Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years

Rex Bernardo*

ABSTRACT

In the mid-1980s, the development of abundant molecular markers, appropriate statistical procedures, and user-friendly computer software that implemented these statistical procedures permitted the detection of molecular markers associated with quantitative trait loci (QTL) for complex traits. Marker-assisted selection was then proposed as a means of exploiting markers linked to QTL to develop improved cultivars. But while thousands of marker-trait associations have been reported for many traits in different plant species, far fewer examples of successfully exploiting mapped QTL have been reported in the literature. Key lessons learned from applying markers in plant breeding include the following: (i) the purpose of detecting QTL should be clearly defined before embarking on QTL mapping; (ii) procedures for marker-based selection depend on the number of QTL; (iii) estimates of QTL effects for complex traits are often inconsistent; and (iv) gain per unit cost and time rather than gain per cycle should be considered. Future applications for complex traits will likely focus on predictive methodologies for marker-based selection before phenotyping and for marker-based selection without QTL mapping. These applications will take advantage of cheaper costs of genotyping than of phenotyping.

Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Cir., Saint Paul, MN 55108. Received 7 Mar. 2008. *Corresponding author (bernardo@umn.edu).

Abbreviations: BLUP, Best linear unbiased prediction; QTL, quantitative trait loci; RFLP, restriction fragment length polymorphisms; SNP, single nucleotide polymorphisms.

In the last two decades we have witnessed the widespread use of molecular markers to study complex, quantitative traits in different crop species. My thesis in this Perspectives article is that while many QTL have been reported in the literature, these reported QTL have not been adequately exploited in breeding programs. My objectives are to discuss practical lessons that I think we, as a scientific community of plant breeders and geneticists, have learned about marker-based selection for quantitative traits in the last 20 years, and to present insights on how we might best use molecular markers to improve complex traits in current and future plant breeding programs.

FROM PHENOTYPE TO (MARKER) GENOTYPE

My view of plant breeding became more complicated one autumn day in 1986, when a Ph.D. student in molecular genetics walked into our basement office at the University of Illinois and began describing a promising new technology called restriction fragment length polymorphisms, or RFLPs or "riflips" for short (Grodzicker et al., 1974). Back then I was a Ph.D. student in quantitative genetics and maize breeding with Professor John W. Dudley. Based on what my adviser had taught me, plant breeding for complex traits seemed fairly simple: a breeder (i) created genetic variation mainly by crossing good by good, (ii)

Published in Crop Sci. 48:1649–1664 (2008).
doi: 10.2135/cropsci2008.03.0131

677 S. Segoe Rd., Madison, WI 53711 USA

selected the best progenies in the cross, and (iii) synthesized the best progenies into a new and improved cultivar (Dudley and Moll, 1969). The breeding decisions pertaining to these tasks were made on the basis of phenotypic data. But this fellow graduate student showed us maize-breeding students a recent article that concluded that "the advent of RFLPs, by greatly increasing the total number of polymorphic genetic markers available to the agricultural community, may signal the advent of a new and promising era for the *understanding* and genetic *improvement* (emphasis mine) of quantitative economic traits through the use of marker-assisted breeding methodologies" (Beckmann and Soller, 1986).

Two parallel developments have allowed the aggressive use of molecular markers for studying quantitative traits. First, new marker systems such as randomly amplified polymorphic DNA, microsatellites or simple sequence repeats (SSR), amplified fragment length polymorphisms (AFLP; Vos et al., 1995), single nucleotide polymorphisms (SNP), and diversity arrays technology markers (DArT; Kilian et al., 2005) have increased the number and decreased the cost of markers in different crop species (Burrow and Blake, 1998; Bhattramakki and Rafalski, 2001). For example, the development of high-throughput technologies for SNP genotyping (Jenkins and Gibson, 2002; Syvänen, 2005) has led to, since 2000, a 40-fold increase in the number of marker datapoints generated and a six-fold decrease in the cost per datapoint in a commercial maize (*Zea mays* L.) breeding program (Eathington et al., 2007). The U.S. small-grains breeding community is served by four USDA regional genotyping centers that have made marker genotyping routine (Chao et al., 2006), and different molecular-marker service labs in North America, Europe, and Australia provide contractual genotyping services.

Second, statistical methods for detecting QTL and computer software for implementing these procedures have been developed. In this article I define "QTL mapping" as the general class of linkage-based methods for finding QTL, typically in a cross between two inbreds (Dudley, 1993). Methods for QTL mapping range from the simplest method of single-marker analysis (Sax, 1923) to more sophisticated methods such as interval mapping (Lander and Botstein, 1989; Haley and Knott, 1992), joint mapping (Kearsey and Hyne, 1994), multiple regression (Wright and Mowers, 1994; Whittaker et al., 1996), and composite interval mapping (Zeng, 1994). Association mapping, which requires collections of germplasm instead of biparental populations, has also been developed as a method for finding genes underlying quantitative traits (Hästbacka et al., 1992; Lazzeroni, 1997; Yu et al., 2006). Software packages for mapping with F_2 or backcross populations, selfed or recombinant-inbred progenies, or germplasm collections include *MAPMAKER/QTL* (Lincoln et al., 1993), *JoinMap* (Stam, 1993), *QTL Cartographer* (Basten et al., 1994), *PLABQTL* (Utz and Melchinger, 1996), *QGene* (Nelson, 1997), and *TASSEL* (Buckler, 2007). Private breeding companies in turn have developed their own QTL analysis tools that are integrated with company infrastructure for managing both marker and non-marker data (Eathington et al., 2007).

QTL, QTL EVERYWHERE

Reports of mapped QTL are now pervasive in the plant breeding literature. To obtain a rough estimate of the number of QTL mapping studies that have been conducted to date, I searched the Web of Science database for titles that contained the terms (i) "QTL," "QTLs," "quantitative trait locus," "quantitative trait loci," or "markers" + "associated" and (ii) the common names of 12 major crop species [barley (*Hordeum vulgare* L.), bean (*Phaseolus* spp. and *Vicia faba* L.), corn or maize, cotton (*Gossypium* spp.), oat (*Avena sativa* L.), potato (*Solanum tuberosum* L.), rice (*Oryza sativa* L.), soybean [*Glycine max* L. (Merr.)], sorghum (*Sorghum bicolor* L.), sunflower (*Helianthus* spp.), tomato (*Lycopersicon* spp.), and wheat (*Triticum aestivum* L.)]. This search indicated that more than 1200 QTL mapping studies have been reported for these 12 species. Previous reviews have indicated that QTL mapping studies have typically detected an average of 3 to 5 QTL for each trait (Kearsey and Farquhar, 1998; Bernardo, 2002, p. 309–311). If we assume that QTL mapping studies typically involve ~3 traits (as indicated by Kearsey and Farquhar, 1998), we can then surmise that at least 10,000 marker-trait associations in different plant species have been reported in the literature.

The reported QTL have typically accounted for a total of 40 to 60% of the phenotypic variance for the quantitative trait, and the distribution of the estimated genetic effects of individual QTL has been consistent with a quantitative trait being controlled by few QTL with large effects and many QTL with small effects (Kearsey and Farquhar, 1998; Bernardo, 2002). Overall, the QTL mapping literature has shown that if a breeder can develop a mapping population of $N = 100$–150 progenies derived from an F_2 or backcross population between two inbreds, obtain reasonably good phenotypic data for the traits of interest, and genotype the population with markers spaced about 10 to 15 cM apart, then an analysis of the phenotypic and marker data with an appropriate statistical method as implemented in a user-friendly software package will almost always lead to the identification of at least a few markers associated with each trait of interest. In short, in the last 20 years we have learned how to routinely map QTL.

In contrast, exploiting the QTL that have been mapped has not been routinely done. While the following statement may seem harsh [and I myself, with colleagues, have previously reported a total of 172 QTL in maize (Lu

et al., 2003a,b; Parisseaux and Bernardo, 2004)], the vast majority of the favorable alleles at these identified QTL reside in journals on library shelves rather than in cultivars that have been improved through the introgression or selection of these favorable QTL alleles. This is certainly not to say that there have been no examples of QTL that have been successfully used in breeding; successful examples (some of which will be cited later) provide useful insights on how QTL information can be utilized in a breeding program. Neither is this to say that there is no value in estimating the number, location, and effects of genes underlying quantitative variation. Yet from a plant breeding standpoint, we need to pay much greater attention to how identified QTL can be exploited in a breeding program or, more generally, how molecular markers can best be used to improve (instead of simply study) a complex trait. Toward this end, the following are key lessons learned from the successful as well as unsuccessful use of molecular markers to improve quantitative traits:

1. We should know why we want to find QTL
2. Procedures for marker-based selection depend on the number of QTL
3. Estimates of QTL effects for complex traits are often inconsistent
4. We need to consider gain per unit time and cost rather than gain per cycle

WHY DO WE WANT TO FIND QTL IN PLANTS?

While plant breeding relies heavily on the science of genetics, the primary goal of a plant breeder is fundamentally different from the primary goal of a geneticist. A plant breeder aims to develop improved cultivars, mainly through selection, whereas a geneticist aims to understand the inheritance and variation of traits. Breeding programs obviously require genetic variation for selection to act on, but genetic variation per se is not the main interest of a breeder. Given this context, two general goals of QTL mapping in plants to (i) increase our biological knowledge of the inheritance and genetic architecture (Mackay, 2001) of quantitative traits, both within a species and across related species; and (ii) identify markers that can be used to select for a complex trait. This latter goal, which focuses more on breeding than on pure genetics, can be further subdivided into two subgoals: (ii-a) identify a few major QTL (i.e., with large estimated effects) that can be introgressed by standard breeding procedures into other germplasm, or (ii-b) identify many QTL that can serve as the basis for selection for a complex trait in elite germplasm.

While these goals and subgoals are not mutually exclusive, they require different emphases on gene discovery versus selection, require different levels of stringency for declaring the presence of a QTL, require different levels of resolution for pinpointing QTL location, and usually require different types of germplasm (Fig. 1). As shown later in this article, the most useful breeding procedures differ between the two breeding subgoals. Given that there is no single best way to find and exploit QTL, the purpose of detecting QTL should therefore be very clearly defined before embarking on a QTL mapping study.

The QTL mapping approach has been proposed as a means of increasing our understanding of the genetics underlying quantitative variation (Beckmann and Soller, 1986). The results from QTL mapping have provided information on the genetic architecture of complex traits, i.e., estimated number of QTL and magnitude of their estimated additive, dominance, and epistatic effects in multiple environments (Mackay, 2001; Holland, 2007). But after 20 years of QTL mapping, we need to pause and seriously question how much biological (as opposed to statistical) information we have gleaned from the > 1200 QTL mapping studies that have been conducted in major crop species. Specifically, estimates of QTL locations or effects per se do not give us direct biological information (e.g., the product or function of each gene and the interactions among genes). The models that underlie QTL analysis are an extension of the models in quantitative genetics, and the models in quantitative genetics in turn are not necessarily designed to be biologically meaningful. In particular, the linear additive model, which assumes that the genotypic value for a trait can be modeled as the sum of the effects of unknown individual genes (i.e., additive

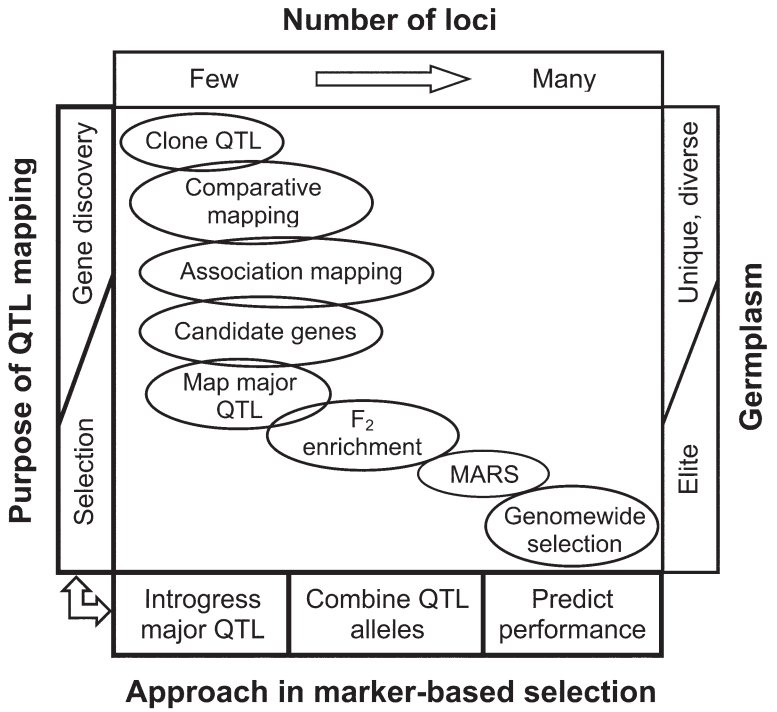

Figure 1. Goals and approaches for using molecular markers to study and select for complex traits in plants.

effects) and of combinations of unknown genes (i.e., dominance and epistatic effects), has been a simple yet useful statistical model for describing the inheritance and behavior of quantitative traits. The estimated genotypic or breeding values, while useful in selection, have limited biological meaning. By extension, a linear additive model has also been used to model the effects of QTL, the difference being that the individual genes or combinations of genes are now identified through known molecular markers. Estimates of the number of QTL and the magnitude of QTL effects are therefore biologically relevant only to the extent that the preconceived linear additive model for QTL effects is biologically accurate.

Furthermore, the wide use of the term "population" in the QTL mapping literature suffers from a double meaning of the term. From a population genetics standpoint, a population refers to a group of potentially interbreeding individuals (as in an F_2 or backcross population). But from a statistics standpoint, we need to remember that we can study only a sample of individuals (e.g., $N = 150$ F_2 plants) rather than the entire (statistical) population to which inferences would apply. Estimates of recombination distances among markers and estimates of the location, number, and effects of QTL are therefore subject to statistical error (Beavis, 1994).

On the other hand, QTL mapping studies have yielded useful biological information in terms of the importance of pleiotropy versus linkage for specific traits (Monforte and Tanksley, 2000; Chung et al., 2003) and collinearity in the organization of crop genomes (Kurata et al., 1994; Gale and Devos, 1998). Furthermore, QTL mapping has served as a springboard for the discovery of the underlying genes through map-based cloning of QTL (Frary et al., 2000), candidate-gene analysis (Pflieger et al., 2001), or comparative mapping (Paterson et al., 1995) (Fig. 1). Knowledge of the approximate locations of QTL has been used as a starting point for fine mapping by non-QTL-mapping approaches or for studying candidate genes that are close to the identified QTL and that may be the actual genes that affect the quantitative trait. At least 20 QTL have been cloned based on their map positions (Price, 2006). If the eventual goal is to clone QTL or identify candidate genes, the penalty of a false positive is severe. The statistical stringency or threshold for declaring the presence of the QTL must therefore be very high. Furthermore, the position of the QTL needs to be mapped precisely relative to closely spaced flanking markers.

As previously mentioned, association mapping in plants typically involves finding marker-trait associations among a diverse collection of inbreds with different genetic backgrounds, instead of among recombinant inbreds derived from an F_2 or backcross population between a pair of inbreds as in QTL mapping (Thornsberry et al., 2001; Flint-Garcia et al., 2003; Breseghello and Sorrells, 2006). The use of markers that represent polymorphisms at candidate genes would lead to a high resolution in association mapping, although random markers could also be used for genomewide association mapping. Spurious marker-trait associations arise due to different genetic backgrounds or pedigrees of the inbreds used, and association mapping needs to account for the population structure among the inbreds that comprise the association-mapping panel (Pritchard et al., 2000; Yu et al., 2006).

Any mapping procedure can detect only those QTL that are polymorphic in the population. The wide assortment of inbreds typically used in association mapping provides the wide genetic diversity needed for discovering a wide array of genes present in the plant species as a whole. This increased genetic diversity, however, often comes at the cost of a decreased mean performance or adaptedness of the germplasm used (Breseghello and Sorrells, 2006). To a geneticist, association mapping is therefore a powerful approach for discovering the genes that underlie quantitative variation (Hästbacka et al., 1992; Lazzeroni, 1997; Fig. 1). But to a breeder, association mapping with diverse, unadapted germplasm, rather than with elite germplasm, could often represent yet another way to discover additional QTL that would remain largely unexploited in selection for a complex trait, particularly if the contrasting QTL alleles detected by association mapping correspond to mutant forms that have no practical value. These consequences again underscore that the purpose of detecting QTL (e.g., gene discovery versus selection) should therefore be very clearly defined before embarking on a QTL mapping study (Fig. 1).

NUMBER OF QTL AND MARKER-BASED SELECTION

Finding and Exploiting a Few Major QTL

The nature of a trait may sometimes suggest that much of the quantitative variation is controlled by a few genes with large effects. In this situation, the objective of QTL mapping is clearly defined as finding a few major QTL. The subsequent breeding strategy is to introduce or pyramid these QTL, via standard breeding procedures, into elite germplasm to develop improved cultivars (Fig. 1). Exploiting a few major QTL therefore requires both gene discovery (i.e., QTL mapping) and selection.

Two examples that illustrate this approach are the *Fhb1* QTL for resistance to Fusarium head blight (caused by *Fusarium graminearum* Schwabe [telomorph: *Giberella zeae*]) in wheat (Anderson et al., 2008) and QTL for resistance to soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) (Concibido et al., 2004). Resistance to both Fusarium head blight and SCN is quantitative (Waldron et al., 1999; Concibido et al., 1994). The *Fhb1* QTL was first reported by Waldron et al. (1999) and the effect of *Fhb1* was confirmed in a second mapping population by

Anderson et al. (2001). With the use of flanking markers, the *Fbh1* QTL was then introduced into 19 different pairs of near-isogenic (for *Fhb1*) lines and was found to have a consistent effect of about 27% reduction in infected grains (Pumphrey et al., 2007). For SCN resistance, QTL have been found near the known *rhg1* and *Rhg4* resistance genes (Concibido et al., 1994, 2004; Webb et al., 1995). Molecular markers linked to these QTL have been routinely used to introduce SCN resistance into elite soybean lines (Cahill and Schmidt, 2004).

This approach exemplified by *Fhb1* in wheat and by *rhg1*- and *Rhg4*-related QTL in soybean therefore relies on (i) identifying unique germplasm as sources of useful QTL alleles (Fig. 1), (ii) finding closely linked markers for a few QTL that account for a substantial portion of the genetic variance for the trait, (iii) confirming the effect of the major QTL alleles in different genetic backgrounds, and (iv) deploying the QTL alleles widely in a breeding program. The *Fhb1* QTL allele originated from the Chinese cultivar Sumai 3 (Wang and Miller, 1988). The QTL alleles for SCN resistance were found predominantly in 'Peking' but were also detected in five plant introductions (Concibido et al., 2004). Given these specific germplasm sources, the linkage phase between the flanking marker loci and the QTL will remain the same across all recipient lines to which the given donor parent is crossed. Alternative marker loci, however, will have to be used to tag the QTL if the donor and recipient lines are not polymorphic at the original marker loci. This limitation will be circumvented after candidate genes for *Fhb1*, *rhg1*, and *Rhg4* are eventually cloned and confirmed (Hauge et al., 2006; Liu et al., 2006), as functional markers will then be available for the genes themselves.

In addition to QTL mapping, other approaches may be used to find major QTL with large effects (Fig. 1). Through comparative mapping, which exploits the collinearity among genomes of related species (Gale and Devos, 1998), the gene responsible for the loss of shattering in sorghum has been mapped to the same genomic locations in rice and in maize (Paterson et al., 1995). By association mapping in a gene bank of 600 potato cultivars, markers for the *R1* candidate gene were found associated with resistance to late blight caused by the *Phytophthora infestans* (Gebhardt et al., 2004). The marker alleles associated with increased resistance were traced to an introgression from the wild species *S. demissum*. This result indicates that while a diverse potato germplasm collection was used in association mapping, late blight resistance was ultimately from a unique source.

Exploiting Many QTL for a Complex Trait

The nature of a trait may sometimes suggest that much of the quantitative variation is controlled by many genes with small effects. An example is grain yield in cereal crops where a long breeding history suggests that if any major QTL were present to begin with, then the favorable alleles at these major QTL would have been fixed during the domestication process (Doebley, 2006) or during previous selection that led to the high-yielding cultivars in the current germplasm base. When much of the variation is controlled by many QTL that mostly have small effects, the find-and-introgress-QTL approach described in the previous section has limited applicability. The reasons for this are two-fold. First, as discussed in the next section, estimates of QTL effects for minor QTL are often inconsistent. Second, even if the effects for a large number of minor QTL were consistent, pyramiding favorable QTL alleles into a single cultivar becomes increasingly difficult as the number of QTL increases. Furthermore, breeders most often select for several traits at time. Even if each trait were controlled by only a few major QTL, selection for multiple traits would most likely involve the more difficult process of selecting for many QTL simultaneously.

To illustrate, suppose the objective is to pyramid the favorable alleles at only four major QTL. At the *i*th QTL, we denote the favorable allele by Q_i and the less favorable allele by q_i. One inbred parent has the Q_i allele at two of the QTL whereas a complementary inbred parent has the Q_i allele at the two other QTL. If the QTL are unlinked, a recombinant inbred with the Q_iQ_i genotype at each of the four QTL will occur an average of once every 2^4 = 16 recombinant inbreds. Now suppose the objective is to pyramid the favorable alleles at 10 unlinked QTL. If one inbred parent has the Q_i allele at five of the QTL whereas a complementary inbred parent has the Q_i allele at the other five QTL, then an inbred with the Q_iQ_i genotype at all 10 QTL will occur an average of once every 2^{10} = 1024 recombinant inbreds.

Unfavorable linkages and multiple sources of the favorable QTL alleles will decrease the frequency of an inbred with all the desired QTL alleles. Suppose that each inbred has the Q_i allele at only one out of 10 unlinked QTL, so that 10 inbred parents will have to be crossed to form a segregating population from which recombinant inbreds will be developed. Because the frequency of Q_i decreases from 0.50 in a biparental F_2 population to 0.10 in the 10-parent population, an inbred with the Q_iQ_i genotype at all 10 QTL will occur an average of once every 10^{10} = 10 billion recombinant inbreds.

When many QTL control the trait, our inability to control how genes segregate and assort during meiosis and how they come together during fertilization therefore severely limits our ability to synthesize the ideal genotype at many QTL. In the last 20 years we have learned that an effective strategy for exploiting multiple QTL is to increase the frequency of favorable marker alleles in the population, so that the probability of obtaining superior genotypes or the ideal genotype is consequently

increased. Suppose that by selection, the frequency of Q_i at each QTL is increased from 0.50 in an F_2 population to 0.75 in an improved population. The frequency of a recombinant inbred with the Q_iQ_i genotype at all 10 QTL will subsequently increase from an average of once every 1024 recombinant inbreds to an average of once every $1/(0.75^{10}) = 18$ recombinant inbreds.

Increasing the frequency of favorable alleles is also the rationale behind traditional recurrent selection based on phenotype (Allard, 1960, p. 283), and the concept of increasing allele frequencies before developing recombinant inbreds has therefore come full circle in the form of both phenotypic recurrent selection in the 1950s and marker-based selection in the 2000s. Marker-based selection to increase QTL-allele frequencies sacrifices an additional amount of time needed to develop improved populations in exchange for smaller population sizes that are required for obtaining superior genotypes. As indicated later in this paper in the Gain per Unit Time and Cost section, the use of off-season nurseries or greenhouses drastically reduces the amount of time needed to increase QTL-allele frequencies.

F_2 Enrichment and Marker-Assisted Recurrent Selection

Two related approaches have been proposed and used to increase the frequency of favorable QTL alleles at multiple loci (Fig. 1): (i) *F_2 enrichment* followed by inbreeding (Howes et al., 1998; Bonnett et al., 2005; Wang et al., 2007) and (ii) *marker-assisted recurrent selection* (MARS; Edwards and Johnson, 1994; Hospital et al., 1997; Koebner, 2003; Johnson, 2004; Bernardo and Charcosset, 2006). In both approaches the base generation is usually an F_2 population from the cross between two inbreds, although backcrosses, three-way crosses, or double crosses may also be used. The objective is to develop a recombinant inbred with superior per se performance for self-pollinated crops or with superior testcross performance for hybrid crops. Whereas F_2 enrichment usually involves only one generation of marker-based selection, MARS involves several cycles of marker-based selection.

In F_2 enrichment, F_2 plants with the q_iq_i genotype at one or more QTL are culled so that the remaining plants are carriers of the favorable alleles (i.e., Q_iQ_i or Q_iq_i genotypes) at all target QTL (Howes et al., 1998; Bonnett et al., 2005). Suppose a total of 10 QTL have been identified through a standard QTL mapping procedure in an F_2 population and that, for simplicity, markers are available for the QTL themselves. The probability that an F_2 plant has the Q_iQ_i or Q_iq_i genotype at a given QTL is 0.75. If the 10 QTL are unlinked, the expected frequency of F_2 plants with the Q_iQ_i or Q_iq_i genotype at all 10 QTL is $0.75^{10} = 0.056$. In other words, about one out of every 18 F_2 plants will be selected. With complete selection against the q_iq_i homozygote at each QTL, the expected frequency of Q_i increases from 0.50 to 0.67 at each locus (Howes et al., 1998). If recombinant inbreds are developed from the F_2 plants that remain after culling (Bonnett et al., 2005), the expected frequency of recombinant inbreds with the Q_iQ_i genotype at all 10 QTL is $0.67^{10} = 0.018$, or one in 55 recombinant inbreds. As previously indicated, without F_2 enrichment, the frequency of a recombinant inbred with the Q_iQ_i genotype at all 10 QTL is one in 1024 recombinant inbreds.

Increasing the frequency of Q_i from 0.50 to 0.67 via F_2 enrichment therefore increases the probability of recovering a recombinant inbred with the Q_iQ_i genotype at all target QTL, or at as many target QTL as practically possible. But a Q_i frequency of 0.67 may still not be sufficiently high if the target number of QTL is large (e.g., > 15 or so QTL). And regardless of the number of target QTL, further increases in the frequency of Q_i will increase the probability of recovering a recombinant inbred with the desired Q_iQ_i genotypes. A second round of enrichment at a later inbreeding generation may be performed to further increase the frequency of Q_i, but studies suggest little added advantage in culling F_3 or F_4 plants with the q_iq_i genotype at any of the QTL (Wang et al., 2007).

This limitation in F_2 enrichment is overcome in MARS, in which multiple cycles of selection are performed based on markers (Edwards and Johnson, 1994; Johnson, 2004; Eathington et al., 2007). Specifically, MARS involves (i) identifying F_2 plants or F_2-derived progenies that have the Q_i allele at most, if not all, of the target QTL; (ii) recombining selfed progenies from these selected individuals; and (iii) repeating the procedure for 2 to 3 cycles. In the same way that phenotypic recurrent selection is an alternative to phenotypic selection during selfing, MARS is therefore an alternative to the cull-and-inbreed process in F_2 enrichment. An obvious possible disadvantage of MARS, however, is the extra number of generations needed for cyclical selection based on markers.

Direct comparisons of F_2 enrichment and MARS have not been reported, but studies have shown that each procedure is effective in increasing the frequencies of favorable QTL or marker alleles. With enrichment in a wheat BC_1 population followed by marker-based selection among haploids (that were subsequently doubled), the frequency of the *Lr34/Yr18* rust-resistance genes increased from 0.25 to 0.60 (Kuchel et al., 2007). But the frequency of the underlying *Lr46/Yr29* rust-resistance genes, as evaluated from disease reactions from field tests, increased from 0.25 to only 0.27. This result was most likely due to the loose linkage between the actual *Lr46/Yr29* genes and the marker used to screen the BC_1 plants (Kuchel et al., 2007). In a sweet corn F_2 population, MARS increased the frequency of the favorable marker allele from 0.50 to ≥ 0.80 at 18 out of 31 markers used in selection (Edwards

and Johnson, 1994). In a second sweet corn F_2 population, the frequency of the favorable marker allele increased to ≥0.80 at 11 out of 35 markers used. Five marker loci in the first F_2 population and one marker locus in the second F_2 population became fixed for the favorable allele. However, frequencies of the favorable allele decreased or remained equal to 0.50 at five loci in the first population and at four loci in the second population.

These differences in the observed changes in marker allele frequencies were due to unfavorable linkages among markers as well as different weights that were given for each marker when ranking candidates for selection in MARS (Edwards and Johnson, 1994). Specifically, the MARS approach uses a selection index that gives weights to markers according to the relative magnitude of their estimated effects on the trait (Lande and Thompson, 1990; Edwards and Johnson, 1994). The selection index typically has the form $M_j = \Sigma\ b_i X_{ij}$, where M_j is the marker score of the jth individual; b_i is the weight given to the ith marker locus; and X_{ij} is equal to 1 if the jth individual is homozygous for the marker allele (at the ith marker locus) with the favorable effect and –1 if the individual is homozygous for the marker allele with the less favorable effect. The b_i weights can be obtained from multiple regression of trait values on X_{ij} (Lande and Thompson, 1990; Hospital et al., 1997).

Combining Favorable QTL Alleles versus Predicting Performance

The use of weights for different markers in MARS but not in F_2 enrichment underscores two approaches for using markers to develop superior germplasm for complex traits (Fig. 1). The first approach focuses on combining favorable QTL alleles in germplasm. Again, the underlying goal in F_2 enrichment followed by inbreeding is to eventually develop a recombinant inbred with the Q_iQ_i genotype at most, if not all of the target QTL. All of the target QTL are treated as equally important, and individuals that carry as many favorable marker alleles as possible are selected (Bonnett et al., 2005; Wang et al., 2007). The second approach focuses not on the number of favorable marker alleles present in a particular individual, but rather on the use of markers to predict the performance of the individuals so that those with the best predicted performance can be selected.

These two approaches would be equivalent if all QTL have equal effects. In this situation, the expected marker weights in MARS would all be equal to $b_i = 1.0$ and the selection index in MARS would be simply equal to the number of marker loci for which the candidate is homozygous for the favorable marker allele, i.e., $M_j = \Sigma\ X_{ij}$. As previously mentioned, however, the QTL mapping literature has indicated that QTL for a given trait do not have equal estimated effects (Kearsey and Farquhar, 1998; Bernardo, 2002). But even if the weights for different QTL vary, the two approaches will still be equivalent if the number of target QTL is small and the population is large. To illustrate, suppose four unlinked target QTL are screened among 150 recombinant inbreds. An average of one in $2^4 = 16$ recombinant inbreds will have the Q_iQ_i genotype at all four QTL. If the breeder selects the best five out of 150 recombinant inbreds, all five are likely to have the ideal Q_iQ_i genotype and, consequently, would also have the highest selection index values regardless of the b_i values used to calculate M_j.

In contrast, suppose 10 unlinked target QTL are screened among 150 recombinant inbreds. It is highly unlikely that one of the 150 recombinant inbreds will have the Q_iQ_i genotype at all 10 QTL (i.e., probability of one in 1024 recombinant inbreds). Based on a binomial distribution, the best five out of 150 recombinant inbreds will likely be fixed for the Q_iQ_i allele at eight out of the 10 QTL. If some QTL are more important than others so that b_i differs among the 10 QTL, recombinant inbreds that carry the same number of QTL alleles will differ in their performance. Furthermore, recombinant inbreds with more Q_i alleles may actually be inferior to recombinant inbreds with fewer Q_i alleles. Suppose the QTL are numbered according to the magnitude of the effects of their alleles, i.e., Q_1 has the largest effect, Q_2 has the second largest effect, and Q_{10} has the smallest effect. Assume that marker-based selection with an unweighted index leads to a recombinant inbred fixed for Q_3, Q_4, Q_5, …, Q_{10} (i.e., eight QTL alleles with the smallest effects). In contrast, marker-based selection with a weighted index leads to a recombinant inbred fixed for Q_1, Q_2, Q_3, … Q_6 (i.e., six QTL alleles with the largest effects). Depending on the distribution of QTL effects, the recombinant inbred fixed for eight favorable QTL alleles may be inferior to the recombinant inbred fixed for six favorable QTL alleles.

When combining favorable QTL alleles in germplasm, the number of target QTL should therefore be kept manageable. Or, the breeder may initially target a large number of QTL and be prepared to accept having fewer QTL alleles fixed in a recombinant inbred. Studies have suggested that for typical population sizes used in wheat, combining favorable marker alleles for more than 9 to 12 unlinked QTL does not seem feasible (Howes et al., 1998; Wang et al., 2007). Given that improvement is targeted at a limited number of QTL, the breeder needs to have a high level of confidence that the target QTL do not represent false positives. This implies that a stringent significance level (e.g., $P \leq 0.0001$) should be used to identify the QTL in the first place. Stringent significant levels unfortunately lead to an upward bias in the estimates of QTL effects (Beavis, 1994; Xu, 2003) and may lead to overly optimistic expected responses from marker-based selection.

For MARS, however, empirical and simulation research has shown that selection responses are increased if relaxed significance levels ($P = 0.20$ to 0.40) are used to identify which markers have significant effects and should therefore be selected (Edwards and Johnson, 1994; Hospital et al., 1997; Koebner, 2003). Relaxed significance levels in MARS allow selection for QTL with smaller effects, and the inclusion of minor QTL more than compensates for the higher frequency of false positives. Because MARS does not aim to directly control changes in QTL allele frequencies (Edwards and Johnson, 1994), the number of marker loci in MARS may be large, with the understanding that improved germplasm from MARS may not have the favorable allele across all QTL included in the selection index.

The relaxed significance levels in MARS are therefore contrary to the stringent significance levels required to identify QTL if the purpose is combining several QTL alleles in a recombinant inbred, introgression of a few major QTL, or gene discovery. Furthermore, less precision is required for pinpointing QTL locations when the purpose is to predict genotypic performance (as in MARS) than when the purpose is to combine favorable QTL alleles. In the latter, the QTL ideally should be tagged by a marker for the QTL itself, by a closely linked marker, or by two flanking markers that will prevent the loss of the underlying QTL because of double recombination. A dense linkage map should therefore be used to map QTL if the objective is to combine favorable marker alleles in a recombinant inbred. In MARS, the multiple markers used in calculating marker scores may account for the effects of one or more nearby QTL. Simulation studies for maize have indicated that for a population size of 144 F_2 plants, the response to MARS is largest when about 128 markers are used (Bernardo and Charcosset, 2006). This result indicates that markers used in MARS should be about 10 to 15 cM apart and that a dense linkage map is unnecessary for predicting performance in MARS. Overall, these differences in the required marker density and significance level for detecting QTL again underscore the need to clearly define the purpose of a QTL mapping study.

INCONSISTENCY OF ESTIMATED QTL EFFECTS

The estimated effects of QTL are often inconsistent and this inconsistency has forced plant breeders to focus on major QTL that tend to have more consistent effects (e.g., *Fhb1*, *rhg1*, and *Rhg4*) or develop breeding strategies that circumvent this inconsistency (e.g., QTL mapping conducted independently within each of several F_2 populations). Reasons for the inconsistency of estimated QTL effects include (i) different QTL segregating in different mapping populations, (ii) QTL × genetic background interaction, (iii) QTL × environment interaction, and (iv) the Beavis effect (Beavis, 1994; Xu, 2003).

We obviously would expect to detect different QTL in different mapping populations if particular QTL are segregating in some populations but not in others. Germplasm of diverse genetic backgrounds and with different selection histories would likely differ in their QTL alleles. Differences in segregating QTL, however, would not be an issue if a unique QTL allele is identified from a specific germplasm source and this source is used as a common donor parent. For example, when Sumai 3 or Sumai 3-derived inbreds were crossed to a series of inbreds that were susceptible to Fusarium head blight, the *Fhb1* locus was segregating in the different crosses because the *Fhb1* resistance gene was found uniquely in Sumai 3 (Waldron et al., 1999; Anderson et al., 2001; Pumphrey et al., 2007). This situation is unlikely, however, for other traits for which trait means of the parents are not widely divergent and for which no parent has a monopoly on favorable QTL alleles across many loci. A prime example would be grain yield among elite inbreds.

Even when a favorable QTL allele comes from a particular donor parent, the effect of the introgressed QTL allele may vary because of a general form of epistasis that has become known as QTL × genetic background interaction (Tanksley et al., 1989; Charcosset et al., 1994; Blanc et al., 2006). When the *Fhb1* allele was introgressed into 13 genetic backgrounds, the allele had its expected positive effect in 12 genetic backgrounds but a negative effect in one genetic background (Pumphrey et al., 2007). This negative effect may have been due to unfavorable interactions between *Fhb1* and unknown background genes in the recipient inbred. Interconnected populations, which are formed by crossing parental inbreds in a way that pairs of the resulting mapping populations have a parent in common, permit the study of QTL × genetic background interaction (Rebai et al., 1994; Charcosset et al., 1994; Jannink and Jansen, 2000). In a study of six interconnected F_2 maize populations among four parental inbreds, the percentage of significant QTL × genetic background interactions was 8% for grain moisture, 9% for silking date, and 42% for grain yield (Blanc et al., 2006). Given that grain yield is arguably the most complex of these three traits, these results suggest that QTL × genetic background interactions are most important for traits that are controlled by a large number of QTL with minor effects.

Yet even within a single mapping population, estimates of QTL effects may be inconsistent because of QTL × environment interaction and sampling error. One of the tenets of quantitative genetics is that genes affecting complex traits are subject to genotype × environment interaction, and as such QTL × environment interaction should also be expected. In a large QTL mapping study (population size of $N = 344$) in maize, a total of 107 QTL were detected for grain yield, grain moisture, kernel weight, protein concentration, and plant height (Melchinger et

al., 1998). About one third of these 107 QTL exhibited significant QTL × environment interaction. Examples of significant QTL × environment interactions or of QTL being detected in some environments but not in others have also been reported for other crop species including barley (Zhu et al., 1999), cotton (Paterson et al., 2003), oat (Zhu and Kaeppler, 2003), rice (Zhuang et al., 1997), soybean (Reyna and Sneller, 2001), sunflower (Leon et al., 2001), tomato (Paterson et al., 1991), and wheat (Campbell et al., 2003). Unlike the estimation and testing of QTL × environment interaction effects, the detection of a QTL in one environment but not in another is not necessarily due to QTL × environment interaction: this may simply be due to an environment having a high error variance that prevented the detection of a QTL in that environment. Regardless of the underlying cause, the detection of a QTL in one environment but not in others hinders the transferability of QTL mapping results.

The same environment or set of environments may be used to map QTL among segregating progeny in the same cross so that QTL × environment interaction, QTL × genetic background interaction, and differences in segregating QTL are not issues. Yet the results of QTL mapping may still differ due to the Beavis effect (Beavis, 1994; Xu, 2003). Beavis (1994) used both simulated data and $N = 400$ maize F_3 families derived from the B73 × Mo17 cross to determine the effects of a small N on the power to detect QTL and the accuracy of estimated QTL effects. On the basis of family means across environments, QTL mapping for plant height was performed (i) with the entire set of $N = 400$ and (ii) in each of four random subsets of $N = 100$ families each. A total of four QTL were detected in the combined mapping population of size $N = 400$. In contrast, only one to three QTL were detected in each of the subsets of $N = 100$ families. Furthermore, the R^2 values for individual plant-height QTL increased from 3 to 8% with $N = 400$, to 8 to 23% with $N = 100$. Other empirical studies in maize have led to similar results. In the Melchinger et al. (1998) study, a total of 31 QTL for plant height were detected in a mapping population of $N = 344$ maize F_3 families derived from a biparental cross. When a smaller but independent set of $N = 107$ families from the same biparental cross was used, only six QTL for plant height were detected. In a study by Schön et al. (2004), a total of 30 QTL for plant height were detected among testcrosses of $N = 976$ maize $F_{2:5}$ families. Multiple subsets of $N = 488$, 244, and 122 $F_{2:5}$ families were obtained by sampling without replacement. The number of QTL detected decreased to a mean (across multiple subsets of size N) of 17.6 with $N = 488$, 12.0 with $N = 244$, and 9.1 with $N = 122$.

These results, along with simulation studies (Beavis, 1994) and analytical results (Xu, 2003), show that a small N leads to (i) fewer QTL being detected and (ii) an upward bias in the estimated effects of the few QTL that are detected. For a trait controlled by 10 unlinked QTL and a heritability ranging from $h^2 = 0.30$ to 0.95, Beavis (1994) found that $N = 500$ progenies were required to detect at least half of the QTL. With 40 unlinked QTL, $N = 1000$ progenies were required to detect at least a quarter of the QTL. The effects of the detected QTL were greatly overestimated with $N = 100$, slightly overestimated with $N = 500$, and were close to their actual values with $N = 1000$.

For complex traits controlled by many minor QTL (rather than by a few major QTL), the inconsistency of estimated QTL effects has three important implications for plant breeders. First, because estimated QTL effects for traits such as grain yield or plant height have limited transferability across populations, QTL mapping for such traits will likely have to be repeated for each breeding population. This specificity for each population is demonstrated in MARS, where genotyping, phenotyping, and construction of a selection index are repeated for each population (Koebner, 2003). Second, because complex traits controlled by many QTL are likely subject to genotype × environment interaction, QTL mapping for the same population will likely have to be performed for each target set of environments. Third, because the effects of sampling error are large, population sizes of $N = 500$ to 1000 are recommended if the objective is QTL mapping per se for highly complex traits that are likely controlled by many loci (Beavis, 1994). Evaluating $N = 500$ to 1000 progenies for each cross of interest is unfortunately prohibitive in plant breeding programs.

GAIN PER UNIT TIME AND COST

We have learned in the last 20 years that marker-based selection can increase the gain per unit time and gain per unit cost in breeding programs, particularly when phenotyping for the traits of interest is time-consuming, expensive, and erratic. For example, screening for resistance to Fusarium head blight in wheat is routinely done in field or greenhouse tests but the results are often inconsistent (Campbell and Lipps, 1998). Whereas a single test is sufficient to discard highly susceptible individuals, multiple field tests at different locations are needed for reliable evaluations of resistance to Fusarium head blight (Fuentes-Granados et al., 2005). Although extensive screening and validation was required to initially identify the *Fhb1* QTL, the deployment of *Fhb1* has subsequently allowed simple marker-based selection among F_2 plants or F_3 families.

Detailed comparisons in soybean have likewise indicated that the cost and time required to screen for SCN resistance are lower with marker-based selection than with phenotypic selection (Concibido et al., 2004). Specifically, marker-based screening required 1 to 2 d at the cost of

$0.25 to 1.00 per sample. In contrast, an SCN greenhouse assay required 30 d at the cost of $1.50 to 5.00 per sample.

The impact on gain per unit time will be largest if the use of markers reduces the amount of time needed per cycle of selection or increases number of cycles of selection that can be grown per year. The reduction in time per cycle will likely be largest for perennial crops. Oil palm (*Elaeis guineensis* Jacq.), for example, requires 19 years per cycle of phenotypic selection based on testcross performance. Implementing MARS in oil palm would allow marker-based selection among physiologically immature palms and would reduce the time per cycle from 19 to 13 years (Wong and Bernardo, 2008).

Only one season of field trials can be done per year for annual crops in temperate regions. Gain per unit time for annual crops can be increased mainly by having multiple cycles of marker-based selection in greenhouses or off-season nurseries. The MARS approach in maize, soybean, and sunflower illustrates how markers can increase gain per unit time (Eathington et al., 2007). The MARS procedure in maize involves two steps. First, markers associated with the traits of interest are identified from field trials of Cycle 0 testcrosses in Year 1 (e.g., May–October 2008). Markers associated with the traits are then used to construct a marker selection index. The best families are recombined (November 2008–February 2009) to form Cycle 1. Second, up to three cycles of selection based on marker scores are conducted in Hawaii or Puerto Rico in Year 2 (Cycle 1 to Cycle 2 from March–June 2009; Cycle 2 to Cycle 3 from July–October 2009; and Cycle 3 to Cycle 4 from November 2009–February 2010). Genotyping during each of the three cycles of selection is done at the seedling stage so that the best plants can be identified before flowering and intermated to form the next cycle. In other words, selection and recombination are performed in the same generation.

The aggressive use of markers in a year-round nursery, where phenotypic measurements do not reflect performance in the target environments (e.g., the U.S. Corn Belt) but where the marker genotypes remain the same, is therefore the key element that increases the gain per unit time in MARS. Furthermore, quantitative genetic theory has indicated that marker-based selection will be most efficient relative to phenotypic selection if (i) the markers to be used in selection are identified in environments where h^2 is high and (ii) selection is subsequently performed in environments where h^2 is low (Dudley, 1993). The MARS scheme satisfies this "catch-22" situation (Holland, 2004): when marker-trait associations are identified in Year 1, a sufficient number of environments should be used so that phenotypic measurements are reliable. But when marker-based selection is performed in Year 2, h^2 is effectively low or near zero because the performance of individual plants in Hawaii or Puerto is a poor indicator of the plants' genotypic value given the U.S. Corn Belt as the target environment.

Simulation and empirical results have suggested that the per-cycle gain is actually lower during marker-based selection in MARS than with phenotypic selection based on testcross performance. Simulation experiments in maize have indicated that the cumulative gain from two cycles of marker-based selection in MARS was about 25 to 50% lower than the gain from one cycle of phenotypic selection (Bernardo and Yu, 2007). Similarly, empirical results for maize grain yield in six F_2 populations indicated that the gain from one cycle of marker-based selection was about 50% lower than the gain from one cycle of selection based on both phenotypic data and marker scores (Johnson, 2004). Because one cycle of testcross selection in maize requires 2 years, the larger number of cycles per year with marker-based selection (up to three cycles per year) than with phenotypic selection (0.5 cycle per year) compensates for the lower per-cycle response to marker-based selection. Overall, the gain per year for grain yield in maize is therefore larger with MARS than with phenotypic selection.

Gains per unit cost are difficult to compare because the cost of phenotyping varies greatly among traits and species, and the cost of genotyping varies according to the number of markers used and individuals genotyped. Assume that, with inflation, the cost of one maize yield-trial plot has increased from $10 in 1998 (Weyhrich et al., 1998) to $15 in 2008. If at least five locations are needed to obtain reliable phenotypic data, the per-entry cost of phenotyping in maize would be at least $75. In contrast, the cost of SNP genotyping ranges from about 3 cents to 15 cents per data point (Schaeffer, 2006; Ha et al., 2007; Hyten et al., 2008, E. Buckler, personal communication, 2008; G.J. Muehlbauer, personal communication, 2008), where one data point corresponds to one plant sample genotyped for one marker locus and where the lower costs per data point are for larger numbers of SNP markers assayed at once (e.g., 1536 SNP markers). Even if the cost per data point for SNP markers remains at 15 cents, the per-entry cost of SNP genotyping for, say, 256 SNP markers would be $38.

The use of standard SNP chips for all F_2 or backcross populations in a breeding program may lower the cost per data point due to an economy of scale. By this we mean that instead of screening the parental inbreds of each population for polymorphic markers and using only the polymorphic markers in marker-based selection, the same set of 256, 384, or 512 SNP markers on one or more standard SNP chips may be used for all breeding applications. Not all SNP markers on a standard SNP chip will be polymorphic for a given population, and the presence of uninformative SNP markers will increase the price per data point (Hyten et al., 2008). In this situation, the price per entry becomes more meaningful than the price per data point. The current price per entry is about $40 to $60

for a SNP chip with 1536 markers, with lower per-entry costs for larger numbers of entries (Hyten et al., 2008; E. Buckler, personal communication, 2008).

The above results regarding the cost of obtaining marker data are consistent with informal discussions I have had with several commercial maize breeders who indicated that in their breeding programs, the cost of genotyping is already less than the cost of phenotyping. In contrast, the large scale that is required to lower the per-sample costs of genotyping indicates that the total cost of genotyping will remain high. This high total cost will be a challenge for noncommercial breeding programs and for minor crops for which research investment has been low. Nevertheless, although marker costs vary among breeding programs, the higher cost of phenotyping than of genotyping in some programs and the higher gains per unit time with MARS than with phenotypic selection suggest that marker-based selection is a resource-efficient breeding methodology for complex traits.

FUTURE APPLICATIONS

The increasing availability of cheap and abundant molecular markers suggests that markers should no longer be viewed as an add-on to a breeding program (Bernardo and Yu, 2007). I speculate that future applications of molecular markers in plant breeding will have three interrelated foci: (i) exploiting marker and phenotypic data routinely generated in a breeding program; (ii) marker-based selection before phenotyping; and (iii) marker-based selection without QTL mapping.

Exploiting Marker and Phenotypic Databases

The aggressive use of marker-based selection in a breeding program will eventually lead to large amounts of marker and phenotypic data. Suppose a maize breeder conducts MARS in 30 F_2 populations each year, and that in each population $N = 144$ F_3 families are genotyped at 384 SNP markers and evaluated for testcross performance in, say, six environments. The breeder may initially deem that the sole purpose of the phenotypic data (for $30 \times 144 = 4320$ testcrosses) and SNP data ($30 \times 144 \times 384 = 1.7$ million data points) is to allow marker-based selection for multiple, complex traits within each of the 30 populations. In this context, after the selections have been made in each population, the breeder may deem that the SNP and phenotypic data have served their purpose in a time- and cost-effective manner and that the data have no further use in the breeding program. This approach, however, would not be the wisest use of marker and phenotypic data that are accumulated in the course of marker-based selection. As indicated in the next two sections, the accumulated marker and phenotypic data could be mined for information that may later be used for marker-based selection before phenotyping or without QTL mapping. As larger amounts of marker and phenotypic data are accumulated over time, the estimates or predictions of marker effects would become more refined and would make marker-based selection more effective.

In addition to F_2 or backcross populations undergoing marker-based selection, experimental inbreds or hybrids are useful for finding marker-trait associations (Parisseaux and Bernardo, 2004). Experimental inbreds (e.g., soybean) or hybrids (e.g., maize) are evaluated in multi-environment trials for their potential as released cultivars. The resulting yield-trial databases are a rich resource for finding marker-trait associations both within and across different genetic backgrounds. Furthermore, the use of many environments in cultivar trials permits the sampling of a large set of genotype × environment interactions. For example, an experimental maize hybrid is typically evaluated in 20 environments, and those that are eventually released as cultivars are evaluated in up to 1500 location-year combinations (Smith et al., 1999). Marker-trait associations may then be identified either for a wide range of environments or for a specific subset of environments.

Combining marker and phenotypic data from multiple F_2 or backcross populations as well as experimental cultivars leads to (i) highly unbalanced data sets and (ii) strong population structures. Methods for mining marker and phenotypic databases should therefore account for these two complicating factors. Mixed-model methods have long been used to handle large, unbalanced data sets as well as account for pedigree relationships (Henderson, 1984), and these methods have been successfully extended and used for finding marker-trait associations (Kennedy et al., 1992; Parisseaux and Bernardo, 2004; Zhang et al., 2005; Yu et al., 2006). Among 6921 maize single-crosses with different genetic backgrounds, a major QTL for resistance to common smut [*Ustilago maydis* (DC.) Cda.] was detected on chromosome 8 via mixed-model analysis (Parisseaux and Bernardo, 2004). Marker-trait associations for other traits in maize, including grain moisture, have been identified through an identity-by-descent approach (Zhang et al., 2005) that traces the inheritance of QTL through a pedigree (Graham and Podlich, 2006). Because the identity-by-descent of markers is considered, the marker-trait associations detected may have greater repeatability across different inbreds or populations.

Marker-Based Selection before Phenotyping

In the MARS scheme described so far, a set of progenies are both genotyped and phenotyped and an ad hoc *index* is used in multiple cycles of selection in the same cross. An *ad hoc index* refers to a marker selection index constructed from marker and phenotypic data for a given cross and used for selection in the same cross. Can an effective *prior index*, however, be constructed from prior marker and phenotypic data on germplasm related to the cross at hand? If so, the prior index can be used for marker-based selection of the

best individuals in the cross at hand, before obtaining any phenotypic data on the cross. Prior indices would likely be most useful in the early stages of selection, during which large numbers of progenies need to be quickly evaluated.

Two studies suggest the feasibility of a prior index for marker-based selection without phenotyping the population at hand. In sweet corn, families selected based on a prior index yielded 4% higher than families selected based on field evaluations (Johnson, 2001). In maize, divergent marker-based selection with a prior index led to a 25 g H_2O kg^{-1} difference in grain moisture between the high and low selections (Eathington et al., 2007). Details are unavailable on how the prior indices were calculated in these two studies. Comparisons of the responses to an ad hoc index vs. a prior index are likewise unavailable. Nevertheless, prior indices will become increasingly attractive as the cost of genotyping decreases and the cost of phenotyping increases. Furthermore, prior and ad hoc indices may complement each other. Marker-based selection among individual F_2 plants may first be performed with a prior index. If the cross proves superior and warrants further selection, F_2-derived progenies could be phenotyped and an ad hoc index could be developed and used for further selection in the cross.

As previously mentioned, plant breeding for complex traits involves three phases: (i) creating genetic variation mainly by crossing good by good, (ii) selecting the best progenies in the cross, and (iii) synthesizing the best progenies into a new and improved cultivar (Dudley and Moll, 1969). In the literature, most of the applications of molecular markers for improving complex traits have focused on the middle phase of selecting within a population. Little research has been published on the usefulness of marker-trait associations for choosing parents of populations in inbred development for several complex traits, or for choosing the parents of a single-cross cultivar for hybrid crops. Best linear unbiased prediction (BLUP) based only on phenotypic and pedigree data has been useful for choosing parents to maximize the mean performance of F_2 or backcross populations and for predicting the performance of single crosses before field testing (Bernardo, 1996). Simulation studies have indicated that because BLUP is effective for predicting mean performance, marker information adds little to the prediction of single-cross performance (Bernardo, 2001).

On the other hand, pedigree-based BLUP is not useful in within-population selection because individuals within the same F_2 or backcross population have the same pedigree (Bernardo, 2002). Marker applications reported in the literature have therefore already focused on the phase of breeding where markers are potentially most useful. Furthermore, results for grain moisture in maize suggest that a prior index is useful for predicting the performance of individuals within a cross (Graham and Podlich, 2006). These results suggest that prior indices may be used to create virtual F_2 or backcross populations that can then serve as a basis for choosing which F_2 or backcross populations to create in a breeding program. This topic deserves much further study.

Marker-Based Selection without QTL Mapping

As previously mentioned, F_2 enrichment can target up to 9 to 12 unlinked QTL whereas MARS can target a larger number of marker loci (e.g., 30), with the understanding that recombinant inbreds eventually developed from MARS might not be fixed for the favorable allele at all target loci. One may argue that MARS does not truly entail QTL mapping because the procedure does not require mapping the specific positions of QTL (e.g., as in interval mapping) relative to the markers with significant effects. A third approach, *genomewide selection* (Meuwissen et al., 2001), focuses purely on prediction of performance and avoids QTL mapping altogether (Fig. 1).

To illustrate, suppose an F_2 population is genotyped with 512 SNP markers. Further suppose that in MARS, significance tests at $P = 0.20–0.40$ subsequently identify 30 SNP markers associated with grain yield. The multiple-regression coefficients for these 30 markers are then used as weights (b_i) in calculating marker scores (M_j) in MARS. In genomewide selection, however, the joint effects on a quantitative trait of all 512 SNP markers are fitted as random effects in a linear model. Trait values are still predicted from a weighted index with the form $M_j = \Sigma\, b_i\, X_{ij}$, but a b_i is calculated for each of the 512 markers instead of only for those markers that were found significant (30 in this example) in MARS (Bernardo and Yu, 2007). Because marker effects are fitted as random effects, the number of markers used can exceed the population size. Whereas MARS involves a two-step process of model selection (i.e., which markers to use) and model estimation (i.e., effect of each significant marker), genomewide selection avoids model selection altogether.

Simulation studies have indicated that across different numbers of QTL (20, 40, and 100) and levels of h^2, responses to genomewide selection were 18 to 43% larger than the corresponding responses to MARS (Bernardo and Yu, 2007). Genomewide selection was found most useful for complex traits controlled by many QTL and with a low h^2. Furthermore, genomewide selection can be implemented in the same way as MARS (i.e., three cycles of marker-based selection in Hawaii or Puerto Rico) with the obvious exception of having to genotype all individuals with a larger number of markers. Empirical studies comparing genomewide selection and MARS in maize are underway (R. Bernardo and H.G. Jung, in progress). Genomewide selection methods that utilize a prior index rather than an ad hoc index need further study.

In addition to genomewide selection, *machine learning* methods represent a potentially useful class of procedures

for maximizing the power of molecular markers to predict performance. Some plant breeders may already have unknowingly used machine learning methods in chemometric analysis, e.g., finding near-infrared reflectance spectroscopy (NIRS) calibrations for nondestructive measurement of chemical composition of seeds or other plant parts. Breeders who have been involved in calibrating NIRS machines know that a representative, training sample is needed to develop calibrations. In the same way, machine learning methods as applied to marker-based selection would focus on finding rules or patterns in massive sets of phenotypic and marker data. Machine learning methods such as *artificial neural networks* have been used to map disease loci in humans (Lucek and Ott, 1997), whereas *support vector machine regression* has been used in maize (Maenhout et al., 2007). Results indicated that support vector machine regression and BLUP based on phenotypic and pedigree data (Bernardo, 1996) were equally good at predicting maize single-cross performance. The usefulness of machine learning methods for predicting performance based on prior marker data and prior phenotypic data needs further study. Machine learning methods may be particularly useful in accounting for epistatic interactions (Thornton-Wells et al., 2004) that may be important in the context of long-term selection for certain traits (Carlborg et al., 2006; Dudley, 2008).

In a previous Perspectives article, I concluded that because of the difficulty in estimating the joint effects of many QTL from finite data sets, knowing the number and location of the QTL themselves has little value in selecting for a quantitative trait (Bernardo, 2001). Predictive, black-box methodologies, such as genomewide selection, ignore information on the number and location of QTL and focus on the genetic improvement of quantitative traits rather than on understanding their genetic basis. The usefulness of marker-based selection procedures that focus on predicting performance indicates that markers can be used simply as a selectable tool to improve a complex trait, without a clear understanding of the underlying genetics of the trait. The promise of predictive methods such as genomewide selection, however, does not imply that QTL discovery should no longer be done. Rather, the data used in genomewide selection can also be used to map QTL. Any markers found to have large, highly significant effects can subsequently be exploited by introgressing such major QTL into other germplasm.

SUMMARY AND CONCLUSIONS

The main points presented in this article were as follows

1. For complex traits, QTL mapping is (too) routine but marker-based selection is not.
2. At the outset, one needs to determine if he or she is interested primarily in gene discovery or in selection to improve a complex trait.
3. The germplasm, procedures, stringency, and resolution required for QTL mapping depend on how the results of QTL mapping will be exploited.
4. For traits controlled by few QTL, an effective strategy is to find major QTL in unique germplasm and to introgress these QTL in breeding germplasm.
5. For traits controlled by several QTL, an effective strategy is to select for carrier F_2 individuals and develop recombinant inbreds with most, if not all, of the target QTL alleles.
6. For traits controlled by many QTL, the results of QTL mapping are often inconsistent.
7. For single or multiple traits controlled by many QTL, an effective strategy is to increase the frequency of favorable marker alleles via cyclical marker-based selection.
8. For single or multiple traits controlled by many QTL, prediction of performance based on multiple markers is more effective than pyramiding specific QTL alleles.
9. In major commercial breeding programs, costs of genotyping are now lower than the costs of phenotyping.
10. Future applications will focus on predictive methodologies for marker-based selection before phenotyping and for marker-based selection without QTL mapping.

In conclusion, we have learned in the last 20 years that finding QTL for complex traits is easy but exploiting these QTL in selection is more difficult. Gains per cycle are not necessarily greater with marker-based selection than with phenotypic selection, but markers can increase the gain per year and per unit cost. As marker data become more readily available than phenotypic data, plant-breeding decisions will become more genotype-driven than phenotype-driven. Plant breeders then would need to design marker-based breeding schemes that consider both the routine availability of marker data and the continuing challenges in obtaining good phenotypic information.

Acknowledgments

I thank the following colleagues for engaging me in helpful discussions or reviewing an earlier draft of this manuscript: James A. Anderson, Jan Eric Backlund, William D. Beavis, Todd Piper, Bruno Poupard, Kevin P. Smith, and Jianming Yu.

References

Allard, R.W. 1960. Principles of plant breeding. John Wiley & Sons, New York.

Anderson, J.A., S. Chao, and S. Liu. 2008. Molecular breeding using a major QTL for Fusarium head blight resistance in wheat. Crop Sci. 47(S3):S112–S119.

Anderson, J.A., R.W. Stack, S. Liu, B.L. Waldron, A.D. Fjeld, C. Coyne, B. Moreno-Sevilla, J. Mitchell Fetch, Q.J. Song, P.B. Cregan, and R.C.

Frohberg. 2001. DNA markers for Fusarium head blight resistance QTLs in two wheat populations. Theor. Appl. Genet. 102:1164–1168.

Basten, C.J., B.S. Weir, and Z.-B. Zeng. 1994. QTL Cartographer. Dep. of Statistics, North Carolina State Univ., Raleigh, NC.

Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 250–266. In Proc. Corn Sorghum Ind. Res. Conf., 7–8 Dec. 1994, Chicago, IL. Am. Seed Trade Assoc., Washington, DC.

Beckmann, J.S., and M. Soller. 1986. Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111–124.

Bernardo, R. 1996. Best linear unbiased prediction of maize single-cross performance. Crop Sci. 36:50–56.

Bernardo, R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci. 41:1–4.

Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Bernardo, R., and A. Charcosset. 2006. Usefulness of gene information in marker-assisted recurrent selection: A simulation appraisal. Crop Sci. 46:614–621.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082–1090.

Bhattramakki, D., and A. Rafalski. 2001. Discovery and application of single nucleotide polymorphism markers in plants. p. 179–191 *In* R.J. Henry (ed.) Plant genotyping: The DNA fingerprinting of plants. CABI Publishing, Oxford, UK.

Blanc, G., A. Charcosset, B. Mangin, A. Gallais, and L. Moreau. 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. Theor. Appl. Genet. 113:206–224.

Bonnett, D.G., G.J. Rebetzke, and W. Spielmeyer. 2005. Strategies for efficient implementation of molecular markers in wheat breeding. Mol. Breed. 15:75–85.

Breseghello, F., and M.E. Sorrells. 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. 46:1323–1330.

Buckler, E. 2007. TASSEL: Trait Analysis by aSSociation, Evolution, and Linkage. User manual. http://www.maizegenetics.net/tassel; verified 10 July 2008.

Burrow, M.D., and T.K. Blake. 1998. Molecular tools for the study of complex traits. p. 13–29 *In* A.H. Paterson (ed.) Molecular dissection of complex traits. CRC Press, Boca Raton, FL.

Cahill, D.J., and D.H. Schmidt. 2004. Use of marker assisted selection in a product development breeding program. Proc. 4th Int. Crop. Sci. Cong., 26 Sep.–1 Oct. 2004, Brisbane, Australia. http://www.cropscience.org.au/icsc2004; verified 10 July 2008.

Campbell, B.T., P.S. Baenziger, K.S. Gill, K.M. Eskridge, H. Budak, M. Erayman, I. Dweikat, and Y. Yen. 2003. Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. Crop Sci. 43:1493–1505.

Campbell, K.A.G., and P.E. Lipps. 1998. Allocation of resources: Sources of variation in Fusarium head blight screening nurseries. Phytopathology 88:1078–1086.

Carlborg, Ö., L. Jacobsson, P. Åhgren, P. Siegel, and L. Andersson. 2006. Epistasis and the release of genetic variation during long-term selection. Nat. Genet. 38:418–420.

Chao, S., J. Anderson, K. Glover, and K. Smith. 2006. Use of high throughput marker technologies for marker-assisted breeding in wheat and barley. Proc. Plant Anim. Genomes XIV Conf., 14–18 Jan. 2006, San Diego, CA.

Charcosset, A., M. Causse, L. Moreau, and A. Gallais. 1994. Investigation into the effect of genetic background on QTL expression using three recombinant inbred lines (RIL) populations. p. 75–84 *In* J.W. van Ooijen and J. Jansen (ed.) Biometrics in plant breeding: Applications of molecular markers. CRPO-DLO, Wageningen, The Netherlands.

Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053–1067.

Concibido, V.C., R.L. Denny, S.R. Boutin, R. Hautea, J.H. Orf, and N.D. Young. 1994. DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). Crop Sci. 34:240–246.

Concibido, V.C., B.W. Diers, and P.R. Arelli. 2004. A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci. 44:1121–1131.

Doebley, J. 2006. Unfallen grains: How ancient farmers turned weeds into crops. Science 312:1318–1319.

Dudley, J.W. 1993. Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. Crop Sci. 33:660–668.

Dudley, J.W. 2008. Epistatic interactions in crosses of Illinois high oil × Illinois low oil and of Illinois high protein × Illinois low protein corn strains. Crop Sci. 48:59–68.

Dudley, J.W., and R.H. Moll. 1969. Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci. 9:257–262.

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular markers in a commercial breeding program. Crop Sci. 47(S3):S154–S163.

Edwards, M., and L. Johnson. 1994. RFLPs for rapid recurrent selection. p. 33–40. In Analysis of molecular marker data. Joint Plant Breed. Symp. Ser., Am. Soc. Hort. Sci., CSSA, Madison, WI.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler, IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357–374.

Frary, A., T.C. Nesbitt, A. Frary, S. Grandillo, E. van der Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert, and S.D. Tanksley. 2000. fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88.

Fuentes-Granados, R.G., H.R. Mickelson, R.H. Busch, R. Dill-Macky, C.K. Evans, W.G. Thompson, J.V. Wiersma, W. Xie, Y. Dong, and J.A. Anderson. 2005. Resource allocation and cultivar stability in breeding for Fusarium head blight resistance in spring wheat. Crop Sci. 45:1965–1972.

Gale, M.D., and K.M. Devos. 1998. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA 95:1971–1974.

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schüler. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. 13:93–102.

Graham, G.I., and D. Podlich. 2006. The use of molecular markers in Pioneer's corn breeding program. ASA-CSSA-SSSA annual meetings, 12–16 Nov. 2006, Indianapolis, IN.

Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harb. Symp. Quant. Biol. 39:439–446.

Ha, B.-K., R.S. Hussey, and H.R. Boerma. 2007. Development of SNP assays for marker-assisted selection of two Southern

root-knot nematode resistance QTL in soybean. Crop Sci. 47(S2):S73–S82.

Haley, C.S., and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324.

Hästbacka, J., A. de la Chapelle, I. Kaitila, P. Sistonen, W. Weaver, and E. Lander. 1992. Linkage disequilibrium mapping in isolated founder populations: Diastrophic displasia in Finland. Nat. Genet. 2:204–211.

Hauge, B.M., M.L. Wang, J.D. Parsons, and L.D. Parnell. inventors; Monsanto Company, assignee. 2006. Nucleic acid molecules and other molecules associated with soybean cyst nematode resistance. U.S. patent 7154021.

Henderson, C.R. 1984. Applications of linear models in animal breeding. Univ. of Guelph, Canada.

Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in breeding programs- challenges and opportunities. Proc. 4th Intl. Crop Sci. Cong., 26 Sep.–1 Oct. 2004, Brisbane, Australia. http://www.cropscience.org.au/icsc2004/symposia/3/4/203_hollandjb.htm; verified 10 July 2008.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10:156–161.

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

Howes, N.K., S.M. Woods, and T.F. Townley-Smith. 1998. Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs. Euphytica 100:225–230.

Hyten, D.L., Q. Song, I.-Y. Choi, J.E. Specht, L.K. Matukumalli, R.L. Nelson, R.C. Shoemaker, N.D. Young, and P.B. Cregan. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116:945–952.

Jannink, J.L., and R.C. Jansen. 2000. The diallel mating design for mapping interacting QTLs. p. 81–88. *In* Quantitative genetics and breeding methods: The way ahead. INRA, Paris.

Jenkins, S., and N. Gibson. 2002. High-throughput SNP genotyping. Comp. Funct. Genom. 3:57–66.

Johnson, L. 2001. Marker assisted sweet corn breeding: A model for specialty crops. p. 25–30 *In* Proc. 56th Annu. Corn Sorghum Ind. Res. Conf. 5–7 Dec. 2001, Chicago, IL. Am. Seed Trade Assoc., Washington, D.C.

Johnson, R. 2004. Marker-assisted selection. Plant Breed. Rev. 24:293–309.

Kearsey, M.J., and A.G.L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity 80:137–142.

Kearsey, M.J., and V. Hyne. 1994. QTL analysis: A simple 'marker-regression' approach. Theor. Appl. Genet. 89:698–702.

Kennedy, B.W., M. Quinton, and J.A.M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70:2000–2012.

Kilian, A., E. Huttner, P. Wenzl, D. Jaccoud, J. Carling, V. Caig, M. Evers, K. Heller-Uszynska, G. Uszynski, C. Cayla, S. Patarapuwadol, L. Xia, S. Yang, and B. Thomson. 2005. The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. p. 443–461. *In* R. Tuberosa et al. (ed.) Proc. Int. Cong. "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution," 27–31 May 2003, Bologna, Italy.

Koebner, R. 2003. MAS in cereals: Green for maize, amber for rice, still red for wheat and barley. *In* Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding? 17–18 Oct. 2003, Turin, Italy. FAO, Rome. http://www.fao.org/biotech/docs/Koebner.pdf; verified 10 July 2008.

Kuchel, H., R. Fox, J. Reinheimer, L. Mosionek, N. Willey, H. Bariana, and S. Jefferies. 2007. The successful application of a marker-assisted wheat breeding strategy. Mol. Breed. 20:295–308.

Kurata, N., G. Moore, Y. Nagamura, T. Foote, M. Yano, Y. Minobe, and M. Gale. 1994. Conservation of genome structure between rice and wheat. Biotechnology 12:276–278.

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

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Lazzeroni, L.C. 1997. Empirical linkage-disequilibrium mapping. Am. J. Hum. Genet. 62:159–170.

Leon, A.J., M. Lee, and F.H. Andrade. 2001. Quantitative trait loci for growing degree days to flowering and photoperiod response in sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 102:497–503.

Lincoln, S.E., M.J. Daly, and E.S. Lander. 1993. Mapping genes controlling quantitative traits using MAPMAKET/QTL version 1.1: A tutorial and reference manual. Whitehead Inst., Cambridge, MA.

Liu, S., X. Zhang, M.O. Pumphrey, R.W. Stack, B.S. Gill, and J.A. Anderson. 2006. Complex microcolinearity among wheat, rice, and barley revealed by fine mapping of the genomic region harboring a major QTL for resistance to Fusarium head blight in wheat. Funct. Integr. Genomics 6:83–89.

Lu, H., J. Romero-Severson, and R. Bernardo. 2003a. Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor. Appl. Genet. 107:494–502.

Lu, H.-J., H.W. Ohm, and R. Bernardo. 2003b. Mapping QTL for popping expansion volume in popcorn with simple sequence repeat markers. Theor. Appl. Genet. 106:423–427.

Lucek, P.R., and J. Ott. 1997. Neural network analysis of complex traits. Genet. Epid. 14:1101–1106.

Mackay, T.F.C. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303–339.

Maenhout, S., B. De Baets, G. Haesaert, and E. Van Bockstaele. 2007. Support vector machine regression for the prediction of maize hybrid performance. Theor. Appl. Genet. 115:1003–1013.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Melchinger, A.E., H.F. Utz, and C.C. Schön. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403.

Monforte, A.J., and S.D. Tanksley. 2000. Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: Breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor. Appl. Genet. 100:471–479.

Nelson, J.C. 1997. QGENE: Software for marker-based genomic analysis and breeding. Mol. Breed. 3:239–245.

Parisseaux, B., and R. Bernardo. 2004. *In silico* mapping of quantitative trait loci in maize. Theor. Appl. Genet. 109:508–514.

Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir, H.D.

Rabinowitch, S.E. Lincoln, E.S. Lander, and S.D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. Genetics 127:181–197.

Paterson, A.H., Y.-R. Lin, Z. Li, K.F. Shertz, J.F. Doebley, S.R.M. Pinson, S.-C. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718.

Paterson, A.H., Y. Saranga, M. Menz, C.-X. Jiang, and R. Wright. 2003. QTL analysis of genotype × environment interactions affecting cotton fiber quality. Theor. Appl. Genet. 106:384–396.

Pflieger, S., V. Lefebvre, and M. Causse. 2001. The candidate gene approach in plant genetics: A review. Mol. Breed. 7:275–291.

Price, A.H. 2006. Believe it or not, QTLs are accurate! Trends Plant Sci. 11:213–216.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pumphrey, M.O., R. Bernardo, and J.A. Anderson. 2007. Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Sci. 47:200–206.

Rebai, A., B. Goffinet, B. Mangin, and D. Perret. 1994. QTL detection with diallel schemes. p. 170–177. *In* J.W. van Ooijen and J. Jansen (ed.) Biometrics in plant breeding: Applications of molecular markers. CRPO-DLO, Wageningen, The Netherlands.

Reyna, N., and C.H. Sneller. 2001. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. Crop Sci. 41:1317–1321.

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223.

Schön, C.C., H.F. Utz, S. Groh, B. Truberg, S. Openshaw, and A.E. Melchinger. 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498.

Smith, O.S., K. Hoard, F. Shaw, and R. Shaw. 1999. Prediction of single-cross performance. p. 277–285. *In* J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, Madison, WI.

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. 3:739–744.

Syvänen, A.C. 2005. Toward genome-wide SNP genotyping. Nat. Genet. 37:s5–s10.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: New tools for an old science. BioTechnol. 7:257–264.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler, IV. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28:286–289.

Thornton-Wells, T.A., J.H. Moore, and J.L. Haines. 2004. Genetics, statistics and human disease: Analytical retooling for complexity. Trends Genet. 20:640–647.

Utz, H.F., and A.E. Melchinger. 1996. PLABQTL: A program for composite interval mapping of QTL. Inst. Plant Breed., Seed Sci., Pop. Genet., Univ. of Hohenheim, Stuttgart, Germany.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

Waldron, B.L., B. Moreno-Sevilla, J.A. Anderson, R.W. Stack, and R.C. Frohberg. 1999. RFLP mapping of QTL for *Fusarium* head blight resistance in wheat. Crop Sci. 39:805–811.

Wang, J., S.C. Chapman, D.G. Bonnett, G.J. Rebetzke, and J. Crouch. 2007. Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci. 47:582–588.

Wang, Y.Z., and J.D. Miller. 1988. Screening techniques and sources of resistance to *Fusarium* head blight. p. 239–250. *In* A.R. Klatt (ed.) Wheat production constraints in tropical environments. CIMMYT, Mexico City.

Webb, D.M., B.M. Baltazar, A.P. Rao-Arelli, J. Schupp, K. Clayton, P. Keim, and W.D. Beavis. 1995. Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. Theor. Appl. Genet. 91:574–581.

Weyhrich, R.A., K.R. Lamkey, and A.R. Hallauer. 1998. Responses to seven methods of recurrent selection in the BS11 maize population. Crop Sci. 38:308–321.

Whittaker, J.C., R. Thompson, and P.M. Visscher. 1996. On the mapping of QTL by regression of phenotypes on marker-type. Heredity 77:23–32.

Wong, C.K., and R. Bernardo. 2008. Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. Theor. Appl. Genet. 116:815–824.

Wright, A.J., and R.P. Mowers. 1994. Multiple regression for molecular-marker, quantitative trait data from large F_2 populations. Theor. Appl. Genet. 89:305–312.

Xu, S. 2003. Theoretical basis of the Beavis effect. Genetics 165:2259–2268.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208.

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.

Zhang, Y.-M., Y. Mao, C. Xie, H. Smith, L. Luo, and S. Xu. 2005. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics 169:2267–2275.

Zhu, H., G. Briceno, R. Dovel, P.M. Hayes, B.H. Liu, C.T. Liu, and S.E. Ullrich. 1999. Molecular breeding for grain yield in barley: An evaluation of QTL effects in a spring barley cross. Theor. Appl. Genet. 98:772–779.

Zhu, S., and H.F. Kaeppler. 2003. Identification of quantitative trait loci for resistance to crown rust in oat line MAM17–5. Crop Sci. 43:358–366.

Zhuang, J.-Y., H.-X. Lin, J. Lu, H.-R. Qian, S. Hittalmani, N. Huang, and K.-L. Zheng. 1997. Analysis of QTL × environment interaction for yield components and plant height in rice. Theor. Appl. Genet. 95:799–808.