Theor Appl Genet (1996) 92:191–203

S. D. Tanksley · J. C. Nelson

Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines

Received: 5 July 1995 / Accepted: 4 August 1995

Abstract Advanced backcross QTL analysis is proposed as a method of combining QTL analysis with variety development. It is tailored for the discovery and transfer of valuable QTL alleles from unadapted donor lines (e.g., land races, wild species) into established elite inbred lines. Following this strategy, QTL analysis is delayed until the BC2 or BC3 generation and, during the development of these populations, negative selection is exercised to reduce the frequency of deleterious donor alleles. Simulations suggest that advanced backcross QTL analysis will be effective in detecting additive, dominant, partially dominant, or overdominant QTLs. Epistatic QTLs or QTLs with gene actions ranging from recessive to additive will be detected with less power than in selfing generations. QTL-NILs can be derived from advanced backcross populations in one or two additional generations and utilized to verify QTL activity. These same QTL-NILs also represent commercial inbreds improved (over the original recurrent inbred line) for one or more quantitative traits. The time lapse from QTL discovery to construction and testing of improved QTL-NILs is minimal (1–2 years). If successfully employed, advanced backcross QTL analysis can open the door to exploiting unadapted and exotic germplasm for the quantitative trait improvement of a number of crop plants.

Key words Molecular markers · Introgression · Plant breeding · Quantitative trait loci

Introduction

Most traits of agronomic importance, including yield, nutritional quality and stress tolerance, are quantitatively inherited (Allard 1960; Hallauer and Miranda 1988). The ability to manipulate genes responsible for quantitative traits is a prerequisite for sustained improvement of crop plants. In the past 10 years there have been numerous reports on the use of molecular marker-based methods to detect, map and characterize the loci responsible for quantitative traits in crops (Paterson et al. 1988; Keim et al. 1990; Fatokun et al. 1992; Stuber et al. 1992; Anderson et al. 1993a; Hayes et al. 1993; Wang et al. 1994). Despite these successes, there are few examples where molecular marker techniques have led to the creation of new crop varieties enhanced for one or more quantitative traits.

Two factors may be contributing to the less-than-expected impact of marker-based QTL analysis on the development of varieties with enhanced quantitative traits.

(1) QTL discovery and variety development are currently separate processes

QTL analysis, as applied to plant breeding, typically involves a two-step process. The first step is QTL discovery. Parental lines are identified which differ for one or more quantitative traits of agronomic importance (e.g., yield, quality). The parents are hybridized and a segregating population(s) is created in which markers can be used to identify linked QTLs. The second step is to utilize knowledge of QTL map locations to create a superior variety.

The most common populations for QTL analysis are F2 or F3 families, backcrosses to one or both parents or recombinant inbred lines (e.g., F7 single-seed descent lines). Each of these populations has different strengths with regard to QTL analysis. However, with respect to plant breeding, they all share a common weakness: once potentially valuable QTLs are discovered, considerable backcrossing and/or intercrossing is likely to be required for the development of commercial varieties/inbreds.

Communicated by G. Wenzel

S. D. Tanksley (✉) · J. C. Nelson
Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA

Separating QTL discovery and variety development into discrete and sequential steps not only increases the time required for new variety development, but also (for other reasons to be discussed later) reduces the probability of successfully using QTL information to create a superior crop variety.

(2) Most breeding-related QTL studies are targeted toward manipulating quantitative trait variation already existing within elite germplasm

In every breeding program there exists a primary adapted pool of elite germplasm from which most new variety selections are made. Breeders have observed that working within this adapted gene pool normally increases the probability of making successful selections. Excluded from this pool are a wide range of genetic resources, including primitive varieties, land races and wild species, all of which are inferior to commercial varieties in one or more respects.

Even when one is working within the primary gene pool, improvement of quantitative traits is often a difficult and time-consuming task. There has been an implicit expectation that marker-based QTL analysis will make it easier and faster for breeders to manipulate these traits (Soller and Beckman 1983; Tanksley 1983). Theoretically, marker-aided selection can lead to the accumulation of valuable QTLs into new varieties within elite germplasm. However, in reality there are several practical problems with this strategy. Frequently, elite germplasm (especially in self-pollinated crops) has reduced levels of genetic variation making it difficult to find the necessary polymorphism with the molecular markers required for QTL analysis (Helentjaris et al. 1985; Miller and Tanksley 1990; Wang et al. 1992; Anderson et al. 1993 b). A second problem is that focusing on elite germplasm means that molecular markers are being used to manipulate the same alleles that breeders have been manipulating for many years through classical breeding procedures. While the molecular marker approach may in some instances be faster than traditional selection, it is often more expensive and, if applied in this manner, will not result in a general enhancement of the gene pool of crop varieties. Without such enhancement, it is unlikely that molecular markers can contribute to the long-term and sustained improvement of quantitative characters in crop plants.

Need for an alternative approach

Before QTL analysis will have a major impact on the genetic improvement of crop plants it seems likely that the situation described above will have to change. Specifically, QTL discovery and variety development must be integrated into a single process. In addition, QTL analysis must be used to identify and selectively introduce new and beneficial alleles into elite crop varieties, thereby broadening the genetic base of the crop species and accelerating the rate of genetic improvement. The topic of this paper is a proposal of how these goals might be achieved through a new strategy which we refer to as "advanced backcross QTL analysis".

Background

All crop species were originally domesticated from wild plants by humans – a process that inherently reduced genetic variation (Simmonds 1976; Ladizinsky 1985). Intensive breeding of crop varieties by modern science has further narrowed the gene pool in many crops. This problem is especially acute in self-pollinated crops where the level of genetic variation in cultivated varieties is often a small fraction of that available in nature (Miller and Tanksley 1990; Wang et al. 1992). The limited genetic variation among modern crop varieties makes them more susceptible to disease epidemics. For example, Southern corn leaf blight destroyed a large portion of the U.S. corn crop in 1971 due to the widespread use of a single source of cytoplasmic male sterility in hybrid maize (Ullstrup 1978). Reduced genetic variation is likely to have another, more subtle, effect: a slower rate of crop improvement by plant breeders. The lower the genetic variation in breeding populations, the less likely breeders are to identify new and useful combinations of genes.

Unadapted and wild germplasm: an under-utilized reservoir of genetic variation

There is no shortage of genetic variation in nature. The wild ancestors and related land races of most crop plants can still be found in their natural habitats and national and international centers have been established to collect and maintain this material. For most major crop plants, germplasm banks contain tens of thousands of individual seed accessions. However, only a few of these accessions have actually contributed to the development of new elite varieties. Nevertheless, we maintain these collections with the tacit belief that the alleles contained in these wild and unadapted accessions will someday fuel crop plant improvement. The unfortunate reality is that, for the most part, we have been unable to exploit the majority of the genetic potential stored in germplasm repositories.

Problems with using unadapted germplasm to improve quantitative traits

Linkage drag

The productivity, uniformity and quality of modern, elite cultivars is based on decades of selection by breeders. As this elite germplasm becomes more differentiated from primitive cultivars, the option of utilizing

wild or unadapted germplasm, which is likely to contribute many undesirable characters and require intensive selection to recover acceptable cultivars, becomes less and less attractive. For this reason, unadapted or wild species germplasm has been used mainly as a source of major genes for disease and insect resistance which can be incorporated via a backcross breeding scheme. Even in these cases, the transfer of the resistance genes has often been fraught with problems associated with linkage drag. Theoretical and empirical studies have shown that, even after 20 or more years of traditional breeding, a single gene transferred from a wild species will be associated with enough linked chromosomal DNA to contain more than 100 other, potentially undesirable genes (Young and Tanksley 1989; unpublished results). Breeding to introduce polygenic characters from wild germplasm, where epistasis becomes a problem and linkage drag is compounded, has been generally avoided.

Molecular linkage maps provide a method to reduce the problems with linkage drag by allowing selection for individuals containing recombinant chromosomes which break linkage drag. It has been estimated that the use of molecular maps can reduce linkage drag at least tenfold in a fraction of the time needed for traditional breeding (Tanksley et al. 1989). In tomato, the molecular linkage map has been used to identify varieties carrying minimal linkage drag around several disease resistance genes transferred from wild species (Young and Tanksley 1989; Messeguer 1991; Zamir et al. 1994)

Most unadapted germplasm has inferior phenotypes with respect to key quantitative traits

While it is relatively straightforward to identify wild accessions that contain genes for resistance to pathogens, it is difficult to identify accessions that are likely to contain genes for the improvement of quantitative traits such as yield. In this respect, unadapted germplasm is almost always inferior to elite varieties.

Report of occasional transgressive variation for yield in populations derived from crosses of elite lines to wild species suggests that, despite their inferior phenotypes, wild accessions do contain genes that can improve quantitative traits (Frey et al. 1981). Recent studies in tomatoes utilizing molecular markers have confirmed this hypothesis and have demonstrated that QTLs isolated from wild accessions, can substantially improve the phenotype of commercial tomato varieties for most quantitative traits (de Vicente and Tanksley 1993; Eshed and Zamir 1994). In these studies, molecular markers were used to screen the genome of the wild species *L. pennellii* for QTLs that modify both botanical and horticultural traits. Altogether more than 20 traits were studied. Regardless of the character, 10–50% of the QTLs detected in the wild species improved the trait of interest, even though the wild species phenotype was inferior to that of the cultivated parent. Specific QTLs for increased yield and soluble solids were subsequently transferred into the cultivated tomato using linked molecular markers (de Vicente and Tanksley 1993; Eshed and Zamir 1994).

The experiments described above demonstrate that positive (with regard to agronomic value) QTLs can be identified and transferred from an unadapted or wild species into elite cultivated varieties using molecular markers. If wild species, regardless of their phenotype, contain alleles that can improve most traits of interest, then the problem of identifying the source of new genes for the genetic improvement of quantitative traits disappears. It is replaced by the task of locating and selectively transferring valuable QTLs from unadapted germplasm into elite crop varieties. Theoretically this can be accomplished via QTL mapping and marker-based selection using the molecular linkage maps that are well-developed for many crop species. However, the genetic design used to identify and transfer useful QTLs from unadapted germplasm may have to differ from that followed by researchers in the past.

Advanced backcross QTL analysis – a strategy to identify and transfer beneficial QTLs from an unadapted line into an elite breeding line

Most QTL studies have utilized populations in which the alleles of both parents occur at a relatively high frequency (e.g., F2, BC1, RI). While such populations (hereafter referred to as balanced populations) are well-suited for QTL mapping, they have significant drawbacks when employed to detect and transfer useful QTLs from unadapted germplasm into elite breeding lines. In this context, the shortcomings of balanced populations are as follows: (1) Undesirable QTL alleles from the unadapted parent occur in high frequency and can seriously reduce (or eliminate altogether) the ability to collect meaningful data on yield and other field performance traits. For example, the presence of shattering and/or sterility genes would make it difficult to measure yield in balanced populations derived from crosses of cultivated wheat or barley with one of their wild relatives. (2) Epistatic interactions are statistically difficult to detect, yet are likely to occur in balanced populations where donor alleles occur in a high frequency. For breeding purposes it is desirable to identify QTLs not requiring epistatic interactions among donor alleles – a goal which is more difficult to achieve in balanced populations. (3) Subtle (and often negative) pleiotropic effects may go unnoticed in balanced populations due to the large genetic and phenotypic variance created by the segregation of donor alleles in high frequency. These pleiotropic effects may not become obvious until the QTL has been backcrossed into an elite line and the genetic variance reduced.

A possible solution to the above problems would be to delay QTL analysis until an advanced generation (e.g., BC2, BC3, etc). Researchers have already demon-

strated that advanced backcross populations can be used to increase the probability of successful selections when breeding with wild germplasm (Bliss 1981; Sullivan and Bliss 1983 a,b). With regards to QTL analysis the potential benefits of advanced backcross populations versus more balanced populations (e.g., F2, BC1, Rl) are: (1) The genotype (and phenotype) of an average individual would be much more similar to the elite parent in an advanced backcross population than in a balanced population, making measurements of yield and other characters more meaningful. (2) By waiting until an advanced generation, one could allow for phenotypic selection to further reduce the frequency of deleterious or undesirable alleles from the donor. Therefore, problems with deleterious characters (e.g., sterility, seed shattering, etc), often encountered in balanced populations derived from wide crosses, should be reduced or eliminated. (3) One would be less likely to detect QTLs with epistatic effects in advanced generations due to the overall lower frequency of donor alleles. Therefore, the effect measured for each QTL in the advanced generation is likely to be a better predictor of the ultimate effect of that QTL when it is transferred to the cultivated background. (4) Few, if any, additional backcross generations will be needed to create nearly isogenic lines with selected QTLs. Such QTL-NIL lines could then be subjected to extensive field testing in replicated trials. (5) Associated deleterious effects due to linkage drag are less likely to be observed in advanced generations since there have been more opportunities for useful meiotic recombination.

The objective of this paper is to explore, through computer simulations, the properties of QTL analysis in advanced backcrosses populations as compared with more traditional balanced populations. Results from the simulations are discussed in the context of using advanced backcross QTL analysis as the basis for a breeding strategy in which valuable QTLs from unadapted germplasm are efficiently discovered and transferred into elite breeding lines.

Materials and methods

Simulations for QTL detection

Simulations were performed on a Macintosh Power PC computer. Genetic maps were simulated as arrays of integers representing map positions in units of 0.1% recombination (where 100% probability of crossover is represented as 1000) and containing subarrays representing chromosomes, separated from the other subarrays by intervals of size 500. Maps were constructed according to specifications of chromosome number, average number of markers per chromosome, average chromosome length, number of QTLs governing the simulated trait, the magnitude of the weights on the different QTLs, and the nature of gene action. For all simulations two QTLs were used, each being allowed to account for half of the phenotype. They were placed on different chromosomes and any modifier genes also included in an experiment were placed on separate chromosomes from each other and from either QTL, in order to avoid confounding of gene action with linkage. Each QTL (specified at random as one of the loci in the map) was given a uniform numerical weight, applied to either the maternal or the paternal allele.

Genotypes were simulated as strings of bits set to either 1 (maternal allele) or 0 (paternal allele). An F_1 diploid individual consisted of a string of 1s and another of 0s. Gametogenesis was done via "random walk" (all randomness being simulated by the computer's pseudo-random number generator) starting at random on one of the strings, beginning with a gamete with the first bit set as on that string, then proceeding along the map, generating a random number between 0 and 1000 at each interval, and crossing over to read from the other string if the specified recombination probability exceeded this number. Either one (backcrossing) or two (selfing) gametes were produced for each parental genotype; for backcrossing, the second gamete consisted of a string of 1s. Manipulations of this kind are described in several sources; two early ones are Bellman and Ahrens (1966) and Fraser and Burnell (1970). In this manner, populations of 200 diploid genotypes were generated. Their phenotypes were evaluated as the sums of the weights placed on the QTLs having the allele that had been specified for weighting, each weight being multiplied by the number of weighted alleles at the locus. Where other than additive types of genetic action were specified, the corresponding modification was made to the evaluation as follows. For full dominance, heterozygotes were given the same value as the weighted homozygote. For overdominance, only the heterozygotes at a QTL contributed to the phenotypic score for that QTL. For epistasis, an unweighted, unlinked marker was specified for each epistatically modified QTL, such that the specified type of gene action of the QTL was phenotypically expressed subject to the number of modifier alleles present at the modifier locus.

All simulations were done from 20 different maps, constructed at random according to the map specification which was four chromosomes, an average of 50 markers per chromosome, and an average of 1200 for the recombination length of a chromosome, corresponding to approximately 100 cM after expression of the individual recombination fractions as cM according to the Kosambi function (Kosambi 1944). Each simulation run from a map consisted of at least five generations: the first one a population of 200 genotypes constructed from an F_1 individual as described above and each succeeding population of 200 generated by selfing or backcrossing each member of the preceding generation to produce one progeny, analogous to the single-seed-descent method of advancing selfed populations. From each map either 50 or 100 simulation runs were made, corresponding to 5000 or 10000 distinct populations.

Statistics were recorded for each population and averaged for each generation. The computation done for most experiments was a regression of phenotype on marker genotype at every locus within 35 cM of a QTL. For backcross populations the marker genotypes were either 0 or 1 and a simple regression was performed to estimate the additive effect of a single allele substitution. In selfed populations the marker genotypes were 0 and 2 for homozygotes and 1 for heterozygotes. In this case the regression also included a dominance term. The resulting R^2, representing the proportion of genetic variance explained, was recorded as a function of the generation and of the distance of the marker from the QTL. The correctness of segregation, weighting, and statistical computations was verified directly or by sampling. For each genetic model, whether additive, dominant, over-dominant, or epistatic, a multiple-regression model could be constructed that accounted for 100% of the genetic variance. Environmental variance was simulated by the addition to each phenotypic score of a N(0,1) random number scaled by 20% of the range of the phenotype.

For tallying segment lengths, each locus on a map was assigned a length calculated as the sum of the half-intervals on either side (or only the proximal side for the terminal locus of a chromosome) expressed in cM. Runs of consecutive loci with the same marker genotype were tabulated for all individuals by marker genotype, number, length, and generation. Average segment length at a generation was then calculated as the sum of products of run counts by the corresponding lengths, divided by the sum of counts.

Simulations of NIL extraction

For these, as for the other simulations, all populations were of size N = 200. Based on a 12-chromosome map with an average of 50 loci

per 100-cM chromosome, a BC2, BC3, or BC4 population was generated and sorted in descending order based on the proportion of recurrent-parent genome (genome ratio). For the 12 top genotypes bearing the donor allele at a preset pair of loci at least 10 cM apart and at all loci between them, a P value was calculated as the probability that after one more backcross the sole remaining donor alleles would be those in this interval. The genotype with the highest P value of these 12 (usually in the top four with respect to genome ratio) was backcrossed and all progeny in the resulting population were accepted as NILs if they were heterozygous in the specified interval, but over no more than 30 cM total, and homozygous for the recurrent parent everywhere else in the map.

In 50 such runs from each of 200 independent maps, the numbers of such heterozygous or homozygous NILs were recorded for each of 12 P-value classes indexed by the negative $\log_e$ of the P-value, from 0 to 11. These subclasses were not treated separately in the summary and did not enter into the experiment. They are mentioned only because in practice the best estimate of the number of genotypes to screen for the desired introgression would take into account the quality of the original NIL candidate, expressed as this P value or a similar parameter. The NIL counts were summed for each introgression length from 10 to 30 cM and the frequencies and expected lengths were calculated for each cumulative class; that is, the frequency (f) of finding a genotype with an introgression of less than a given length (calculated as the cumulative sum of counts divided by the total genotypes in the experiment, N × runs × maps = 2 000 000) and the length expectation within this class, calculated as the sum of products of counts and lengths, divided by the sum of counts. These frequencies were then used to calculate n, the number of plants to genotype in the final population to obtain with probability p, a NIL with an introgression in the specified length range, by the formula $n = \log(1-p)/\log(1-f)$. This n was plotted for a minimal acceptable size of introgression (10 cm) as a function of the backcross generation at which the NIL candidate was selected for subsequent advancement.

The software (unpublished) used for simulations may be obtained from the junior author via electronic mail (jcn5@cornell.edu).

Results

Average size of donor chromosome segments in backcross and selfing generations

The size of intact donor chromosome segments in a segregating population determines the degree of linkage drag (Stam and Zeven 1981). Shorter donor segments not only reduce linkage drag but should also increase the resolution with which QTLs are mapped. The size of intact donor segments is influenced by chromosome length (as measured in centiMorgans) as well as the number of generations in which the donor and recurrent parent genomes are allowed to recombine by meiosis (Hanson 1959; Stam and Zeven 1981). Recombination leading to a reduction in linkage drag can occur only in heterozygous regions of the genome. During backcrossing all regions of the genome containing donor DNA will be heterozygous and any recombination in these regions of the genome will result in a reduction in the size of the donor segment. In contrast, selfed populations contain individuals homozygous for donor segments. Recombination occurring in these homozygous areas will result in no net reduction in linkage drag.

Figure 1 depicts the results from simulations in which the average size of donor segments is plotted against the number of generations for both selfing and backcrossing. The results indicate that for the BC1 and F2 generations the average length of donor segments is the same (approximately 40 cM). However, after the first generation, the length of donor segments decreases rapidly with additional backcrosses as compared with additional selfing generations. By the BC5 the average length of donor segments is approximately 14 cM – 50% of that obtained with the same number of selfing generations (29 cM).

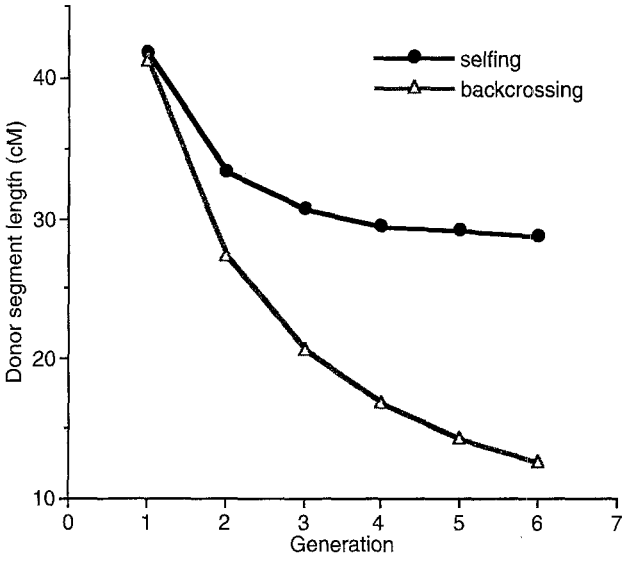


Fig. 1 Average length in centimorgans (cM) of intact donor chromosomal segments (y-axis) versus number of selfing or backcross generations (x-axis) subsequent to production of an F1 hybrid (e.g., for selfing, generation 1 = F2; for backcrossing, generation 1 = BC1). See Materials and methods for details of simulation parameters

Power of detecting QTLs in backcross versus selfing generations

The statistical detection of QTLs is likely to depend not only on the type of population utilized, but also on the intralocus and interlocus interactions of the segregating QTLs. For this reason, simulations were done to determine the power of detecting different types of QTLs in both selfing and backcrossing populations (Fig. 2). It should be noted that because additive and dominant effects are inseparable in backcross populations, the plots for additive, dominant and over-dominant QTLs in backcross generations are effectively identical to one another.

Additive QTLs

A single selfing generation (e.g., F2) was approximately twice as powerful (based on R^2 values) as a BC1 generation in detecting a QTL with additive gene action and no epistatic interactions (Fig. 2 A). Additional selfing slightly increased the power of detecting a QTL com-

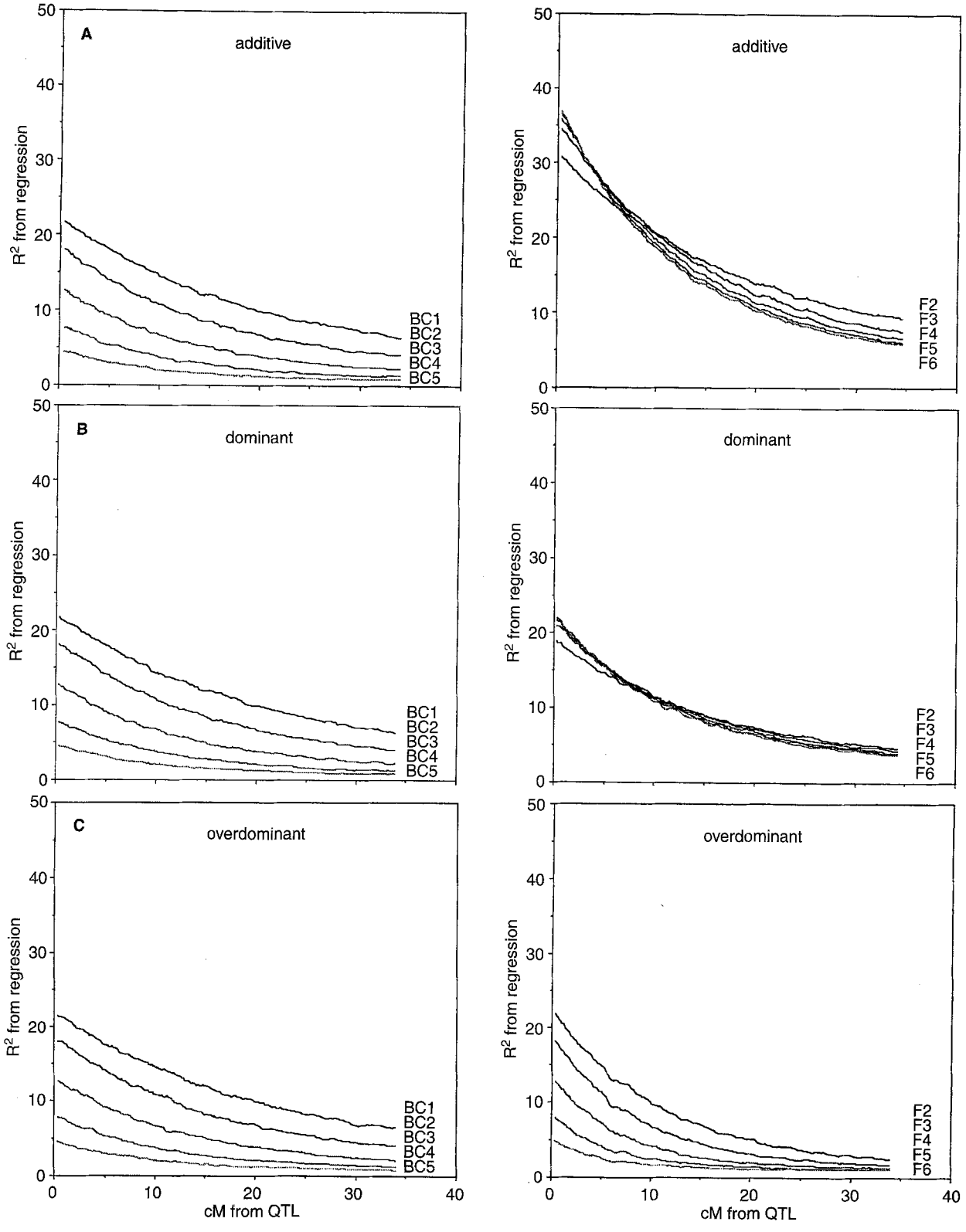
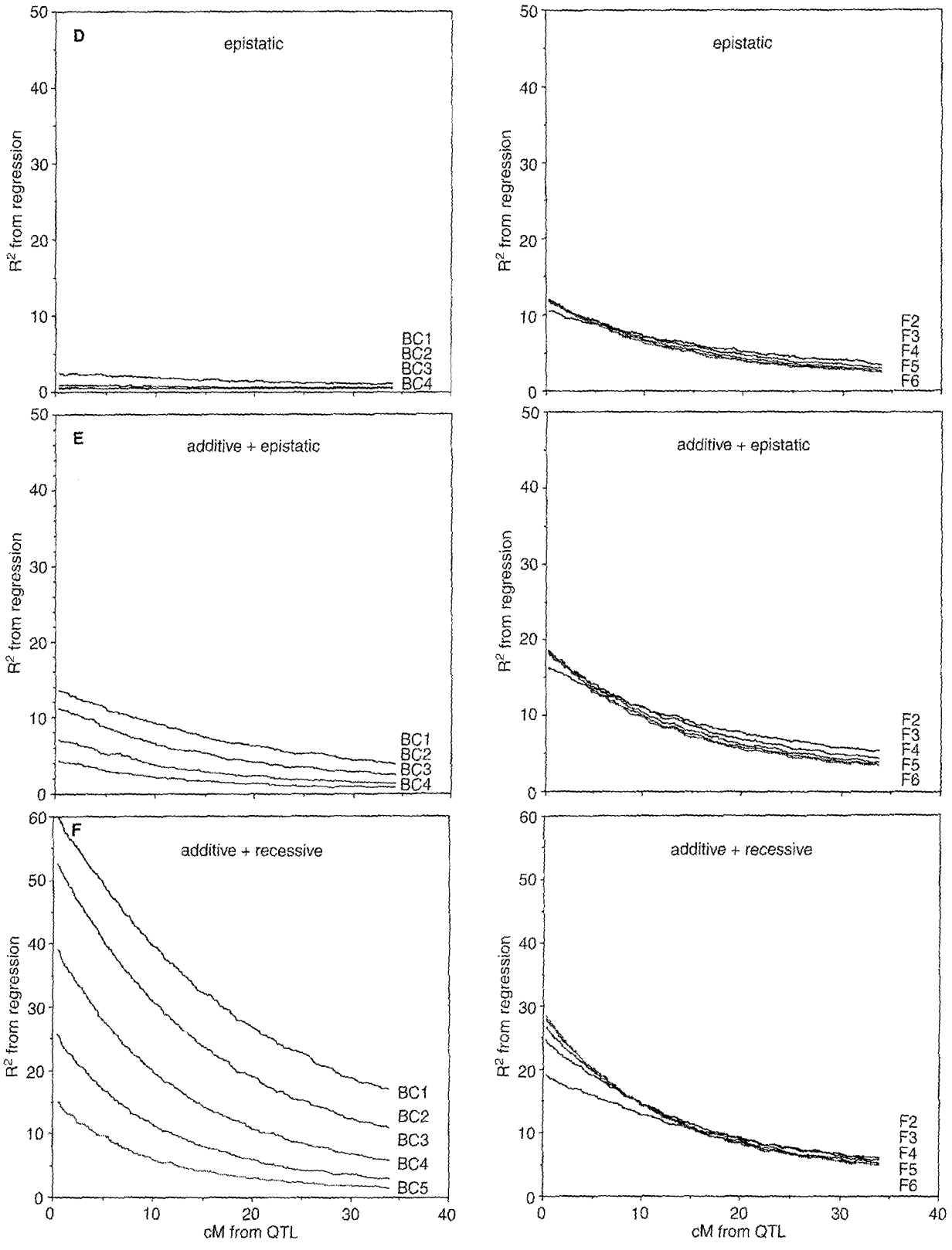


Fig. 2A–F Plots of R^2 values (x 100) (y-axes) from regression of marker genotype on phenotype for the detection of QTLs located at specified centimorgan (cM) distances from the marker. Detection of: **A** additive QTL, **B** dominant QTL, **C** overdominant QTL, **D** epistatic QTL, **E** additive QTL in the presence of an unlinked, epistatic QTL, **F** additive QTL in the presence of an unlinked recessive QTL

pletely linked with the test marker, but resulted in decreased power as the distance between the QTL and marker increased to more than 10 cM. Additional generations of backcrossing consistently decreased the power of detecting the linked QTL, regardless of the distance from the marker to the QTL.



Recessive QTLs

A recessive QTL coming from the donor genome would go undetected in any backcross generation since no genotypes homozygous for the donor allele would occur in such generations (simulation not shown).

Dominant QTLs

A QTL for which the donor genome contained a completely dominant allele would be detected in both selfing and backcrossing generations (Fig. 2 B). The efficiencies for all selfing generations are very similar and equally

diminished with increasing distance between the QTL and the marker. For backcrossing the power of detecting a dominant QTL in the first or second generation is very similar to that in the selfing generations. However, with every additional generation of backcrossing the power is reduced and by BC5 the power is reduced by nearly 80% of that in the BC1 generation. This is likely because only a very few individuals (approximately 3%) would still contain donor alleles at any given locus by the BC5 generation.

Over-dominant QTLs

An over-dominant QTL is defined as a locus for which individuals heterozygous for a donor and recurrent parent allele have a greater phenotypic value than either homozygous genotype. For simulation purposes, the homozygous genotypes were considered to have the same mean with respect to the phenotypic trait (see Materials and methods for details). Because over-dominance requires heterozygous individuals, one would expect detection of an overdominant QTL to be less powerful as a population becomes more inbred. The simulation for selfing generations confirms this prediction (Fig. 2 C). Each generation of selfing reduces the power (based on R^2 values) by approximately 20% each generation. By the F6 generation, R^2 values are so low that detection of an overdominant QTL would be unlikely (Fig. 2 C).

Backcross populations are, on average, more powerful in detecting overdominant QTLs than selfing generations (Fig. 2 C). When the QTL and marker are completely linked (0 cM), the R^2 for each backcross generation is the same as for the corresponding selfing generation. However, as the distance between the marker and QTL increases, the power of the backcross generation remains relatively high as compared with the corresponding selfing generation. For example, when the QTL and marker are 10 cM apart, the $R^2 = 12$ for a BC2 generation; however, for the corresponding generation of selfing (F3) $R^2 = 8$. As the distance between the QTL and marker increases, selfing becomes even less powerful. In selfing generations recombination between the QTL and marker may have occurred in either parental gamete. For backcrossing, recombination is restricted to gametes from only one parent. This difference probably explains the decreased power with selfing as compared with backcrossing.

Epistatic QTLs

A QTL whose allelic effects depend on the genotype of another locus (loci) elsewhere in the genome is referred to herein as an epistatic QTL. In the simulations the donor allele for a given QTL was expressed only in the presence, in the same individual, of the donor allele from an unlinked modifier locus (see Material and methods for details). The probability that the donor allele for both the QTL and modifier locus will occur simultaneously in any given individual in the population depends in part on the frequency of the donor alleles in the population. For example, in any selfing generation the genotypes with the donor allele at both QTL and modifier locus will constitute 1/4 of the population. In contrast, the fraction is only 1/16 in the BC1 and becomes negligibly small in succeeding generations. Thus, one would predict that the power of detecting an epistatic QTL to be much lower in backcrossing generations versus selfing generations. The simulations displayed in Fig. 2 D confirm this prediction. An epistatic QTL is detected with nearly equal power in any selfing generation. In contrast, the R^2 values for any backcross generation are nearly ten-fold less than for selfing generations. These results suggest that one is unlikely to detect an epistatic QTL in backcross populations, especially in advanced generations ($\geqslant$ BC2).

Detection of an additive QTL in the presence of an unlinked, epistatic QTL

As described above, a major difference in selfing and backcross populations is the amount of genetic variance contributed by epistatic QTLs. Backcross populations (especially in advanced generations) should have only a small portion of the genetic variance contributed by interaction of donor QTL alleles. In contrast, a much larger portion of the genetic variance should be due to epistatic interactions in selfing populations due to the overall higher frequency of donor alleles. For this reason, it is important to ask the question whether detection of non-epistatic QTLs (e.g., additive or dominant QTL) would be differentially affected by the segregation of epistatic QTLs (elsewhere in the genome) in backcross versus selfing generations. To test this notion simulations were made in which the power of detecting a non-epistatic, additive QTL was measured in populations in which an epistatically expressed, unlinked QTL was also segregating (Fig. 2 E).

For both backcross and selfing populations the power of detecting an additive QTL is lower in the presence of epistatic QTLs (Fig. 2 E) than in the case where the additive QTL is segregating alone (Fig. 2 A). However, selfing generations appear to be more adversely affected by the presence of the epistatic QTLs and the R^2 values were reduced by approximately 50% (Fig. 2 A,E). In comparison the R^2 was reduced approximately 35% for the BC1 generation, and in subsequent backcross generations the reduction was even less severe.

Detection of an additive QTL in the presence of a recessive QTL

QTLs with recessive donor alleles will not contribute to the genetic variance in backcross populations but will

contribute significantly to the variance in selfing generations. Whether the increased genetic variance, due to recessive QTLs in selfing generations, affects the ability to detect an additive QTL was tested. The power of detecting an additive QTL was measured in populations in which another unlinked QTL with a recessive donor allele was also segregating. The results indicate that the presence of a segregating recessive QTL does affect the ability to detect an unlinked additive QTL in selfing generations (Fig. 2F). In contrast, backcross generations are unaffected since the recessive QTL is not expressed (Fig. 2F). This difference would most likely enhance the power of backcross populations in detecting additive or dominant QTLs in situations where the donor genome contains recessive alleles at other QTL loci.

Probability of recovering QTL-NILs

The goal of QTL analysis as described in this manuscript is the detection of valuable QTL alleles from a donor genome and their transfer into elite breeding lines. In this context, it is important to consider how many additional generations of backcrossing, and how many individuals would need to be sampled, in order to obtain lines in which the segment of the donor chromosome containing the valuable QTL is present in an otherwise recurrent parent genome. We refer to such lines as QTL/nearly isogenic lines or QTL-NILs.

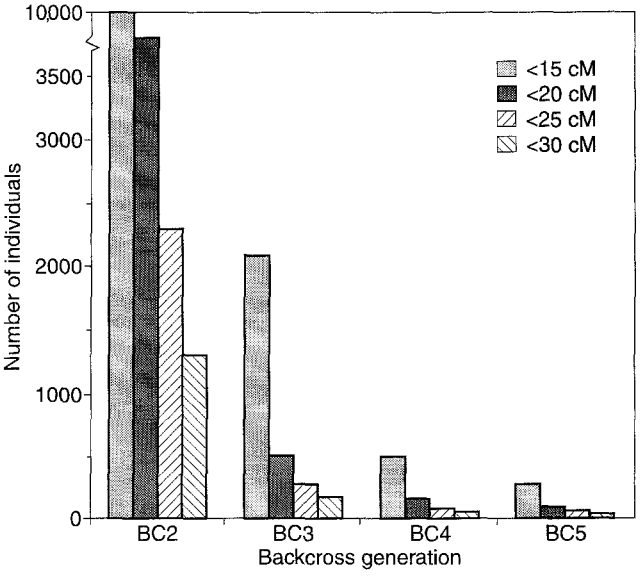
Two factors are important in determining how many generations and how many individuals are needed to create a QTL-NIL. These are: (1) the maximum size (in centimorgans) of the donor segment specified to contain the QTL and (2) the amount of residual donor genome (unlinked to the targeted QTL) still present in the genome. In this regard backcross populations are skewed toward recurrent parent alleles and should therefore be superior to selfing generations.

Simulations were employed to test the efficiencies of generating QTL-NILs from backcross versus selfing populations. For these simulations we assumed that a QTL has been localized to a marker-defined segment of donor DNA at least 10 cM long. We examined the probability of recovering after one backcross a genotype heterozygous only in a segment < 15 cM, < 20 cM, < 25 cM, or < 30 cM long that included the target segment. The lower the maximum donor segment size, the less the linkage drag (and hence undesirable effects) a QTL-NIL is likely to have.

Using the concept of graphical genotypes for whole-genome selection (Young and Tanksley 1989) a single individual was selected from a population of 200 individuals. This individual was determined to contain the donor allele for the specified QTL and also to be the individual most likely to yield the desired heterozygous QTL-NIL with one additional cross to the recurrent parent. The number of individuals that would need to be screened to yield the desired heterozygous QTL-NIL with a probability greater than 90% was plotted against the generation from which the initial individual was selected (Fig. 3).

The simulations reveal two main points. First, to select for QTL-NILs with small target segments (e.g., < 15 cM) would require, in most instances, very large populations. For example, more than 10 000 BC2-derived progeny would be required to have aa 90% chance of recovering at least one such QTL-NIL. If one accepts a larger target segment (e.g., < 30 cM) the number of individuals to sample decreases dramatically. For example, in the BC2 the number of individuals drops to approximately 1000. The second point is that selecting QTL-NILs from BC1-derived populations would require prohibitively large populations (> 50 000 individuals, data not shown). QTL-NILs could be isolated from BC2 derived populations, but a large number of individuals (usually > 1000) would need to be screened. More realistically, one would probably screen a smaller number of individuals over two sequential generations (e.g., a backcross followed by a selfing) in which selection would be exerted to remove non-target donor segments. In contrast to the BC1 and BC2, QTL-NILs can be derived directly from BC3–BC5 selections from a relatively small number of individuals.

Fig. 3 Number of individuals that would need to be screened to have a $> 90\%$ chance of identifying at least one QTL-NIL. The population for selection is derived by hybridization of a single selected individual from primary populations (e.g., BC2, BC3, etc) to the recurrent parent (see Materials and methods and Results section for details)



Discussion

Selfing populations and backcross populations differ in the power with which different types of QTLs can be detected. Compared with selfing populations, backcross populations will be less powerful in detecting a donor QTL having some degree of recessive gene action. However, as the donor QTL becomes more dominant it will be detected with relatively greater power in backcross populations. Fully dominant or over-dominant QTLs will be detected in the BC1 and BC2 generations with efficiencies nearly equivalent to or greater than that of selfing populations. Backcross populations are not very powerful for detecting donor QTLs requiring epistatic interactions and it is very unlikely that such QTLs would be detected in advanced backcross generations. Recessive and epistatic QTLs contribute little to the phenotypic variance of backcross generations compared with selfing generations. As a result, detection of additive, dominant and over-dominant QTLs is enhanced in backcross generations as compared with selfing generations in cases where recessive or epistatic QTLs are segregating simultaneously in the population.

Alleles of both parents occur in high frequency for all loci in all selfing generations as well as in the BC1 generation. This creates a problem if one is trying to identify a useful QTL(s) from a donor line and then to transfer that QTL(s) into the elite parent. Any given individual from a selfing population or a BC1 population is likely to contain, in addition to the useful donor QTL, many other donor segments that are undesirable. To recover the recurrent parent genome while maintaining the useful QTL (i.e., create a QTL-NIL) would require very large progeny sizes and/or a number of backcrosses to the recurrent parent (Fig. 3). In contrast, from advanced backcross generations one is much more likely to recover a plant containing the useful QTL, with few or no undesirable donor segments. This is especially true for advanced backcross populations (e.g., $\geqslant$ BC2). The more advanced the backcross population, the easier it will be to recover desired QTL-NILs (Fig. 3).

Based on the above considerations, we propose the application of "advanced backcross QTL analysis" in which QTL mapping is delayed to either the BC2 or BC3 generations. Generations beyond the BC3 are likely to have too low statistical power to detect most QTLs. However, additive, dominant and over-dominant QTLs can be detected statistically in BC2 or BC3 generations and the plants are sufficiently similar to the recurrent parent to allow ready isolation of QTL-NILs which may serve directly as improved varieties (or as a parent of a variety in the case of hybrid crops) or used to further confirm and fine map selected QTLs. The potential applications and limitations of AB-QTL analysis with respect to plant breeding are discussed below.

Strategies for the application of AB-QTL analysis in crop improvement

Self-pollinated crops

For the application of AB-QTL analysis to a self-pollinated crop, a single elite inbred variety could be crossed to an unrelated donor line (e.g., land race or wild species). One hundred or more BC1 progeny could be generated. Selection could be exercised on the BC1 population to remove any individuals with obviously undesirable characteristics (e.g., sterility, seed shattering, undesirable growth habit, small fruit, etc). The remaining BC1 could be crossed again to the recurrent parent to produce a BC2 population of > 200 individuals. If a sufficiently large BC2 population were generated, selection could be exercised again to remove obviously undesirable plants. A minimum of 200 selected BC2 individuals could be measured in replicated trials. By using BC2S1 or BC2S2 families it should be possible to detect some recessive QTL donor alleles in addition to the expected dominant and additive donor QTL alleles. BC2 marker data would be used to search for QTL associations from the BC2S1 or BC2S2 family performance data.

Once putatively beneficial QTLs were identified from such an analysis, one would use whole-genome selection to identify the BC2S1 line(s) from which QTL-NILs could be isolated. No more than two generations should be required to isolate targeted QTL-NILs and such QTL-NILs could be evaluated in replicated trials for performance against the original elite inbred variety. As little as 1 year might elapse from the time the QTL analysis is originally conducted (i.e., BC2S2 generation) until field trials could be performed on selected QTL-NILs. QTL-NIL(s) outperforming the original elite variety would replace that elite variety as the recurrent parent in future experiments. Thus the process leads to a stepwise improvement of elite lines through the introgression of valuable QTLs. In addition, QTLs of proven benefit could be easily combined by intercrossing different QTL-NILs.

Hybrid crops

Advanced backcross QTL analysis could be applied to hybrid crops via a slight modification of the strategy presented above. Assume one is starting with a commercial hybrid derived from a cross of two inbreds, A and B. The donor line would be hybridized and backcrossed to inbred A to produce a BC2 population as described above. Marker analysis would be conducted on the BC2 plants. However, instead of selfing the selected BC2 plants, these plants would instead be crossed to inbred B to produce BC2F1 families which would be field tested for yield and other traits. Based on QTL analysis of these data, QTL-NILs could be generated in the A inbred background and crossed to inbred B, or other inbreds, for replicated trials to compare performance

against the original elite hybrid. In this manner a stepwise improvement of one (or both) inbreds could be accomplished.

Crops for which AB-QTL analysis is not likely to be useful

Advanced backcross QTL analysis would be most readily applied to annual crops with relatively short generation times (< 2 years). Development and exploitation of the necessary advanced backcross populations and QTL-NILs would very likely limit the use of this method in long-generation perennial crops. It would also be difficult to apply this method to crops for which inbred lines do not exist. The success of this method requires that the commercial varieties to be improved are either inbred themselves or hybrids derived from the crossing of inbreds. It would be difficult to apply the method to highly heterozygous, outcrossing crops like alfalfa or clonally propagated crops like potato for which inbred lines are not commonly employed in breeding programs.

Advantages of AB-QTL analysis for introgression of valuable QTLs

The advantages of AB-QTL analysis as compared with conventional strategies for QTl detection are as follows: (1) Since phenotypic selection can be practiced on single plants in the BC1 and BC2 generations, major negative QTLs which would otherwise interfere with later measurement of field/quality performance in the BC2F1 families can be reduced or eliminated. (2) Because BC2F1 families are skewed toward alleles from the recurrent parent, the probability is reduced for the detection of QTLs requiring epistatic interactions among alleles from the wild parent. Instead, there will be a higher probability of detecting additive QTLs which will continue to function as predicted when they are placed in the nearly-isogenic background of the recurrent parent. (3) Since the mean performance of the BC2F1 families is skewed towards the elite parent, subtle pleiotropic effects are more likely to be detected. (4) It is possible to produce lines nearly isogenic for selected QTLs (QTL-NILs) in as little as a single generation after QTL detection. QTL-NILs (or hybrids produced with QTL-NILs) are candidates for new varieties with one or more enhanced attributes. Hence, unlike conventional QTL analysis which can require up to 5 more years of breeding to create candidate varieties, AB-QTL analysis results in candidate varieties in as little as a year after the initial QTL detection.

Comparison of AB-QTL analysis with the inbred backcross breeding method

Wehrhahn and Allard (1965) proposed the development of inbred backcross lines by crossing a donor line with a recurrent line, followed by several sequential backcrosses to the recurrent parent. These backcross plants and their descendents would then be selfed until a state of near homozygosity. If the trait of interest was affected by a few (< 10) QTLs, then most inbred backcross lines would differ for donor alleles at only a single QTL. Inbred backcross lines would therefore be expected to fall into discrete phenotypic clusters determined by difference at single QTLs. A comparison of phenotypic clusters would then allow inferences about the number and magnitude of effects of QTLs (Wehrhahn and Allard 1965). Bliss (1981) later proposed the use of inbred backcross lines as the basis of a breeding method for the use of exotic germplasm to improve quantitative traits in crop plants. The inbred backcross method has been demonstrated in beans in which high-yielding, high-protein BC2 selfed lines were identified (Bliss 1981; Sullivan and Bliss 1983 a,b). The method has also been utilized in breeding for fruit weight and length in cucumbers (Owens et al. 1985 a,b).

The inbred backcross breeding method and AB-QTL analysis method described in this paper are similar in that they both rely on sequential backcrosses to create populations in which allele frequencies are highly skewed towards the recurrent parent. However, the methods differ in the manner in which these skewed populations are utilized to improve quantitative traits and create improved varieties.

Genotypic versus phenotypic selection

Inbred backcross breeding, like all breeding methods, is based on phenotypic selection and dependent upon the random appearance of inbred backcross lines, or progeny derived from intercrossing selected inbred backcross lines, whose phenotypes are superior to that of the recurrent parent (Sullivan and Bliss 1983a,b; Owens et al. 1985 b). However, if such lines do not appear, there is no obvious way to direct their creation. In contrast, AB-QTL analysis is based on utilizing information about the map locations and effects of favorable QTL alleles from the donor parent to create QTL-NILs which are likely to be superior to the recurrent parent. In other words, AB-QTL analysis is based on genotypic selection versus phenotypic selection. Superior QTL-NILs can be created whether or not superior phenotypes occur in the generation used for QTL mapping (e.g., BC2). This is an important point, since, even though favorable QTL alleles exist in most exotic germplasm, they may not manifest themselves in phenotypes that surpass the recurrent parent until they are isolated away from other, inferior alleles from the donor.

QTL-NILs and linkage drag

Unlike those of inbred backcross breeding, the end-products of AB-QTL analysis will be nearly isogenic

lines (QTL-NILs) which carry recurrent parent alleles throughout their genome, except for the specific targeted QTL. With QTL-NILs, any phenotypic difference between the recurrent parent and the QTL-NIL, beneficial or deleterious, can most likely be attributed to the targeted QTL. Markers linked to the QTL can be used to select for derived QTL-NILs with reduced linkage drag. In contrast, breeding lines derived from inbred backcross breeding will almost certainly contain a number of segments of the donor genome and an overall higher donor genome content than QTL-NILs derived from AB-QTL analysis. While such lines may be improved for one or two target characters, there is a substantial risk that they will have been modified for other untested traits which will ultimately be important in the success of the breeding line. Also, there is no obvious way to reduce linkage drag significantly in such lines. In this regard, QTL-NILs derived from AB-QTL analysis are more likely to match or exceed the recurrent parent with respect to overall performance.

QTL map position and genotypic design

QTL mapping information derived from AB-QTL analysis is cumulative. Every time AB-QTL analysis is applied the map positions of donor QTLs affecting key traits will likely be discovered. Based on this knowledge, it would be straightforward to combine beneficial donor QTLs discovered in one experiment with non-allelic QTLs affecting the same trait from experiments in which a different donor parent was used. It should be possible to design genotypes containing QTL alleles, from several different exotic donor lines, that are likely to transgress the recurrent parent phenotype even further. In the absence of map information and marker technology, the development of NILs containing mixtures of specific QTL alleles from different donor parents would be difficult and time-consuming and there would be no clear way to determine if one had successfully accomplished the goal. Combining QTL alleles from multiple donors in a systematic manner would be nearly impossible with conventional breeding techniques.

AB-QTL analysis and the cloning of QTLs

Map-based cloning has now been demonstrated to be a powerful technique for the isolation of major genes from plants without prior knowledge of their gene products (Arondel et al. 1992; Martin et al. 1993; for review see Tanksley et al. 1995). Unfortunately, many traits of agronomic importance are not represented by major-gene mutations, but are quantitatively inherited (e.g., yield, drought tolerance, horizontal disease resistance). The introduction of molecular maps into quantitative genetics has made it possible to map and characterize QTLs underlying most quantitative traits. If a QTL can be mapped to a precise position in the genome relative to molecular markers, the prospects are good that the QTL can be cloned via the chromosome landing approach (Tanksley et al. 1995).

QTL-NILs provide a starting point for fine mapping of QTLs since associated, flanking markers can be used to generate the necessary recombinants for high-resolution mapping and the uniformity of the genetic background should permit straightforward phenotypic evaluations. AB-QTL analysis can result in the creation of a reservoir of well-defined QTL-NILs which may ultimately be the substrate for map-based cloning of key genes underlying quantitative traits.

Acknowledgments This work was supported in part by grants from the National Research Initiative Cooperative Grants Program, Plant Genome Program USDA (No. 58-1908-5-001) and by the Binational Agricultural Research and Development Fund (No. US-2427-94) to SDT. We thank D. Zamir, A. Frary, S. McCouch, M. Sorrells, J. Xiao, D. Bernachi, S. Grandillo and T. Fulton for helpful comments and discussions.

References

Allard RW (1960) Plant breeding. John Wiley and Sons, New York
Anderson JA et al. (1993 a) Detection of QTLs affecting pre-harvest sprouting resistance in wheat by RFLPs. Crop Sci 33:543
Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993 b) Optimizing parental selection for genetic linkage maps. Genome 31:181–186
Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258:1353–1355
Bellmann K, Ahrens H (1966) Modellpopulationen in der Selektionstheorie und einige Ergebniße aus Simulationsstudien. Der Züchter 36:172–185
Bliss FA (1981) Utilization of vegetable germplasm. HortScience 16:129–132
Eshed Y, Zamir D (1994) Introgressions from *Lycopersicon pennellii* can improve the soluble-solids yield of tomato hybrids. Theor Appl Genet 88:891–897
Fatokun CA, Menancio-Hautea Dl, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841–846
Fraser A, Burnell D (1970) Computer models in genetics. McGraw-Hill, New York
Frey KJ, Cox TS, Rodgers DM, Bramel-Cox P (1981) Increasing cereal yields with genes from wild and weedy species. Journal Paper No. J-11254 of the Iowa Agric and Home Econ Exp Stn, Ames, Iowa 50011, USA. Project 2447, pp 51–68
Hanson WD (1959) Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics 44:833–837
Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa
Hayes PM, Liu BH, Knapp SJ, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson, Sorrells M, Ullrich, SE, Wesenberg D, Kleinhofs A (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germ plasm. Theor Appl Genet 87:392–401
Helentjaris T, King G, Slocum M, Siedenstrang, C, Wegman S (1985) Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118
Keim P, Diers BW, Shoemaker RC (1990) Genetic analysis of soybean hard seededness with molecular markers. Theor Appl Genet 79:465–469
Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Ladizinsky G (1985) Founder effect in crop-plant evolution. Econ Bot 39:191–199
Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436
Messeguer R (1991) High-resolution RFLP map around the root knot nematode resistance gene (*Mi*) in tomato. Theor Appl Genet 80:437–448
Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448
Owens KW, Bliss FA, Peterson CE (1985 a) Genetic analysis of fruit length and weight in two cucumber populations using the inbred backcross line method. J Am Soc Hort Sci 110:431–436
Owens KW, Bliss FA, Peterson CE (1985 b) Genetic variation within and between two cucumber populations derived via the inbred backcross line method. J Am Soc Hort Sci 110:437–441
Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726
Simmons, NW (1976) Evolution of crop plants. Longman, London New York
Soller M, Beckmann JS (1983) Genetic polymorphism in variety identification and genetic improvement. Theor Appl Genet 67:25–33
Stam PA, Zeven C (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238
Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839
Sullivan JG, Bliss FA (1983 a) Genetic control of quantitative variation in phaseolin seed protein of common bean. J Am Soc Hort Sci 108:782–787
Sullivan JG, Bliss FA (1983 b) Expression of enhanced seed protein content in inbred backcross lines of common bean. J Am Soc Hort Sci 108:787–791
Tanksley SD (1983) Molecular markers in plant breeding. Plant Mol Biol Rep 1:3–8
Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257–264
Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11:63–68
Ullstrup AJ (1978) Corn diseases in the United States and thier control. Agricultural Handbook 199:1–55
Vicente MC de, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596
Wang ZY, Second G, Tanksley SD (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryze* as determined by analysis of nuclear RFLPs. Theor Appl Genet 83:565–581
Wang G, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136:1421–1434
Wehrhahn C, Allard RW (1965) The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51:109–119
Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353–359
Zamir D, Ekstein-Michelson I, Zakay Y, Navot N, Zeidan M, Sarfatti M, Eshed Y, Harel E, Pleban T, van-Oss H, Kedar N, Rabinowitch HD, Czosnek H (1994) Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, *TY-1*. Theor Appl Genet 88:141–146