Tanksley et al. 1996; Yousef and Juvik 2001). Unfortunately, there are at least as many examples where MAS was not sufficiently better than conventional selection to justify its cost (Beavis 1998; Eathington et al. 1997; Stromberg et al. 1994; van Berloo and Stam 1999; Willcox et al. 2002).

Why is MAS not more generally effective, and are there situations in which MAS for polygenic traits will be more efficient than phenotypic selection? Barriers to implementing MAS for polygenic traits include problems in accurately localizing and estimating the effects of the quantitative trait loci (QTL), the difficulty of improving the already substantial gains from selection when heritability is high, inability to infer QTL effects from one breeding cross to another, and difficulty in integrating QTL mapping procedures into efficient breeding methods. Some of these difficulties are currently intractable for breeding programs in general, but may be solvable for specific breeding circumstances. Future research goals will include attempting to solve the difficulties of QTL mapping and cross-population inference.

Accuracy of QTL Mapping

The first major difficulty in implementing MAS for polygenic traits is the limited accuracy of QTL position and effect estimates for these traits. Since QTL do not have discrete phenotypic effects, they cannot be mapped as Mendelian loci; instead their chromosomal positions are typically inferred by calculating the likelihood of observing the phenotypic data conditional on a QTL position, repeating this calculation for many positions, and choosing the QTL position and effect with the highest likelihood. Although specific maximum likelihood locations can be stated, there is always a good possibility that the QTL is not located precisely at the maximum likelihood position. Exact confidence intervals for QTL positions are not easily obtained (Visscher et al. 1996), but typical approximate confidence intervals for QTL positions are on the order of 20 cM (Dekkers and Hospital 2002; Kearsley and Farquhar 1998; Lee 1995). The size of QTL confidence intervals cannot be reduced simply by increasing marker density beyond about 10-cM spacing. Instead, improved resolution of QTL positions requires increased mapping population sizes or the creation of near-isogenic lines with overlapping introgressions near the QTL (Stuber 1998; Zamir 2001).

Inaccurate QTL localizations result in the need to select for marker loci covering large genomic segments to be certain that target QTL alleles are retained in selected progeny. This has the unfortunate effect of maintaining potentially unfavorable linkages between QTL that affect one or more target traits. For example, fine-mapping a region that appeared to contain one major QTL for yield in maize revealed that there were at least two yield QTL in the region tightly linked in repulsion phase (Graham et al. 1997). Therefore, selection for the region encompassing both QTL (which would have been the logical procedure if fine mapping had not been performed) would actually limit response to selection by maintaining linkage between favorable and unfavorable alleles in the region. In comparison, phenotypically based selection methods would provide more opportunity to recover a rare recombinant genotype that had favorable alleles at both linked QTL. This same principle can be extended to genome regions affecting multiple traits, as correlated traits are often affected by linked QTL. Unfavorable QTL linkages may be maintained via MAS if a large chromosomal region must be selected.

The differences between the phenotypic effects of the alleles at a QTL are also poorly estimated with typical QTL mapping procedures. Unless huge numbers of genotypes are sampled from a mapping population, effects of QTL even on different chromosomes are not independent. This prevents accurate simultaneous estimation of multiple QTL effects. The result is that in any one sample of progeny from a mapping population, too few QTL are identified as significant, and the effects of those QTL that are identified are overestimated (Beavis 1998). The problem gets worse as heritability and population size decrease and the true number of QTL increases (Beavis 1998). For example, simulation studies indicate that if 10 loci affect a trait of 30% heritability and a population of 100 F_2 progeny is used for mapping, each true QTL will be identified as significant with only 9% probability, and the variance explained by each QTL that is detected is overestimated by 5.6 times its true value (Beavis 1998)! The problem is much worse if 40 QTL, each with a small effect, are segregating. Very large population sizes (on the order of 1000) are required to obtain reasonably accurate QTL position and effect estimations. Thus, despite the frequency of reports of small numbers of QTL with fairly large effects underlying grain yield in maize when fewer than 300 progeny are tested (Beavis 1998), the largest-scale yield QTL mapping study in maize based on nearly 1,000 progeny families tested in 19 environments discovered 28 QTL for yield, all with small effects (Openshaw and Frascaroli 1997). Similarly, Utz. et al. (2000) reported that QTL models explaining about 30% of the genotypic variation for maize grain yield in mapping populations of 230 families only explained about 15% of the genetic variation in independent samples of progeny from the same population.

These results suggest that, first; prediction of relative efficiency of MAS based on QTL results from typical population sizes can be greatly misleading (Bohn et al. 2001; Dekkers and Hospital 2002). Second, if many QTL remain undetected, the potential gain from MAS will be reduced. The problem of poor QTL effect estimation seems to be even more important than the problem of estimating QTL positions. Bernardo (2001) simulated MAS when the positions, but not the effects, of all of the genes controlling yield were known without error. When only a few (e.g., 10) genes affected yield, marker information was helpful and contributed to improved gain from selection, and the relative efficiency of MAS increased as heritability decreased. But if 50 or 100 loci affected yield, then selection on gene information was no better and sometimes worse than phenotypic selection. The problem was that gene effects for many loci could not be reliably estimated even from populations of 500 to 2000 progeny. Gene effect estimates have a high variability and when the true effects are small, they are estimated so poorly as to lead to worse prediction of genotypic values than phenotype data alone would provide.

All of these difficulties increase if epistasis is important. Epistasis causes the allelic effects of one QTL to depend on the genotype at a different locus (Holland 2001b). Accurate characterization of epistatic QTL is more difficult because it requires estimation of QTL main effects plus their epistatic interaction effects with other QTL. Epistasis also complicates the use of MAS because specific combinations of alleles are the ideal targets of selection (Holland 2001b). Although the relative importance of epistasis is still debated, epistasis has sometimes reduced gains from MAS below what was predicted (Holland 2001b).

Limits to relative efficiency of MAS imposed by trait heritability

Lande and Thompson (1990) showed that MAS is most effective when breeding values are predicted by an index of QTL genotypic values (inferred from linked marker genotypes and QTL effect estimates) and phenotypic values. As heritability increases, phenotypic data improve as estimators of genotypic values, and the relative weight placed on marker scores compared to phenotypes decreases in the selection index. Numerous theoretical or simulation comparisons of MAS and phenotypic selection have demonstrated that generally there is little advantage to MAS for traits of heritability around 50% or more (Edwards and Page 1994; Hospital et al. 1997; Knapp 1998; Lande and Thompson 1990; van Berloo and Stam 1998). This is true even without accounting for the additional cost of MAS and even if QTL effects and positions are estimated perfectly. The reason is that when heritability is high, gain from phenotypic selection approaches the maximum possible given the genetic variance, leaving little room for additional improvement by the use of markers. It is perhaps not recognized widely that heritability for complex, polygenic traits such as maize grain yield, although nearly zero on an individual plant basis, is often greater than 50% on a family mean basis when measured in well-defined target environments (Eathington et al. 1997; Openshaw and Frascaroli 1997). Although there is generally an upward bias in these

heritability estimates, it is certain that heritability on a family mean basis can be increased experimentally by increasing the number of replications and environments in which progeny families are tested and by improving experimental precision (Holland et al. 2003). Most traditional plant breeding programs are actually quite efficient at estimating breeding values of quantitative traits by multi-environment phenotypic evaluations.

Because increasing trait heritability requires time-consuming and expensive field testing, it is tempting to think that this stage of the breeding process could be more replaced by selection on markers. Unfortunately, in order to conduct MAS using current procedures, one already must have mapped QTL in the population, and large-scale experiments and good phenotypic data are required to accurately map the QTL (Beavis 1998). Increasing the numbers of progenies, replications, and testing environments in the mapping stage will increase the family mean heritability and permit more accurate identification of QTL positions and effects, thus improving the effectiveness of MAS. Once the effort has been expended to collect good phenotypic data and improve the trait heritability, however, phenotypic selection based on family mean values will also be highly effective (Dekkers and Hospital 2002; Moreau et al. 2000; van Berloo and Stam 1998). **This is the “catch-22” of MAS for polygenic traits: If phenotypic data are poor indicators of genotype values, it is difficult to adequately map QTL to implement MAS, but if phenotypic data are good, MAS is not needed to make genetic improvement.**

Whereas the simulation studies cited above found that MAS was really useful only for traits with heritability lower than 50%, they assumed that all QTL positions and effects were accurately identified. In reality, this would not be true, and would be less so with lower heritabilities. When costs and potential inaccuracies of QTL mapping are simulated as part of the MAS process, MAS is substantially more efficient (>1.2 times) than phenotypic selection only when heritability is below 40% and the true number of QTL is small (e.g., 5) (Moreau et al. 2000). If more QTL affect the trait, their average effects are smaller and MAS often becomes less efficient than phenotypic selection when markers cost more than evaluation plots. MAS becomes relatively more efficient as heritability decreases, but only if investment in the program increases, permitting more genotypes to be analyzed (Moreau et al. 2000). This may explain why MAS is adopted faster by commercial breeding programs with greater access to capital resources than public programs.

Two experiments to test the effectiveness of MAS in early-generation testing in maize illustrate the limits on the relative efficiency of MAS imposed by heritability (Eathington et al. 1997; Stromberg et al. 1994). In both experiments, F_2 plants were crossed to a tester and their hybrid progeny evaluated for grain yield in multiple environments. Inbred lines were derived from each F_2 plant, crossed to the same tester, and the testcross hybrids were evaluated in a later year. Early generation testing is effective if selections based on the F_2 generation result in selection of inbred lines that make superior hybrids in combination with the same tester. In the first experiment, neither conventional phenotypic selection nor MAS improved grain yield (Stromberg et al. 1994). In this case, the F_2 testcrosses were conducted in multiple locations in one year, which happened to be a severe drought year (Stromberg et al. 1994). Thus, neither the phenotypic data *per se*, nor the QTL information derived from that data, were indicative of performance in more typical production environments. This result illustrates that it is critical to obtain phenotypic data that are representative of performance in target environments with adequate environmental sampling, or genotype-by-environment interactions will hinder gain from selection with or without the use of markers. Adding markers to poor quality phenotypic data is not effective (Luby and Shaw 2001). In the second experiment, lines were tested in multiple locations and years, leading to higher heritability for grain yield (77%), but in this case, selection of the top 10% of families using MAS was not better than selection based on early generation phenotypes for predicting yield potential of the later generation lines (Eathington et al. 1997). This illustrates the converse principle, that when phenotypic data are very good and heritability is high, markers add little to gain from selection based on phenotype alone.

These examples illustrate that when MAS for polygenic traits is implemented in the same way as current conventional breeding methods, it is often not much more efficient than phenotypic selection. Once markers have been associated with target traits in a population, however, the markers can be used at any point in later generations, even when phenotypic selection is not effective. For example, MAS for polygenic traits can be conducted in off-season nurseries or greenhouses where the target trait cannot be

accurately phenotyped because of genotype-by-environment interactions. In fact, Hospital et al. (1997) demonstrated that MAS can be substantially more efficient than phenotypic selection on a yearly basis if MAS can be implemented in more generations per year, as in off-season nurseries. In such cases, MAS is even more efficient for high heritability traits, because the QTL can be identified with greater precision, improving the response to MAS in the offseason nursery where phenotypic selection cannot be accurately conducted. Similarly, MAS can be applied to select among individual plants within lines in pedigree breeding programs, even if single-plant phenotypes are not useful for conventional selection.

Integrating QTL mapping into large-scale breeding programs

Current MAS methods for polygenic traits involve mapping QTL in one population and applying that information for MAS in the same population. MAS for polygenic traits has been integrated with varying levels of success into the following breeding methods: early-generation testing (Eathington et al. 1997; Stromberg et al. 1994), recurrent selection (Yousef and Juvik 2001), selection of progeny pairs with high probability of producing the target genotypes from intermating (van Berloo and Stam 1999), selection for a target genotype (Stuber 1998) or “genotype building” (Dekkers and Hospital 2002), and introgression of exotic germplasm into elite lines using advanced backcross inbred selection (Tanksley et al. 1996; Tanksley and McCouch 1997; Zamir 2001). These applications have all been conducted in populations segregating for at most two alleles per locus, or in other words, within one or a few populations derived from biparental crosses. In contrast, plant breeders typically make many, often hundreds, of crosses each season to produce new breeding populations. Thus, even when MAS for polygenic traits has been successful, it has not been conducted on the scale of a whole breeding program. Repeating QTL mapping experiments for every breeding population in a program is simply not practical.

To scale up MAS for application across a whole breeding program will require methods to integrate QTL information across numerous populations (Beavis 1998; Dekkers and Hospital 2002). Instances where this has been achieved with markers for oligogenic disease resistances were discussed previously, but appropriate theory for polygenic traits is only beginning to be developed (Jansen et al. 2003). Polygenic traits like yield present additional complexity because, unlike oligogenic disease resistance, selection for yield is usually conducted exclusively in crosses between elite lines from a restricted germplasm pool. Within this germplasm pool, it is likely that most of the QTL with large effects have been fixed, leaving only minor effect QTL segregating (e.g., Openshaw and Frascaroli 1997). Favorable alleles tend to be dispersed across numerous lines and genotypes with very different QTL allelic composition may have similar yield potentials, resulting in genetic heterogeneity for yield. As a result, QTL mapped in one population may have little relevance to those mapped in other populations. Further, favorable QTL alleles will tend to be in repulsion phase LD with each other (Graham et al. 1997). Favorable QTL alleles that can be detected may represent regions where the net yield effect of alleles at several linked genes are positive. Conversely, regions with no detectable QTL may actually contain several genes linked in repulsion phase whose allelic effects cancel each other out. This genetic heterogeneity may be a critical impediment to MAS for polygenic traits within elite germplasm pools. At worst, MAS for polygenic traits may be limited to introgression programs where rare favorable alleles with relatively large effects may be introduced into adapted populations from exotic germplasm sources (Tanksley and McCouch 1997; Zamir 2001). At best, marker-QTL linkage estimates will have to be updated regularly to account for recombination occurring between many linked QTL as well as between QTL and markers. An added benefit of QTL effect updating should be improved gains from selection on epistatic QTL, since their effects depend on frequencies of alleles at other QTL, which will fluctuate over cycles of selection (Podlich et al. 2004).

One approach to applying QTL information across a whole breeding program is single large-scale MAS. This involves creating mapping populations by crossing each elite breeding line in the cultivar development program to a common tester line, mapping QTL within each population cross, then using this QTL information to make marker-assisted selections in early generation populations derived from intermating the elite lines (Ribaut and Betran 1999). This method has the advantage of requiring only n mapping populations, where n elite lines are used, to apply QTL information to $n(n-1)/2$ selection populations. A drawback to this method is that a substantial proportion of the important QTL for

complementary favorable and deleterious traits segregating in the selection populations may not be segregating in the mapping populations.

A more general approach will be to identify and select on the actual genes underlying QTL or on markers that are so tightly linked to important QTL that they remain in LD with the QTL across a wide spectrum of breeding crosses (Dekkers and Hospital 2002). Identifying the actual QTL genes themselves is still exceedingly difficult, and only a handful of examples exist of QTL that have been resolved to specific DNA sequences (Paran and Zamir 2003). QTL have been cloned by several approaches, including transposon tagging, map-based cloning, and candidate gene association analyses (Buckler and Thornsberry 2002; Paran and Zamir 2003).

Association mapping could be used for higher-throughput QTL cloning by identifying correlations between candidate gene sequence variation and phenotypic variation in breeding lines, without requiring special population development for mapping. To minimize false positive associations that arise due to similarities in genetic background, population structure can be estimated with random background markers and included in the analysis (Buckler and Thornsberry 2002). The extent of LD around candidate genes is the main factor determining resolution of association mapping (Buckler and Thornsberry 2002). If substantial LD extends over regions covering many genes, it will not be possible to statistically separate the effects of those genes. The extent of LD in plant populations is not well known. In a diverse sample of maize germplasm, historical recombinations around most genes reduce LD to near zero within 2 kb (Flint-Garcia et al. 2003), but within a restricted sample of elite maize germplasm, significant LD extends beyond 100 kb in some regions (Rafalski 2002). Populations that have undergone extensive bottlenecks will tend to have more extensive LD (Rafalski 2002). LD can therefore often be decreased in a population sample by including exotic germplasm in the study, but the candidate gene associations discovered in such samples may reflect variation between adapted and exotic germplasm, rather than the important QTL variation within elite germplasm. Association mapping may also be more difficult in self-pollinated species because their reduced opportunities for recombination will result in greater LD (Garris et al. 2003; Hagenblad and Nordberg 2002).

Association mapping can be combined with traditional QTL mapping approaches to take advantage of the strengths of the two approaches. Association mapping can be used with already existing breeding lines, does not require new population development, and can have higher resolution if historical LD around the candidate gene is less than LD in mapping populations. Traditional QTL mapping does not require candidate genes and uses specific populations that greatly reduce the potential for discovering false-positive associations due to population substructure that occurs in association tests. Joint linkage and LD mapping has been employed in animals to improve the resolution and statistical power of QTL mapping over either approach alone (Farnir et al. 2002; Meuwissen et al. 2002). At each genomic position tested for QTL effects, the parents of several mapping populations were grouped according to their identity by descent (IBD) based on identity in state (IIS) of marker haplotypes surrounding the QTL position. Within mapping populations, probability of progeny alleles being IBD with particular parental alleles was based on IIS of marker haplotypes surrounding the QTL and the known pedigree relationship of progeny and parent. In this manner, the phenotypic effects of various allelic classes could be estimated across mapping populations. Jansen et al. (2003) developed theory to combine QTL information across plant mapping populations based on haplotype sharing of their parent lines. This method could be joined with association mapping and pedigree-based linkage analysis by, first, genotyping mapping population parents and key ancestral breeding lines at high density to predict their probability of having IBD QTL alleles at any region based on haplotype sharing. Next, mapping population progeny and other breeding lines could be genotyped at lower density. The probability that a genomic region in a mapping or breeding line belongs to a particular IBD group can then be assigned based on its marker genotype and known pedigree relationship to the densely genotyped mapping parents or ancestral lines. This should be easier to accomplish in many plant species than in animals since highly homozygous lines are often used as parents, making the linkage phase between markers and QTL in those parental chromosomes certain. As in association mapping, background marker genotypes of the key ancestral lines, breeding lines, and mapping parents can be used to adjust the QTL estimates to reduce false positive associations due to population structure among the lines. The resulting data set may be severely unbalanced because different lines and populations were tested in different environments, but this problem can be addressed

using mixed models analysis and best linear unbiased prediction (Bernardo 2002; Holland et al. 2003). A great advantage of this method would be that previously collected phenotypic data on breeding lines could be used for QTL mapping, as long as remnant seed is available for DNA analysis. Furthermore, breeding values of lines based on their phenotype and QTL genotype can be obtained from the same analysis used for QTL mapping (Bernardo 2002).

In populations and genomic regions where LD decays rapidly, it should be possible to identify specific sequences underlying phenotypic effects (Buckler and Thornsberry 2002). Having identified the specific sequence variants that underlay the QTL, one could then use allele-specific markers to select for favorable QTL variants. Allele-specific markers for the QTL genes themselves could be used across existing populations and in newly developed populations without requiring additional mapping. This should be effective as long as the phenotypic effects of the gene are consistent across genetic backgrounds, i.e., if epistasis is not important for the expression of the QTL. These ideal types of markers could be easily integrated into breeding programs at any phase of cultivar development.

In populations where LD does not decay rapidly, fine mapping of QTL will be hindered, but haplotype information from markers may be useful to combine QTL information across populations, even if the markers are not directly at the QTL (Jansen et al. 2003). In either case, the fact that LD can vary among populations or germplasm pools indicates the importance of defining the reference population for association analysis and further use of LD in MAS. Furthermore, when markers and QTL are in linkage equilibrium across populations but in disequilibrium within populations, the phase of LD must vary across populations. In such cases, LD information could perhaps still be used to obtain the probability that a line possesses a particular QTL allele, conditional upon both the surrounding marker genotypes and the population to which the line belongs. Such conditional probabilities would also be necessary when genetic heterogeneity results from the association of different alleles with similar phenotypic effects in different populations. This may occur even for oligogenic traits. For example, Garris et al. (2003) reported that different haplotypes surrounding the *xa5* gene were associated with resistance in different ecotypes of rice. Similarly, Yu (2000) noted that markers for bacterial blight resistance genes were often restricted in utility to one gene pool, or even one race within a gene pool, of common bean because marker-gene LD was gene pool specific. These types of problems will only be more common for polygenic traits.

The adoption of MAS as a general breeding strategy for polygenic traits in plants is currently hindered by numerous difficulties. These include genetic heterogeneity for polygenic traits; the expense of applying high density marker assays across many populations; limited knowledge of how LD varies among species, populations, and genomic regions; lack of specific statistical approaches to combine disparate sources of phenotypic and genotypic data; and the computational difficulties that will arise when complex data sets are analyzed. The cost and difficulty of implementing such approaches means that if they are to be used, large private breeding programs probably will adopt them first. How much additional gain from selection such approaches will provide, and if they will be cost effective are unknown. Fortunately, conventional phenotypic selection is effective for most important agronomic traits and can be relied on for incremental improvements for the foreseeable future.

References

- Beavis WD (1998). In 'Molecular Dissection of Complex Traits'. (Ed. Paterson AH) pp 145-162. (CRC Press, Boca Raton, FL).
- Bernardo R (2001) What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Science* 41, 1-4.
- Bernardo R (2002) Breeding for quantitative traits in plants. (Stemma Press, Woodbury, MN).
- Bohn M, Groh S, Khairallah MM, Hoisington DA, Utz HF, Melchinger AE (2001) Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. *Theoretical and Applied Genetics* 103, 1059-1067.

Bonnet DG, Rebetzke GJ, Spielmeyer W (2004) Strategies for efficient implementation of molecular markers in wheat breeding programs - consideration of population size, allelic frequency, inbreeding and marker expression. *Molecular Breeding* (accepted).

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5, 107-111.

Chen S, Lin XH, Xu CG, Zhang Q (2000) Improvement of bacterial blight resistance 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. *Crop Science* 40, 239-244.

Cregan PB, Mudge J, Fickus EW, Danesh D, Denny R, Young ND (1999) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. *Theoretical and Applied Genetics* 99, 811-818.

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* 3, 22-32.

Eagles HA, Bariana HS, Ogbonnaya FC, Rebetzke GJ, Hollamby GJ, Henry RJ, Henschke PR, et al (2001) Implementation of markers in Australian wheat breeding. *Australian Journal of Agricultural Research* 52, 1349-1356.

Eathington SR, Dudley JW, Rufener GK, II (1997) Usefulness of marker-QTL associations in early generation selection. *Crop Science* 37, 1686-1693.

Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. *Theoretical and Applied Genetics* 88, 376-382.

Farnir F, Grisart B, Coppieters W, Riquet J, Berzi P, Cambisano N, Karim L, et al (2002) Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* 161, 275-287.

Flint-Garcia SF, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54, 357-374.

Frisch M, Bohn M, Melchinger AE (1998) Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Science* 39, 1295-1301.

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics* 165, 759-769.

Graham GI, Wolff DW, Stuber CW (1997) Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. *Crop Science* 37, 1601-1610.

Hagenblad J, Nordberg M (2002) Sequence variation and haplotype structure surrounding the flowering time locus *Fr1* in *Arabidopsis thaliana*. *Genetics* 161, 289-298.

Holland JB (2001a). In 'Dealing with Genetically Modified Crops'. (Ed. Wilson RF, Hou CT, Hildebrand DF) pp 60-83. (Amer. Oil Chem. Soc. Press, Champaign, IL).

Holland JB (2001b) Epistasis and plant breeding. *Plant Breeding Reviews* 21, 27-92.

Holland JB, Nyquist WE, Cervantes-Martinez CT (2003). In 'Plant Breeding Reviews'. (Ed. Janick J). Vol. 22, pp 9-111. (Wiley, New York.).

- Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. *Theoretical and Applied Genetics* 95, 1181-1189.
- Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, et al (1997) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theoretical and Applied Genetics* 95, 313-320.
- Jansen RC, Jannink J-L, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: Use of parental haplotype sharing. *Crop Science* 43, 829-834.
- Jefferies SP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam AKMR, et al (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). *Theoretical and Applied Genetics* 101, 767-777.
- Johnson GR, Mumm RH (1996) Marker assisted maize breeding. *Proceedings of the 51st Ann. Corn & Sorghum Res. Conf., Chicago, IL (Amer. Seed Trade Assoc.)*.
- Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; Where are we now? *Heredity* 80, 137-142.
- Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Science* 38, 1164-1174.
- Koebner RMD, Summer RW (2003) 21st century wheat breeding: plot selection or plate detection? *Trends in Biotechnology* 21, 59-63.
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756.
- Lee M (1995) DNA markers and plant breeding programs. *Advances in Agronomy* 55, 265-344.
- Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience* 36, 872-879.
- Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161, 373-379.
- Moreau L, Lemarie S, Charcosset A, Gallais A (2000) Economic efficiency of one cycle of marker-assisted selection. *Crop Science* 40, 329-337.
- Morris M, Dreher K, Ribaut J-M, Khairallah M (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. *Molecular Breeding* 11, 235-247.
- Ogbonnaya FC, Subrahmanyam NC, Moullet O, de Majnik J, Eagles HA, Brown JS, Eastwood RF, et al (2001) Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat. *Australian Journal of Agricultural Research* 52, 1367-1374.
- Openshaw S, Frascaroli E (1997) QTL detection and marker-assisted selection for complex traits in maize. *Proceedings of the 52nd Ann. Corn and Sorghum Res. Conf., Chicago, IL (Amer. Seed Trade Assoc.)*.
- Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. *Trends in Genetics* 19, 303-306.
- Peleman JD, van der Voort JR (2003) Breeding by design. *Trends in Plant Science* 8, 330-334.

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go. An effective approach for marker-assisted selection of complex traits. *Crop Science* 44, 1560-1571.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* 5, 94-100.

Ribaut JM, Betran J (1999) Single large-scale marker-assisted selection (SLS-MAS). *Molecular Breeding* 5, 531-541.

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS (2000) Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Science* 40, 792-797.

Schneider KA, Brothers ME, Kelly JD (1997) Marker-assisted selection to improve drought resistance in common bean. *Crop Science* 37, 51-60.

Singh S, Sidhu JS, Huang N, Vikal Y, Li Z, Brar DS, Dhaliwal HS, et al (2001) Pyramiding three bacterial blight resistance genes (xa5, xa13 and Xa21) using marker-assisted selection into indica rice cultivar PR106. *Theoretical and Applied Genetics* 102, 1011-1015.

Stromberg LD, Dudley JW, Rufener GK (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. *Crop Science* 34, 1221-1225.

Stuber CW (1998). In 'Molecular Dissection of Complex Traits'. (Ed. Paterson AH) pp 197-206. (CRC Press, Boca Raton, FL).

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, et al (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics* 92, 213-224.

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

Utz HF, Melchinger AE, Schon CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154, 1839-1849.

van Berloo R, Stam P (1998) Marker-assisted selection in autogamous RIL populations: A simulation study. *Theoretical and Applied Genetics* 96, 147-154.

van Berloo R, Stam P (1999) Comparison between marker-assisted selection and phenotypical selection in a set of *Arabidopsis thaliana* recombinant inbred lines. *Theoretical and Applied Genetics* 98, 113-118.

Visscher PM, Thompson RM, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143, 1013-1020.

Willcox MC, Khairallah MM, Bergvinson D, Crossa J, Deutsch JA, Edmeades GO, Gonzalez de Leon D, et al (2002) Selection for resistance to southwestern corn borer using marker-assisted and conventional backcrossing. *Crop Science* 42, 1516-1528.

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* 5, 505-510.

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. *Theoretical and Applied Genetics* 77, 353-359.

Yousef GG, Juvik JA (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. *Crop Science* 41, 645-655.

Yu K, Park SJ, Poysa V (2000) Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. *Plant Breeding* 119, 411-415.

Zamir D (2001) Improving plant breeding with exotic genetic libraries. *Nature Reviews* 2, 983-989.