

QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analyses

Alizera Navabi · D. E. Mather · J. Bernier ·
D. M. Spaner · G. N. Atlin

Received: 17 December 2007 / Accepted: 23 September 2008 / Published online: 15 October 2008
© Springer-Verlag 2008

Abstract Selective genotyping of one or both phenotypic extremes of a population can be used to detect linkage between markers and quantitative trait loci (QTL) in situations in which full-population genotyping is too costly or not feasible, or where the objective is to rapidly screen large numbers of potential donors for useful alleles with large effects. Data may be subjected to ‘trait-based’ analysis, in which marker allele frequencies are compared between classes of progeny defined based on trait values, or to ‘marker-based’ analysis, in which trait means are compared between progeny classes defined based on marker genotypes. Here, bidirectional and unidirectional selective genotyping were simulated, using population sizes and selection intensities relevant to cereal breeding. Control of

Type I error was usually adequate with marker-based analysis of variance or trait-based testing using the normal approximation of the binomial distribution. Bidirectional selective genotyping was more powerful than unidirectional. Trait-based analysis and marker-based analysis of variance were about equally powerful. With genotyping of the best 30 out of 500 lines (6%), a QTL explaining 15% of the phenotypic variance could be detected with a power of 0.8 when tests were conducted at a marker 10 cM from the QTL. With bidirectional selective genotyping, QTL with smaller effects and (or) QTL farther from the nearest marker could be detected. Similar QTL detection approaches were applied to data from a population of 436 recombinant inbred rice lines segregating for a large-effect QTL affecting grain yield under drought stress. That QTL was reliably detected by genotyping as few as 20 selected lines (4.5%). In experimental populations, selective genotyping can reduce costs of QTL detection, allowing larger numbers of potential donors to be screened for useful alleles with effects across different backgrounds. In plant breeding programs, selective genotyping can make it possible to detect QTL using even a limited number of progeny that have been retained after selection.

Communicated by M. Sillanpaa.

A. Navabi · J. Bernier · D. M. Spaner
Department of Agricultural, Food, and Nutritional Science,
University of Alberta, Edmonton, AB T6G 2P5, Canada

Present Address:

A. Navabi (✉)
Agriculture and Agri-Food Canada,
Department of Plant Agriculture, University of Guelph,
Guelph, ON N1G 2W1, Canada
e-mail: navabia@agr.gc.ca

D. E. Mather
Molecular Plant Breeding Cooperative Research Centre
and School of Agriculture, Food and Wine,
University of Adelaide, PMB 1,
Glen Osmond, SA 5064, Australia

J. Bernier
IRRI, DAPO 7777, Metro Manila, Philippines

G. N. Atlin
CIMMYT, Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico

Introduction

Many useful alleles with large effects on economically important quantitative traits are likely present in crop germplasm, yet few have been deployed in breeding programs. One reason for this is that the experimental approaches used to discover quantitative trait loci (QTL) which typically involve assaying trait phenotypes and marker genotypes on several hundred progeny from a

cross, can be too costly to apply across a wide range of germplasm.

Selective genotyping is an alternative approach for QTL detection, in which DNA markers are assayed only on the most genetically informative progeny: those with extremely high and/or low phenotypic values for a trait of interest. This allocation of genotyping resources only to selected progeny can reduce genotyping costs with little loss of information, freeing resources for investigation of more and larger populations, and/or for validation and fine-mapping of QTL that have been detected. This concept was introduced by Lebowitz et al. (1987), who used the term ‘trait-based analysis’ to refer to approaches to QTL mapping in which marker allele frequencies are compared between groups of progeny selected based on trait values. Lander and Botstein (1989) introduced the more general term ‘selective genotyping’ for QTL mapping based on selected groups of progeny, and suggested that QTL analysis in this case could also be based on the usual marker-based approaches that compare phenotypic values among marker genotype classes.

Lebowitz et al. (1987) and Gallais et al. (2007) have discussed the theory and experimental design for analysis of marker allele frequencies in classes of progeny defined on the basis of quantitative trait values. Both groups concluded that trait-based analysis of selectively genotyped progeny can be a useful alternative to marker-based analysis of all individuals in a population, when only one quantitative trait is of interest. Gallais et al. (2007) concluded that, for a given population size of phenotyped individuals, the optimal proportion selected for genotyping is around 30% from each tail.

Lander and Botstein (1989) discussed the application of marker-based analysis to trait and marker data from selectively genotyped populations, suggesting that maximum-likelihood QTL detection methods could be applied, with the genotypes of non-selected progeny considered as missing values. Darvasi and Soller (1992) investigated this approach, concluding that for detection of marker-QTL linkage for an individual trait, it is rarely useful to genotype more than the upper and lower 25% of the phenotypic frequency distribution. Xu and Vogl (2000) developed an exact maximum likelihood approach to map QTL by selective genotyping using phenotypic values of genotyped individuals only. Liu (1998), however, argued that selection will bias hypothesis testing and parameter estimation in maximum-likelihood QTL analysis, and Lee (2005) used genetic simulation to show that selection can reduce the accuracy of QTL detection and bias the estimation of QTL effects.

Trait-based analysis has been applied for QTL detection with bidirectional selective genotyping (i.e., analysis of allele frequencies in both the high and low tails of the phenotypic distribution; e.g., Zhang et al. 2003) and with unidirectional selective genotyping (i.e. analysis of allele

frequencies from only one selected tail; e.g., Foolad et al. 2001). Unidirectional selective genotyping is of particular interest for application within breeding programs, because it has the potential to permit QTL detection using superior progeny that have been retained under selection in breeding programs. If effective, this could help integrate QTL detection with plant breeding, addressing concerns that the treatment of QTL discovery and cultivar development as separate processes may have limited the impact of marker-aided selection in plant breeding (Tanksley and Nelson 1996).

Although both marker-based and trait-based methods have been proposed for analysis of data from selective genotyping, the two approaches have not been compared in detail. Further, analyses of the effects of various factors (such as proportion selected, QTL effect, marker-QTL distance) on the QTL detection power of selective genotyping have been based on asymptotic approximations of theoretical distributions. These may differ appreciably from the actual effects on power in small samples relevant to actual experimental and breeding programs. Here, we present results of simulation studies conducted to (1) investigate the power and precision of QTL effect estimation of trait-based and marker-based analysis in unidirectional and bidirectional selective genotyping and (2) examine the effects of the proportion selected for genotyping, the magnitude of QTL effects, population size, and marker-QTL distance on the power of QTL detection, using values relevant to breeding programs in self-pollinated cereal crop species. We also present results from the application of some of these methods to data from a rice (*Oryza sativa* L.) mapping population. The overall objective of this work is to provide guidance in the design of low-cost trait-based selective genotyping experiments that can be applied to detect large-effect QTL alleles in crop germplasm collections and in ongoing breeding programs.

Materials and methods

Genetic simulations

Unpublished *Perl* scripts (kindly provided by Hai Pham and Nicholas Tinker) were used to simulate a single QTL with an additive effect located at the centre of a 150-cM linkage group. One marker locus was placed at the QTL position, and 15 markers were placed on either side of the QTL, with the probability of recombination between each pair of adjacent markers set to 0.05, resulting in intervals of 5.0 cM (Kosambi mapping function: Kosambi (1944)) and marker-QTL distances ranging from 0 to 75.0 cM. An additional marker locus, not linked to the QTL or any other marker locus, was also simulated in order to permit

calculation of Type I error rates. The reference population was a set of doubled haploid lines (inbreeding coefficient = 1) derived without selection from the F_1 of a cross between two parental lines homozygous for alternative alleles at the QTL and at each marker. The model for the phenotypic variance of a trait in a population of random doubled haploid lines was:

$$\sigma_P^2 = \sigma_{QTL}^2 + \sigma_{BG}^2 + \sigma_E^2 \quad (1)$$

in which σ_P^2 is the phenotypic variance, σ_{QTL}^2 is the genetic variance due to segregation of the QTL to be simulated, σ_{BG}^2 is the genetic variance due to segregation of an unspecified number of other QTL affecting the trait and not linked to the simulated QTL or any of the simulated markers, and σ_E^2 is the non-heritable variance. With σ_P^2 set at a standard value of 1, with completely additive effects and with equal frequencies ($p = q = 0.5$) for two alternative alleles at the QTL, the additive effect of the QTL, half the difference between alternative homozygotes at the QTL (Mather and Jinks 1982), in a doubled haploid population is:

$$a = \sqrt{\sigma_{QTL}^2} \quad (2)$$

Using Eqs. 1 and 2, values of a and of the sum of σ_{BG}^2 and σ_E^2 were set to provide models in which the simulated QTL was responsible for 1, 3, 5, 9, 15 and 25% of σ_P^2 (R_{QTL}^2 from 0.01 to 0.25). This required the additive effect of the QTL to be set at 0.1, 0.1732, 0.2236, 0.3, 0.3873 and 0.5, respectively. For each marker linked to the QTL, the proportion of σ_P^2 expected to be associated with variation at that marker was calculated as $R_P^2 = R_{QTL}^2 (1 - 2r)^2$.

In each simulation run, a population of either 200 or 500 lines was generated. A total of 1000 simulation runs was conducted for each combination of population size (200 or 500), and R_{QTL}^2 (from 0.01 to 0.25). A phenotypic value was computed for each line as the sum of a mean value μ set at 0, the additive effects of the alleles present at the QTL and a random value drawn from a normal distribution of mean 0 and variance $\sigma_E^2 + \sigma_{BG}^2$. Lines were ranked on the basis of these phenotypic values, and those with extreme phenotypic values were selected. From the populations of 500 lines, subsets of the highest and lowest ranking 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, 115, and 125 lines were selected for bidirectional selective genotyping (i.e., selected proportions of 0.02, 0.06, 0.10, 0.14, 0.18, 0.22, 0.26, 0.30, 0.34, 0.38, 0.42, 0.46, and 0.50, respectively) and the highest ranking 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, 210, 230 and 250 lines were selected for unidirectional selective genotyping (i.e., selected proportions of 0.02, 0.06, 0.10, 0.14, 0.18, 0.22, 0.26, 0.30, 0.34, 0.38, 0.42, 0.46, and 0.50, respectively). Similarly, from populations of 200 lines, subsets of the

highest ranking and lowest ranking 5, 15, 25, 45 and 55 lines were selected for bidirectional selective genotyping (i.e., selected proportions of 0.05, 0.15, 0.25, 0.35, 0.45, and 0.55, respectively) and the highest ranking 10, 30, 50, 70, 90, and 110 lines were selected for unidirectional selective genotyping (i.e., selected proportions of 0.05, 0.15, 0.25, 0.35, 0.45, and 0.55, respectively).

Statistical analyses

Statistical analyses, including testing of marker-QTL linkage and estimation of Type I error rates, power of QTL detection, and QTL effects were conducted using SAS procedures (SAS Institute 2003). Marker- and trait-based analyses were used for QTL detection using a per-marker significance level of $\alpha = 0.01$:

Marker-based analysis

One-way analysis of variance (ANOVA) was applied to test for differences in quantitative trait values between contrasting marker genotypic classes. This test was applied to data from bidirectional selective genotyping only.

Trait-based analysis

A test based on a normal approximation of a binomial distribution of allele frequencies was applied to data from unidirectional and bidirectional selective genotyping. A QTL was considered to have been detected if $|d_q| \geq z_{(\alpha/2)} s_q$, where $|d_q|$ is the absolute value of the difference in marker allele frequencies, $z_{(\alpha/2)}$ is the ordinate of the standard normal distribution such that the area under the curve from $-\infty$ to $z_{(\alpha/2)}$ equals $1 - \alpha/2$, and s_q is the standard error of the difference between marker allele frequencies. For bidirectional selective genotyping, $|d_q|$ was estimated as the difference in allele frequencies between the two tails and s_q was estimated as:

$$s_q = \sqrt{\frac{p_u q_u}{n_u} + \frac{p_l q_l}{n_l}} \quad (3)$$

For unidirectional selective genotyping, $|d_q|$ was estimated as the difference in allele frequency between the selected tail and the expected frequency (0.5) and s_q was estimated following Lebowitz et al. (1987) as:

$$s_q = \sqrt{\frac{p_u q_u}{n_u}} \quad (4)$$

In Eqs. 3 and 4, p_u and q_u are alternate allele frequencies in selected samples from the upper tail, p_l and q_l are alternate allele frequencies in selected samples from the lower tail, and n_u and n_l are the numbers of lines in the

upper and lower tails, respectively. Unlike tests used by Lebowitz et al. (1987) and Zhang et al. (2003), the test used here for the bidirectional selective genotyping does not assume symmetrical changes in allele frequencies between the tails, nor does it assume equal variances for the two selected subsets. This test is similar to the selective genotyping approach implemented by Gallais et al. (2007).

For each combination of population size, R_{QTL}^2 , proportion selected for genotyping and selective genotyping design (unidirectional and bidirectional), the power of QTL detection was expressed as the proportion of simulation runs in which the simulated QTL was detected. Similarly, the Type I error rate was expressed as the proportion of simulation runs in which a significant effect was detected at the marker that was not linked to the QTL. In each case, marker-QTL linkage detection was considered reliable if the power was greater than 0.8 ($\beta < 0.20$) and Type I error rate was smaller than 0.01 ($\alpha < 0.01$).

For each marker at which a significant effect was detected by trait-based analysis of data from bidirectional selective genotyping, the proportion of phenotypic variance explained (R_p^2) was estimated as:

$$R_p^2 = \frac{d_q^2}{i^2[q(1-q)]} \quad (5)$$

where d_q is the difference in allele frequencies between the upper and lower selected tails and i is standardized selection differential (Falconer 1989). Similarly, for QTL detected by trait-based analysis of data from unidirectional selective genotyping, R_p^2 was estimated as:

$$R_p^2 = \frac{d_q^2}{4[i^2q(1-q)]} \quad (6)$$

where d_q is the difference between the observed allele frequency in the selected tail and the expected allele frequency for a random sample.

Analyses of data from a rice population

The effectiveness of selective genotyping was also investigated using data from a mapping study (Bernier et al. 2007) involving F_3 -derived recombinant inbred lines from a cross between the upland rice cultivars ‘Vandana’ (a moderately drought-tolerant Indian cultivar) and ‘Way Rarem’ (a drought-susceptible but high-yielding Indonesian cultivar). Bernier et al. (2007) evaluated 436 Vandana/Way Rarem lines for grain yield under severe drought stress at Los Baños, Philippines, over 2 years. They selected lines for genotyping at random (92 lines) or based on their grain yield under drought stress in 2005 (57 high-yielding lines and 48 low-yielding lines). There was some overlap between the random and selected subsets, and the

total number of lines genotyped was 169. These lines were genotyped with 131 DNA markers.

For each marker, allele frequencies were calculated for the randomly selected subset of 92 lines and for subsets of the 5, 10, 15, 20, 25 and 30 lines with the highest and lowest grain yields under drought stress (i.e., selected proportions of 0.01, 0.02, 0.03, 0.05, 0.06, and 0.07 for unidirectional selection and selected proportions of 0.02, 0.04, 0.06, 0.10, 0.12, and 0.14 for bidirectional selection). For each marker, the allele frequency observed in the random subset was tested against the expected frequency of 0.5 by a χ^2 test (Steel et al. 1997). Marker-based ANOVA and the trait-based test based on the normal approximation of the binomial distribution were applied for both bidirectional and unidirectional selective genotyping. For trait-based analysis of bidirectional selective genotyping, the allele frequencies in the high-yielding and low-yielding subsets were tested against each other. For unidirectional selective genotyping, allele frequencies in the high-yielding subset were tested against the expected frequency of 0.5 and also against the frequencies observed in the randomly selected subset.

Results

Simulation study

With selective genotyping of only 10 lines from a population of 200 or 500, observed Type I error rates were above the target rate of 0.01, especially when trait-based analysis as applied to data from bidirectional selection, for which the Type I error rate was above 0.06 (Table 1). With genotyping of 30 or more lines, the Type I error never exceeded 0.02, and converged on the target level of 0.01, as increasing numbers of lines were genotyped.

As expected, power increased with the proportion of the phenotypic variance explained by the simulated QTL (R_{QTL}^2) (Fig. 1) and with proximity of the marker locus to the QTL (not shown). With equal numbers of lines genotyped, power was consistently greater for bidirectional than for unidirectional selection (Fig. 1) and was greater for selection from a population of 500 than from a population of 200 (Fig. 1).

With bidirectional selection, QTL detection power was somewhat better with the trait-based analysis than with marker-based analysis, especially when small numbers of lines were genotyped (Fig. 1). With bidirectional selection in a population of 500 lines, a QTL explaining as little as 3% of the phenotypic variance could be reliably detected ($\beta \leq 0.20$), but only if there was a marker coinciding with the QTL and if a large number of lines (38% or more) were genotyped (Table 2). QTL explaining a larger proportion

Table 1 Type I error rates, computed as the proportion of 6,000 simulation runs (1,000 simulation runs of each of the 6 R^2_{QTL} values) in which a spurious marker-QTL linkage was detected, with bidirectional or unidirectional selective genotyping of progeny from populations of 200 or 500 doubled haploid lines, using marker-based ANOVA or a trait-based analysis using the normal approximation of the binomial distribution trait-based analysis

Number of lines selected	Bidirectional selective genotyping		Unidirectional selective genotyping Trait-based analysis
	Marker-based ANOVA	Trait-based analysis	
Population size = 200			
10	0.022	0.065	0.024
30	0.013	0.020	0.016
50	0.013	0.020	0.016
70	0.011	0.013	0.011
90	0.011	0.015	0.016
110	0.010	0.010	0.010
Population size = 500			
10	0.022	0.066	0.021
30	0.014	0.015	0.014
50	0.011	0.012	0.011
70	0.011	0.009	0.011
90	0.009	0.011	0.011
110	0.009	0.010	0.010
130	0.010	0.009	0.010
150	0.010	0.011	0.011
170	0.009	0.011	0.011
190	0.011	0.011	0.010
210	0.010	0.010	0.009
230	0.011	0.011	0.011
250	0.011	0.010	0.011

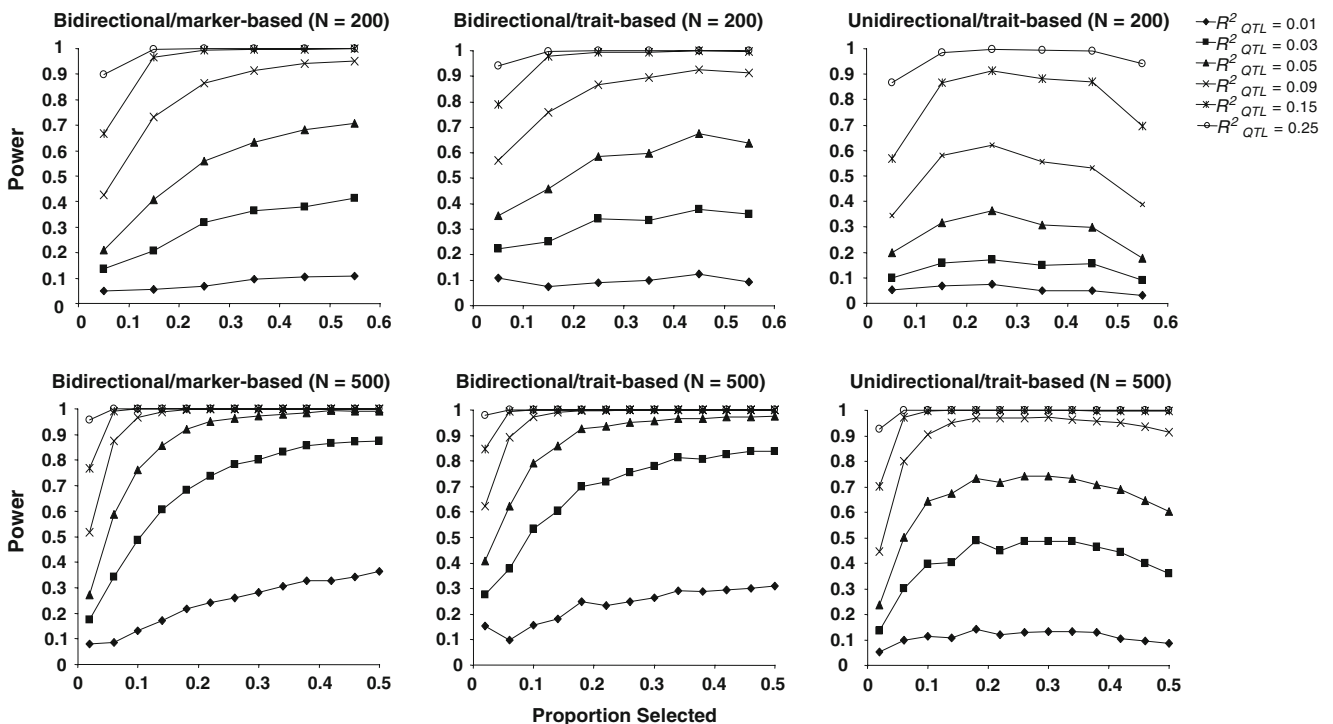


Fig. 1 The observed QTL detection power of trait-based and marker-based ANOVA of data from unidirectional and bidirectional selective genotyping of a population of 500 or 200 doubled haploid lines

derived from a cross between two homozygous lines for markers at 0 cM from the QTL, and for QTL with a range of effect sizes in 1,000 simulation runs

Table 2 Smallest proportion of a population of 200 or 500 doubled haploid lines required to be selectively genotyped for reliable ($\beta \leq 0.2$) detection of QTL with a range of effect sizes (R^2_{QTL}) using bidirectional and unidirectional selective genotyping for markers at different distances from a QTL in 1,000 simulation runs

Marker-QTL distance (cM)	Bidirectional selective genotyping						Unidirectional selective genotyping					
	ANOVA/ R^2_{QTL}						Trait-based analysis/ R^2_{QTL}					
	0.03	0.05	0.09	0.15	0.25	0.45	0.03	0.05	0.09	0.15	0.25	0.45
Population size = 200												
0	-	-	0.25	0.15	0.05	-	-	-	0.25	0.15	0.05	-
5	-	-	0.35	0.15	0.15	-	-	0.35	0.15	0.05	-	0.15
10	-	-	-	0.25	0.15	-	-	-	0.25	0.15	-	0.25
15	-	-	-	0.35	0.15	-	-	-	0.25	0.15	-	0.15
20	-	-	-	0.45	0.25	-	-	-	0.45	0.25	-	0.25
25	-	-	-	-	0.35	-	-	-	-	0.25	-	-
30	-	-	-	-	0.45	-	-	-	-	0.45	-	-
Population size = 500												
0	0.30	0.14	0.06	0.06	0.02	0.38	0.14	0.06	0.06	0.02	-	0.06
5	0.50	0.18	0.10	0.06	0.06	-	0.18	0.10	0.10	0.06	-	0.10
10	-	0.26	0.10	0.06	0.06	-	0.30	0.10	0.10	0.06	-	0.18
15	-	0.46	0.14	0.10	0.06	-	-	0.14	0.10	0.06	-	0.10
20	-	-	0.22	0.14	0.06	-	-	0.22	0.10	0.06	-	0.14
25	-	-	0.30	0.18	0.10	-	-	0.34	0.14	0.06	-	0.26
30	-	-	0.50	0.26	0.14	-	-	-	0.18	0.10	-	0.14
35	-	-	-	0.38	0.14	-	-	-	0.26	0.10	-	0.18
40	-	-	-	-	0.18	-	-	-	0.46	0.14	-	-
45	-	-	-	-	0.26	-	-	-	-	0.18	-	-
50	-	-	-	-	0.38	-	-	-	-	0.26	-	-
55	-	-	-	-	-	-	-	-	-	0.34	-	-

- Reliable detection ($\beta \leq 0.2$) not achieved

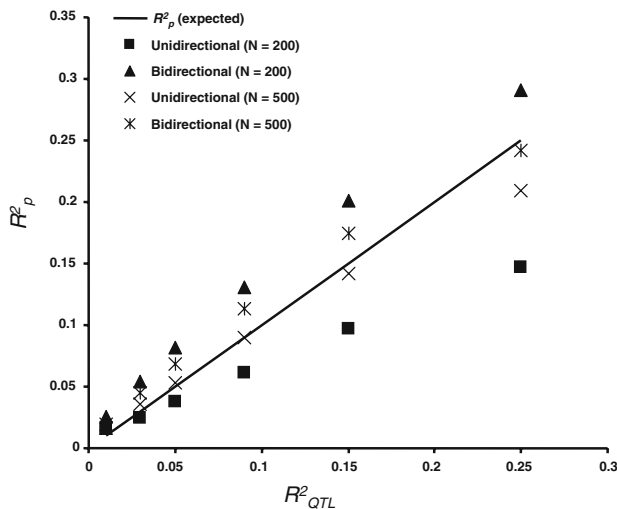


Fig. 2 Estimated versus expected value of the proportion of phenotypic variance accounted for by the QTL (R_p^2) (averaged over 1,000 simulation runs) in bidirectional and unidirectional selective genotyping of 50 lines from a population of 500 or 200 doubled haploid lines for markers coinciding with the QTL with a range of effect sizes (R_{QTL}^2 from 0.01 to 0.25)

of the phenotypic variance could be reliably detected with genotyping of a smaller proportion of lines and (or) by testing at markers at some distance from the QTL. For example, a QTL explaining 25% of the phenotypic variance could be reliably detected by genotyping only 2% of a population of 500 lines at a marker 5 cM from the QTL or even by genotyping 10% of the lines at a marker 35 cM from the QTL (Table 2). Bidirectional selective genotyping in a population of 200 lines was less effective, but still adequate for reliable detection of large-effect QTLs (Table 2).

With unidirectional selection, the number of lines genotyped had little effect on detection power for QTL with very small effects (which were rarely detected) and for QTL with very large effects (which were almost always detected) (Fig. 1). For QTL with intermediate effects, QTL detection power reached a maximum at an intermediate proportion of lines genotyped, and declined as additional lines were genotyped. With unidirectional selection in a population of 500 lines, a QTL explaining as little as 9% of the phenotypic variance could be reliably detected, at most of the selection proportions tried, provided that there was a marker within 10 cM of the QTL (Table 2). QTL explaining a larger proportion of the phenotypic variance could reliably be detected by testing at more distant markers (Table 2). With unidirectional selection in a population of only 200 lines, however, QTL explaining only 9% of the phenotypic variance were not reliably detected, regardless of their proximity to markers (Table 2).

With bidirectional selective genotyping in a population of 500 lines, R_p^2 was usually slightly overestimated. With

bidirectional selective genotyping in a population of 200 lines, R_p^2 was more seriously overestimated, particularly for large-effect QTL. With unidirectional selective genotyping R_p^2 was underestimated, particularly for large-effect QTL and for small population size (Fig. 2).

Rice dataset

Using data from the Vandana/Way Rarem rice population, the QTL on chromosome 12 was detectable using selective genotyping involving small numbers of progeny selected from the tail(s) of the phenotypic distribution. With either marker-based or trait-based analysis, genotyping of 15 low-yielding lines and 15 high-yielding lines was sufficient to detect a QTL at both RM28048 and RM28130, the two markers flanking the 8.4-cM interval on chromosome 12 in which Bernier et al. (2007) mapped a large-effect QTL based on analysis of data from 158 lines (Table 3). Even with only 10 lines genotyped from each tail of the distribution, the QTL was detected at RM28130 and at markers up to 18.1 cM away from the estimated QTL position.

With unidirectional selection, and with trait-based testing of allele frequencies against the expected frequency of 0.5, genotyping of 10 lines from the upper tail of the phenotypic distribution was sufficient to detect the QTL at markers between 4.4 cM (RM28130) and 18.1 cM (RM28166) from the estimated QTL position (Table 4). However, QTL detection at the other flanking marker (RM28048, 4.0 cM from the estimated QTL position) required genotyping of 20 lines.

In the random sample of 92 lines from the Vandana/Way Rarem rice population, segregation distortion was detected at 39% of the marker loci. When allele frequencies in selected tails were tested against the frequencies observed in a random sample of lines, the QTL was detected consistently at RM28130, provided 10 or more lines were genotyped. However, with this test, the QTL was never detected at the other flanking marker (RM28048), even when 30 high-yielding lines were genotyped, even though that marker's allele frequencies did not deviate significantly from 0.5 in the random sample.

In addition to detecting the QTL on chromosome 12, all four selective genotyping approaches declared marker-QTL linkage for the loci at which there was significant segregation distortion and for one locus (RM290 on chromosome 2) at which genotypic frequencies did not deviate significantly from the expectations (Tables 3, 4). The number of loci with significant segregation distortion for which a QTL-marker linkage was declared was higher in unidirectional selection when observed ratios were tested against the expected ratios than when they were tested against the ratios observed in the random sample (Table 4).

Table 3 Microsatellite loci identified by bidirectional selective genotyping as being significantly ($P < 0.01$) associated with grain yield under drought-stress conditions in F_3 -derived lines from the Vandana/Way Karem rice population, when subsets of different number of high- and low-yielding lines were genotyped and QTL detection was performed using trait-based or marker-based analysis

Locus	Map location ^a Chromosome	cM	$\chi^2_{(1,1)}$ ^b	Number of lines genotyped												
				10 ^c	20 ^c	30 ^c	40 ^c	50 ^c	60 ^c	10 ^d	20 ^d	30 ^d	40 ^d	50 ^d	60 ^d	
RM290	2	50.9	0.02	–	0.60	0.50	0.50	0.50	0.42	0.40	–	–234	–193	–198	–176	–164
RM3103	12	34.5	2.25	–	0.08	0.07	0.08	0.08	0.06	0.06	–	–	–	–	207	226
RM7195	12	40.8	1.52	–	–	0.05	0.05	0.06	0.08	0.08	–	–	–	253	301	328
RM28048	12	45.2	1.07	–	–	–	0.06	0.09	0.11	0.11	–	–	187	212	244	278
RM28130	12	53.6	1.91	–	–	0.06	0.07	0.09	0.12	0.12	–	–	241	280	284	322
RM511	12	55.5	2.25	–	0.08	0.08	0.09	0.10	0.10	0.14	–	–	210	265	300	325
RM1261	12	57.5	0.57	–	0.06	0.05	0.08	0.09	0.13	0.13	–	–	211	251	288	316
RM28166	12	67.3	0.57	–	0.07	0.08	0.10	0.12	0.15	0.15	–	–	211	255	291	302
Number of additional loci at which both significant marker-trait association and significant segregation distortion e were detected				0	1	0	0	1	0	0	0	3	3	2	3	0

– Non significant test

^a Bernier et al. (2007) used interval mapping and estimated the position of the QTL to be at the 49.2-cM position of chromosome 12, between RM28048 and RM28130

^b The χ^2 values are for allele frequencies in random sample ($n = 92$) tested against the 1:1 ratio. For all loci presented, ratios conform to 1:1 ratios ($P > 0.05$)

^c Trait-based analysis; values are allele frequency shifts (upper value) and R^2_p

^d Marker-based ANOVA; values are estimated additive effect of QTL (kg ha^{-1})

Table 4 Microsatellite loci identified by unidirectional selective genotyping as being significantly ($P < 0.01$) associated with grain yield under drought-stress conditions in F_3 -derived lines from the Vandana/Way Karem rice population, when subsets of different number of high-yielding lines were genotyped and marker allele frequencies were tested against the expected frequency (0.5) or against allele frequencies estimated by genotyping a random sample of 92 lines

Locus	Map location ^a		$\chi^2_{(1)}$ ^b	Number of lines genotyped											
	Chromosome	cM		5 ^c	10 ^c	15 ^c	20 ^c	25 ^c	30 ^c	5 ^d	10 ^d	15 ^d	20 ^d	25 ^d	30 ^d
<i>RM290</i>	2	50.9	0.02	0.40	0.45	0.37	0.35	0.32	0.28	0.39	0.44	0.37	0.34	0.31	0.28
<i>RM3103</i>	12	34.5	2.25	–	–	–	–	–	–	–	–	–	–	–	–
<i>RM7195</i>	12	40.8	1.52	–	–	–	–	–	–	–	–	–	–	–	–
<i>RM28048</i>	12	45.2	1.07	–	–	–	–	0.26	0.27	–	–	–	–	–	–
<i>RM28130</i>	12	53.6	1.91	–	–	–	–	0.03	0.04	–	–	–	–	–	–
				–	0.40	0.36	0.34	0.33	0.34	–	0.35	0.34	0.29	0.27	0.29
<i>RM511</i>	12	55.5	2.25	–	0.05	0.04	0.05	0.05	0.05	–	0.04	0.04	0.04	0.04	0.04
				–	0.40	0.33	0.33	0.34	0.34	–	0.34	–	0.27	0.28	0.28
<i>RM1261</i>	12	57.5	0.57	–	0.05	0.05	0.04	0.04	0.05	–	0.04	–	0.03	0.04	0.04
				–	0.35	0.30	0.30	0.32	0.32	–	0.32	–	0.27	0.29	0.29
<i>RM28166</i>	12	67.3	0.57	–	0.04	0.04	0.04	0.05	0.05	–	0.03	–	0.03	0.04	0.04
				–	0.35	0.30	0.30	0.33	0.31	–	0.32	–	0.27	0.35	0.28
				–	0.04	0.04	0.04	0.05	0.05	–	0.03	–	0.03	0.05	0.04
Number of additional loci at which both significant marker-trait association and significant segregation distortion were detected				5	5	11	7	9	8	5	2	2	2	0	0

Values are allele frequency shifts (upper value) and R_p^2 (lower value)

– Non significant test

^a Bernier et al. (2007) used interval mapping and estimated the position of the QTL to be at the 49.2-cM position of chromosome 12, between RM28048 and RM28130

^b The χ^2 values are for allele frequencies in random sample ($n = 92$) tested against the 1:1 ratio. For all loci presented, ratios conform to 1:1 ratios ($P > 0.05$)

^c Tested against expected ratios

^d Tested against ratios in random sample

Discussion

In most of the cases investigated here, observed rates of Type I error were at or near the target rate of 0.01. In contrast, we found (results not shown) that other possible tests (those involving estimates of S_q obtained according to the formula of Lebowitz et al. (1987) for bidirectional selective genotyping, the test used by Foolad et al. (2001), and the test used by Zhang et al. (2003)) gave high rates of false positives, probably due to failure of data from selectively genotyped populations to satisfy assumptions on which the tests are based. The formula given by Lebowitz et al. (1987) for bidirectional selection assumes symmetrical changes in marker allele frequencies. The test used by Zhang et al. (2003) assumes equal variances of allele frequencies between the selected sub-set(s) while the test used by Foolad et al. (2001) assumes equal variances between the selected and random samples. In our simulation experiments, the false positive rates were highest when trait-based analysis was applied for bidirectional selective genotyping of only 10 lines. This can be explained by the reliance of this test on a normal approximation of the binomial distribution, which is not adequate for small samples from the extreme tails of a distribution.

Although trait-based and marker-based analyses have more or less the same statistical power, trait based analysis has two advantages relative to marker-based analysis: (1) trait based analysis can readily be adapted to analysis of selective DNA pooling data (Darvasi and Soller 1994), something that is not possible for marker-based analysis, and (2) trait based analysis can deal with the unidirectional selective genotyping with its potential for use within breeding programs, again something that is not possible for marker-based analysis. Moreover, when small numbers of progeny are selected bidirectionally, marker-based analysis is not as powerful as trait-based analysis, even though it uses more information (i.e., trait data in addition to marker data). Trait data may not be useful unless the selected subsets are large enough for precise estimation of means.

When a given number of progeny is selected based on their phenotypic values for a trait of interest, a bidirectional selection strategy will select more progeny that are phenotypically extreme (and therefore genetically informative) than a unidirectional strategy. Therefore, bidirectional selective genotyping can be expected to be more powerful than unidirectional selective genotyping. Nevertheless, unidirectional selective genotyping provided adequate power to detect QTL with moderate to large effects, provided there were marker loci close to the QTL positions. The unidirectional genotyping strategy has the advantage that it can be applied in breeding programs, using lines that have been retained under phenotypic selection, or in cases where only part of a population has survived after exposure

to stress: situations in which bidirectional selective genotyping and full-population mapping are not possible.

With unidirectional selective genotyping, power increased with the size of selected subset to a certain point and decreased thereafter. The optimum selection proportion was 0.2–0.3, which is in agreement with Darvasi and Soller (1992) and Gallais et al. (2007). The decreased power of QTL detection in unidirectional selective genotyping after an optimum selection proportion is due to the increased frequency of lines with alternate genotype in the selected subset. This reduces the value of the shift in marker allele frequencies which in turn reduces the power of QTL detection.

In selective genotyping experiments, segregation distortion (deviation from the expected allele frequency; Zamir and Tadmor 1986, Lyttle 1991) can lead to detection of false associations with the more frequent marker allele or failure to detect true associations with the less frequent allele. With genotyping of a randomly selected subset of the population, it is possible to obtain expected frequencies that are appropriate for the population. Application of this approach in the Vandana/Way Rarem population revealed a high level of segregation distortion (39% of marker loci), which is not unusual in rice (Wan et al. 1996). By testing against the allele frequencies estimated from the random sample, we were able to eliminate many of the apparently spurious associations that were detected using an expected frequency of 0.5. However, most of these were also eliminated with bidirectional selection. Even if allele frequencies differ from expectations, selection will cause them to diverge in the selected tails. Even when there is substantial segregation distortion, the additional resources required to genotype a random subset might be better allocated to bidirectional genotyping.

With bidirectional selective genotyping, it was possible to obtain reasonable estimates of the proportion of phenotypic variation explained at each marker position, provided a large enough population size was used. With unidirectional selective genotyping, this proportion tended to be underestimated. As might be expected, the deviations from expected values were greatest for large-effect QTL: those for which a quasi-infinitesimal model is least appropriate. Consistent with the results of the simulation experiments, estimates of R_p^2 in the rice experiment were greater with bidirectional selection than for unidirectional selection. It should be noted that this estimation method does not account for effects and interactions of other QTL that may affect the trait (Lin and Ritland, 1996; Foolad et al. 2001; Xu et al., 2008), nor does it account for residual dominance that may be a factor in recombinant inbred line populations.

The broad-sense heritability and R_{QTL}^2 values considered in some previous studies (e.g., heritability of 0.05 by Xu

and Vogl (2000) and R_{QTL}^2 of 0.05 by Tenesa et al. (2005)) are low compared to heritability and R_{QTL}^2 values of some agronomically important traits when evaluated in well-designed screens of pure lines in self-pollinated crop species (Holland et al. 2003). For example, based on the variance components reported by Bernier et al. (2007) for the Vandana/Way Rarem rice population, the heritability of grain yield evaluated under severe upland drought stress in single-row plots was 0.45 for selection units consisting of single unreplicated plots, and 0.56 for the means of two replicates in a single trial, and there was a single QTL explaining over 30% of the phenotypic variance. QTL with such large effects are of immediate interest to plant breeders for application in marker-assisted selection. We therefore included large-effect QTL in our simulation experiments. Our results indicate that both marker-based and trait-based analysis of selectively genotyped progeny should be powerful enough to detect QTL with moderate to large effects, with trait heritabilities and selection intensities that are relevant to plant breeding programs.

Selective genotyping is particularly attractive for applications in which the objective is to screen a large sample of potential donors for large-effect alleles, rather than to try to detect many small-effect QTL, or when it is desirable to detect QTL alleles from a donor with effects across a range of backgrounds. Resources that might otherwise be used to conduct both phenotyping and genotyping in one experimental population in a conventional mapping experiment could be reallocated to selective genotyping in several populations. This opens up the possibility of assaying a wide range of germplasm sources for useful QTL alleles. With increasing availability of highly multiplexed and array-based genotyping technologies, consideration can now be given to routine high-density genotyping of small numbers of selected lines from the extremes of breeding populations, providing opportunities to detect QTL with moderate to large effects on traits of particular interest. If progeny with both low and high phenotypic values have been retained, then a bidirectional approach will be preferable to a unidirectional one. In cases where only the superior progeny have been retained, there is still scope for QTL detection via unidirectional selective genotyping, but with a greater risk of both Type I and Type II error.

It is important to note, however, that selective genotyping is limited in its applicability to multiple uncorrelated traits (Lebowitz et al. 1987; Tenesa et al. 2005; Ronin et al. 1998). If selective genotyping is applied on lines selected for more than one trait from breeding populations, detected QTL must be considered to influence a composite trait with adaptive or commercial value.

In practice, a QTL detection strategy based on selective genotyping might initially apply selection at an intensity

appropriate to detect QTL of a certain effect size, given the density of available markers that could be economically assayed on an individual lines. For example, if the aim was to have 80% power to detect QTL explaining 10% or more of the phenotypic variation in a population of 500 lines, and markers were available at 10-cM intervals, 25 lines from each tail of the phenotypic distribution could be genotyped. Subsequently, if significant associations were detected in one or more chromosome regions, markers in those regions could be genotyped across the entire population, for validation purposes and fine-mapping.

Acknowledgments This research was conducted with financial support from research grants provided by the Canadian International Development Agency, Natural Sciences and Engineering Research Council of Canada, the Alberta Agricultural Research Institute, and the Alberta Crop Industry Development Fund. Genetic simulation experiments were conducted in part by using *Perl* scripts written by Hai Pham. We thank Hai Pham and Nicholas Tinker for providing access to this unpublished software and we thank Hai Pham for helping with computer programming. We are also grateful to Chris-Carolin Schön for critical review of an earlier version of the manuscript. We are grateful for the insightful suggestions of several anonymous reviewers.

References

- Bernier J, Kumar A, Ramaiah V, Spaner D, Atlin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Sci* 47:507–518
- Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor Appl Genet* 85:353–359
- Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* 138:1365–1373
- Falconer DS (1989) *Introduction to quantitative genetics*, 3rd edn. Longman, New York
- Foolad MR, Zhang LP, Lin GY (2001) Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. *Genome* 44:444–454
- Gallais A, Moreau L, Charcosset A (2007) Detection of marker-QTL association by studying change in marker frequencies with selection. *Theor Appl Genet* 114:669–681
- Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* 22:9–111
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Lebowitz RJ, Soller M, Beckmann JS (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theor Appl Genet* 73:556–562
- Lee C (2005) Selection bias in quantitative trait loci mapping. *J Hered* 96:363–367
- Lin JZ, Ritland K (1996) The effects of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci. *Theor Appl Genet* 93:1261–1266

- Liu HB (1998) Statistical genomics: linkage, mapping, and QTL analysis. CRC Press, New York, pp 493–517
- Lyttle TW (1991) Segregation distorters. *Ann Rev Genet* 25:511–557
- Mather KN, Jinks JL (1982) Biometrical genetics: the study of continuous variation, 3rd edn. Chapman and Hall, New York, pp 135–175
- Ronin YI, Korol AB, Weller JI (1998) Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis. *Theor Appl Genet* 97:1169–1178
- SAS Institute (2003) Release 9.1. SAS Institute, Inc, Cary
- Steel RGD, Torrie JH, Dickey DA (1997) Principles and procedures of statistics: a biometrical approach, 3rd edn. McGraw Hill Inc, New York
- Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191–203
- Tenesa A, Visscher PM, Carothers AD, Knott SA (2005) Mapping quantitative trait loci using linkage disequilibrium: marker-versus trait-based methods. *Beh Genet* 35:219–228
- Wan J, Yamaguchi Y, Kato H, Ikehashi H (1996) Two new loci for hybrid stability in cultivated rice (*Oryza sativa* L.). *Theor Appl Genet* 92:183–190
- Xu S, Vogl C (2000) Maximum likelihood analysis of quantitative trait loci under selective genotyping. *Heredity* 84:525–537
- Xu Y, Wang J, Crouch J (2008) Selective genotyping and pooled DNA analysis: an innovative use of an old concept. In: Recognizing past achievement, meeting future needs. Proceedings of the 5th international crop science congress, April 13–18, 2008, Jeju, Korea. Published on CDROM, Website <http://www.crops2008.com>.
- Zamir D, Tadmor Y (1986) Unequal segregation of nuclear genes in plants. *Bot Gaz* 147:355–358
- Zhang LP, Lin GY, Niño D, Foolad MR (2003) Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. *Mol Breed* 12:3–19