Mol Breeding (2010) 26:493–511
DOI 10.1007/s11032-010-9390-8

Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement

Yanping Sun · Jiankang Wang · Jonathan H. Crouch · Yunbi Xu

Received: 14 January 2009 / Accepted: 11 January 2010 / Published online: 28 January 2010

Abstract Selective genotyping of individuals from the two tails of the phenotypic distribution of a population provides a cost efficient alternative to analysis of the entire population for genetic mapping. Past applications of this approach have been confounded by the small size of entire and tail populations, and insufficient marker density, which result in a high probability of false positives in the detection of quantitative trait loci (QTL). We studied the effect of these factors on the power of QTL detection by simulation of mapping experiments using population sizes of up to 3,000 individuals and tail population sizes of various proportions, and marker densities up to one marker per centiMorgan using complex genetic models including QTL linkage and epistasis. The results indicate that QTL mapping based on selective genotyping is more powerful than simple interval mapping but less powerful than inclusive composite interval mapping. Selective genotyping can be used, along with pooled DNA analysis, to replace genotyping the entire population, for mapping QTL with relatively small effects, as well as linked and interacting QTL. Using diverse germplasm including all available genetics and breeding materials, it is theoretically possible to develop an "all-in-one plate" approach where one 384-well plate could be designed to map almost all agronomic traits of importance in a crop species. Selective genotyping can also be used for genomewide association mapping where it can be integrated with selective phenotyping approaches. We also propose a breeding-to-genetics approach, which starts with identification of extreme phenotypes from segregating populations generated from multiple parental lines and is followed by rapid discovery of individual genes and combinations of gene effects together with simultaneous manipulation in breeding programs.

Keywords Selective genotyping · Pooled DNA analysis · Genetic mapping · Inclusive composite interval mapping · Marker-assisted selection

Two contrasting approaches have been routinely used for marker-trait association analysis (viz. genetic mapping): (1) testing the average phenotypic

Y. Sun · J. Wang
Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, 100081 Beijing, China

Y. Sun · J. Wang
Crop Research Informatics Lab., CIMMYT China, Chinese Academy of Agricultural Sciences, 100081 Beijing, China

J. H. Crouch · Y. Xu (✉)
International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, DF, Mexico
e-mail: y.xu@cgiar.org

difference between groups of individuals with distinct marker genotypes and (2) comparing marker allele frequencies amongst groups of individuals with distinct phenotypes. The first approach is usually based on genotyping an entire germplasm collection or segregating population with markers evenly covering the genome (Edwards et al. 1987; Soller and Beckmann 1990). However, this approach is extensive, time-consuming and expensive, while generating precision multilocational phenotype data at this scale may be logistically difficult or even impossible for some traits. Unfortunately, reducing the overall size of a mapping population will, in general, decrease the power of QTL detection (Charcosset and Gallais 1996), and increase the confidence interval related to the estimated position of the QTL, as well as increasing the probability of detecting false positive QTL. The second approach provides a partial solution to this problem by focusing on the individuals from the high and low tails of the phenotypic distribution across the germplasm collection or segregating population ('selective genotyping'; Lebowitz et al. 1987; Lander and Botstein 1989). In selective genotyping, marker analysis only needs to be carried out on individuals from those tails, and further economic savings can be made by generating pooled DNA from groups of individuals with similar phenotypes. Marker-trait association is then inferred by analyzing the differences in allelic frequency between the two tails (Stuber et al. 1980,

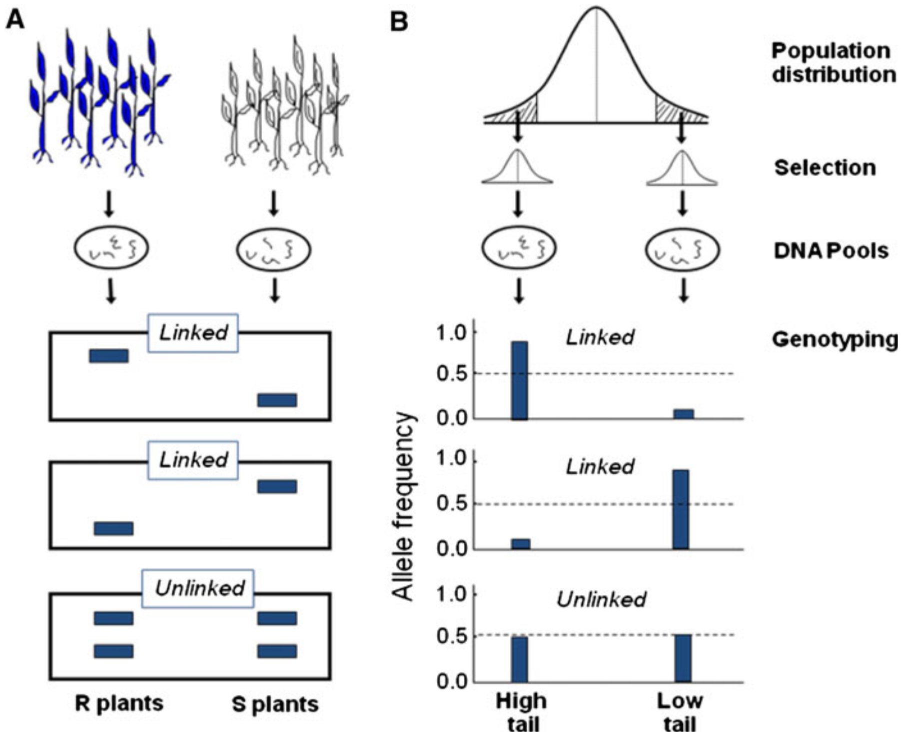

Fig. 1 Selective genotyping and pooled DNA analysis. **A** Pooled analysis using disease resistance (R) and susceptive (S) plants as example. DNA pools are constructed from R and S plants selected from a mapping population and then genotyped by molecular markers. When two DNA pools show different alleles at a specific marker locus, the marker is linked with the disease response, while when the both pools show the same heterozygous genotype, the marker is unlinked with the disease response. **B** Pooled DNA analysis using extreme plants selected for a target quantitative trait from two tails of a normal distribution in the mapping population. Marker-trait linkage is revealed by allele frequency at specific marker loci. When allele frequencies at a marker locus are significantly different between the two pools at a marker locus, the marker is linked with the target trait, while when the allele frequencies are very close to each other (each approximately 0.5), the marker is unlinked with the target trait. In both **A** and **B**, assume that the marker is dominant and reveals polymorphism between the parental lines that are used to derive the mapping population

1982) or the difference in signal strength between the DNA pools from those two tails (‘pooled DNA analysis’) (Fig. 1). Of course, this approach does not reduce the need (and associated cost) for accurate phenotyping of the entire original population in order to accurately identify the individuals with genotypes for extreme phenotypes of the target trait.

The underlying approach of selective genotyping has also been used in ‘tail analysis’ (Hillel et al. 1990; Dunninnton et al. 1992; Plotsky et al. 1993), ‘bulked segregant analysis’ (Giovannoni et al. 1991; Michelmore et al. 1991), and ‘selective DNA pooling’ (Darvasi and Soller 1994). It can be bidirectional if the two tails of the distribution are considered, or unidirectional if only one tail is considered (Navabi et al. 2009). The unidirectional approach is more suitable for traits that have been subjected to strong negative or lethal selection pressure in unfavorable environments. But, bidirectional selective genotyping is generally more effective and commonly used in practice as the effect of segregation distortion can be properly avoided. Selective genotyping and pooled DNA analysis have been widely used in genetic mapping in plants with numerous reports for single major genes (e.g., Barua et al. 1993; Hormaza et al. 1994; Villar et al. 1996; van Treuren 2001; Zhang et al. 2002) and for detection and validation of quantitative trait loci (QTL) (Foolad and Jones 1993; Zhang et al. 2003; Wingbermuehle et al. 2004; Coque and Gallais 2006), including traits controlled by a few major-effect QTL (Quarrie et al. 1999). This approach has also been used to detect significant changes in marker allele frequency through two cycles of recurrent selection (Moreau et al. 2004). A fractioned DNA pooling approach has also been used in which the tails of the population distribution are randomly allocated among a number of independent sub-pools (Sham et al. 2002; Brohede et al. 2005; Korol et al. 2007; Shifman et al. 2008).

Selective genotyping and pooled DNA analysis have been shown to have significant advantages in terms of cost savings, compared to entire population analysis, with negligible practical disadvantages in terms of power of detection in medical genomics research (Knight and Sham 2006; Macgregor et al. 2008). For example, for analysis of a large population with 1,000 individuals where 30 individuals from each tail are selected, selective genotyping will only cost 6% of that required for genotyping the entire population. Combining this with pooled DNA analysis would provide a substantial further saving in genotyping costs now equating to 0.2% of the cost of individually genotyping the entire population. Clearly, the larger the original population size, the greater the power of this approach and the higher the savings compared to entire population genotyping.

For small to moderate sized populations ($n = 200$–500), the optimum size for each tail, in terms of power of detection, is 20–30% of the entire population (Darvasi and Soller 1992; Gallais et al. 2007; Navabi et al. 2009). As the population size increases, the proportion of individuals required for a given power of QTL detection will decrease such that an absolute optimum number of plants from each tail can be defined. Gallais et al. (2007) simulated the detection of marker-trait association by studying changes in marker frequencies amongst groups of individuals with distinct phenotypes and analyzed the effect that different ratios of genotyping and phenotyping had on QTL detection power and overall cost. They found that the optimum size of the tail population (as a proportion of the entire population), which determines the level of cost savings that can be realized by this approach, is mainly determined by the ratio between cost of genotyping and cost of phenotyping.

There are several issues that have confounded many applications of selective genotyping and pooled DNA analysis (Xu and Crouch 2008): (1) a relatively small number of markers has often been used to cover the whole genome with the assumption that genes can be readily identified using a marker density of 15–25 cM, while frequently this is not feasible; (2) contrasting individuals have been selected from a relatively small-size population (e.g., 100–300 individuals), this reduces the power of QTL detection such that only large-effect genes/QTL may be detected, which depending on the germplasm used may not exist for many complex agronomic traits; (3) when the allele signal is determined through a gel-based genotyping system, allele frequency in each pool cannot be quantified accurately and the signal generated by a rare allele present in only a small number of individuals of a pool may not be detected which substantially reduces the power of the approach; (4) pools are often based on a relatively small number of individuals (10–15) from each tail, causing a high level of false positive marker-trait associations. These confounding factors have led to an apparent mixed success of selective genotyping and pooled DNA analysis in literature.

It can be hypothesized that plants with extreme phenotypes chosen for selective genotyping would be those with an accumulation of favorable alleles from multiple loci with various additive effects. However, several significant issues remain to be resolved before the full potential of selective genotyping and pooled DNA analysis can be achieved, including: (1) can selective genotyping be used to replace entire population analysis for both qualitative and quantitative traits; (2) how many independent genes can be identified simultaneously; (3) can selective genotyping be used for mapping linked genes and/or genes with epistatic interactions; and (4) can selective genotyping be used for fine mapping to the level required for map-based cloning? The objectives of this study were to use computer simulation to better understand the effects of potential confounding factors on the power of selective genotyping, including entire population size, relative and absolute tail size, and marker density under various genetic models (including linkage and interaction effects between target QTL as well as different levels of the phenotypic variation explained by the target QTL), by comparison with other QTL mapping methods based on entire population analysis. Our assumption for selective genotyping is that allelic frequencies could be determined for the two selected tails either based on individual genotyping or pooled DNA analysis. Based on this analysis we also provide a detailed discussion of potential uses of selective genotyping (and pooled DNA analysis) in genetic analyses and plant breeding.

Materials and methods

Genetic models and mapping populations used in simulations

Simulations were based on a genome consisting of 10 chromosomes, each of 150 cM in length with four levels of marker density (MD); one marker every 1, 2, 5 or 15 cM. Markers were assumed to be evenly distributed on each chromosome, so the actual number of markers per chromosome at the four MD levels was 151, 76, 31 or 11, respectively. A recombinant inbred line (RIL) population was chosen as the mapping population for this study. The genetic effect of a QTL is measured as the proportion of phenotypic variation explained by that QTL, abbreviated as PVE (%). By definition, PVE of a QTL is the genetic variation caused by the QTL, i.e., V_Q,

Table 1 Six independent QTL used in simulation study and their distances to the nearest markers under four marker densities (MD)

QTL	Chr.	Position (cM)	Additive effect	PVE (%)	Distance of each QTL to its nearest marker under four marker densities (MD)			
					MD = 1 cM	MD = 2 cM	MD = 5 cM	MD = 15 cM
IQ1	1	18	0.1000	1	0	0	2	3
IQ2	2	28	0.1732	3	0	1	2	3
IQ3	3	34	0.2236	5	0	0	1	4
IQ4	5	48	0.2646	7	0	0	2	3
IQ5	7	22	0.3162	10	0	1	2	7
IQ6	9	76	0.3873	15	0	1	2	2

Table 2 Genetic effects of two pairs of linked QTL in the coupling and repulsive linkage phases

QTL	Chr.	Position (cM)	PVE (%)	Additive genetic effect	
				Model 1 (coupling linkage)	Model 2 (repulsive linkage)
LQ1	3	12	5	0.2236	0.2236
LQ2	3	32	5	0.2236	−0.2236
LQ3	5	12	5	0.2236	0.2236
LQ4	5	62	5	0.2236	−0.2236

Table 3 Genetic effects of the three models of digenic interaction used in the simulated study

QTL	Chr.	Position (cM)	Model 1		Model 2		Model 3	
			EQ1	EQ2	EQ1	EQ2	EQ1	EQ2
EQ1	1	18	0.2236		0.0000		0.2739	
EQ2	2	33	0.2236	0.2236	0.3873	0.0000	0.2739	0.0000

divided by the total phenotypic variation, V_P, i.e., $\mathrm{PVE}_Q = \frac{V_Q}{V_P} \times 100\%$.

Three QTL models were compared; (1) independent QTL (Table 1), (2) linked QTL (Table 2), and (3) QTL with epistatic effects (Table 3). For the independent QTL, our purpose was to investigate the detection power of selective genotyping for QTL with various genetic effects. We assumed a quantitative trait was controlled by a total of six QTL distributed on chromosomes 1, 2, 3, 5, 7 and 9, and the PVE of these QTL were 1, 3, 5, 7, 10, and 15%, respectively (Table 1). The distance of QTL to the nearest marker changed with marker density, ranging from 0 to 7 cM (Table 1). When linkage and epistasis are ignored, the total genetic variance is the sum of genetic variances from all QTL. For the independent QTL model in Table 1, the total genetic variance is $V_g = 0.41$, and the random error variance is $V_\varepsilon = 0.59$. Therefore, the phenotypic variance is $V_P = 1.0$, and the broad-sense heritability is $V_g/V_P\,(\%) = 41\%$. The additive effect of each QTL used in the simulation is, therefore, equal to the square root of PVE_Q (Table 1). In addition, two selective genotyping strategies were compared for a QTL with PVE = 10%: (1) 'rough' mapping with population size = 200, marker density = 15 cM and selected proportion (SP) = 5%; and, (2) 'fine' mapping with population size = 500, marker density = 1 cM and SP = 5%.

For the linked QTL, our purpose was to investigate the effect of linkage distance and linkage phase on QTL detection when using selective genotyping. Thus, we considered a trait controlled by two pairs of linked QTL, where LQ1 and LQ2 were linked on chromosome 3 at a distance of 20 cM apart, and where LQ3 and LQ4 were linked on chromosome 5 at a distance of 50 cM apart. When linked in coupling phase, the four QTL have the same amount and direction of genetic effect; but when linked in repulsion phase, LQ1 and LQ2 have the same amount but opposite genetic effects, and similarly for LQ3 and LQ4 (Table 2). The genetic effects of the four QTL given in Table 2 were used in the simulation together with a random error variance $V_\varepsilon = 0.80$, resulting in a broad-sense heritability of 20% if the four QTL were unlinked. It should be noted that the total genetic variance is not equal to the summation of individual genetic variances under linkage (Li et al. 2008). For the QTL linked in coupling phase, genetic variation is greater than the sum of individual QTL variations, leading to a higher heritability. For QTL linked in repulsion phase, the genetic variation is less than the sum of individual QTL variance, leading to a lower heritability. The broad-sense heritabilities for the two models in Table 2 were 25.48 and 13.64%, respectively. To understand the effect of population size and marker density on the power of QTL detection, two linked QTL (20 cM apart), each of PVE = 5%, were mapped using different entire population sizes, tail sizes that were different proportions of the entire population and different marker densities.

For the QTL with epistatic effects, we considered a trait controlled by two interacting QTL distributed across two chromosomes, i.e., EQ1 and EQ2 on chromosomes 1 and 2 (Table 3), respectively. In this case, the total genetic variance $V_G = a_1^2 + a_2^2 + aa^2$, where a_1 and a_2 are the additive effects of the two interacting QTL, and aa is the additive by additive epistatic effect between the two QTL. In the epistasis model, we assume EQ1 and EQ2 together explain 15% of the phenotypic variance. In Epistasis Model 1, all genetic effects are present, i.e., $a_1 = a_2 = aa = 0.2236$, each accounting for 5% of the phenotypic variation. In Epistasis Model 2, only the epistatic effect is present, i.e., $a_1 = a_2 = 0$, $aa = 0.3873$, where epistasis is the only genetic effect influencing phenotypic variation and accounts for 15% of the phenotypic variation. In Epistasis Model 3, one additive effect and the epistatic effect are present, i.e., $a_1 = 0.2739$, $a_2 = 0$, and $aa = 0.2739$ (Table 3), each of which account for 15% of the phenotypic variation. The random error variance is $V_\varepsilon = 0.85$ for the three epistasis models.

Entire population sizes ranged from 100 to 3,000 with 100, 150, 200, 250, 300, 350, 400, 500 and 600 steps used for most models. Tail population sizes used represented SP of the entire population from 5 to 50% for each tail (i.e., 5, 10, 15, 20, 25, 30, 35, 40, and 50%). RIL populations were simulated by crossing two inbred parental lines, and QTL mapping was carried out using QTL IciMapping software (Li et al. 2007; Wang 2009; available from http://www.isbreeding.net). A total of 100 simulation runs were conducted for each combination of population size and selected proportion.

QTL detection based on selective genotyping

For selective genotyping, a *t*-test comparing the marker frequency in each selected tail population is normally used to analyze the association between QTL and markers. In contrast, a likelihood ratio test is normally used in interval mapping-based methods to compare mean phenotypes of groups of individuals with the same genotype. In order to compare the two approaches in this study, a likelihood ratio test was derived for selective genotyping. Thus, in the two-tail selective genotyping situation used in this study (Table 4), p_{H} and p_{L} represent the frequencies of a marker allele (i.e. M in Table 4) in the two tail populations. The null hypothesis is H_0: $p_{\mathrm{H}} = p_{\mathrm{L}}$, indicating that there were no QTL associated with the marker, and the alternative hypothesis is H_{A}: $p_{\mathrm{H}} \neq p_{\mathrm{L}}$, indicating the association between QTL and the marker. The number of the plants with two marker types in each tail follows a binomial distribution, and the likelihood function under H_{A} is therefore, $L_A = C_{n_{\mathrm{H}}}^{n_{1\mathrm{H}}}(p_{\mathrm{H}})^{n_{1\mathrm{H}}}(1-p_{\mathrm{H}})^{n_{2\mathrm{H}}} \times C_{n_{\mathrm{L}}}^{n_{1\mathrm{L}}}(p_{\mathrm{L}})^{n_{1\mathrm{L}}}(1-p_{\mathrm{L}})^{n_{2\mathrm{L}}}$.

Table 4 The two-tail selective genotyping

Marker type	MM	mm	Sum
The high tail			
Sample size	$n_{1\mathrm{H}}$	$n_{2\mathrm{H}}$	n_{H}
Frequency of M	p_{H}	$1 - p_{\mathrm{H}}$	1.0
The low tail			
Sample size	$n_{1\mathrm{L}}$	$n_{2\mathrm{L}}$	n_{L}
Frequency of M	p_{L}	$1 - p_{\mathrm{L}}$	1.0

The two marker alleles segregating in the two parental lines were represented by M and m

The likelihood function under H_0 is, $L_0 = C_{n_{\mathrm{H}}}^{n_{1\mathrm{H}}}(p_0)^{n_{1\mathrm{H}}}\ (1-p_0)^{n_{2\mathrm{H}}} \times C_{n_{\mathrm{L}}}^{n_{1\mathrm{L}}}(p_0)^{n_{1\mathrm{L}}}(1-p_0)^{n_{2\mathrm{L}}}$, where $p_0 = \frac{n_{1\mathrm{H}}+n_{1\mathrm{L}}}{n_{\mathrm{H}}+n_{\mathrm{L}}}$, representing the frequency of the marker under H_0. Therefore, the LOD score can be defined as the natural logarithm of the ratio of the two likelihoods, as defined in other QTL mapping methods. The above procedure was implemented using the QTL IciMapping software.

Comparison with empirical mapping data from a barley mapping population

A barley mapping population, derived from a cross between two-row barley (*Hordeum vulgare* L.) genotypes (Harrington × TR306), was used to provide an empirical comparison with the simulated mapping results from selective genotyping versus entire population mapping. The mapping population consisted of 145 random doubled haploid (DH) lines (Tinker et al. 1996). A linkage map was constructed using 127 markers covering all seven chromosomes with a total map length of 1,274 cM. This mapping population was evaluated in 1992 and/or 1993 at 17 locations (Tinker et al. 1996) providing average kernel weight data from 25 environments. The average kernel weight was 38.7 mg for Harrington, and 45.0 mg for TR306. The minimum, mean and maximum kernel weight of the 145 DH lines were 35.8, 42.0, and 48.1 mg, respectively. The family-level broad-sense heritability was 0.71 for this population (also see Li et al. 2008).

For comparison, other QTL mapping methods such as simple interval mapping (SIM; Lander and Botstein 1989) and inclusive composite interval mapping (ICIM; Li et al. 2007; Zhang et al. 2008; Wang 2009), were applied to the simulated and empirical datasets. For ICIM, marker selection was conducted only once through stepwise regression by considering all marker information simultaneously, and the phenotypic values were then adjusted by all markers retained in the regression equation except the two markers flanking the target mapping interval. In the first step of ICIM for additive QTL, a probability value for entering variables (PIN) of 0.01, and a probability value for removing variables (POUT) of 0.02 were used to select the significant markers. There was no background genetic control when SIM was used. The threshold LOD of 2.5 was used to declare significant QTL for all methods.

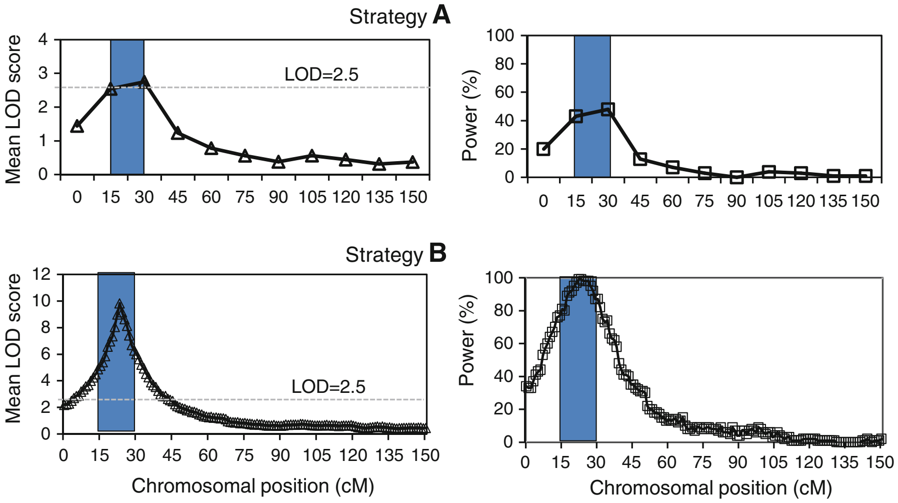

Fig. 2 Effects of selective genotyping strategies on detection power and mean LOD score around the target region (15 cM, *grey area*) assuming the QTL explain 10% of phenotypic variation. Strategy **A** population size = 200, selected proportion = 0.05 (bidirectional selection), resulting no clear peak because the QTL are located in the middle of two makers; Strategy **B** population size = 500, selected proportion = 0.05 (bidirectional selection), marker density = 1 cM, resulting in multiple markers showing positive in the target region with LOD = 9.82 and power = 99%, which is proposed for selective genotyping-based fine mapping

Results

Use of selective genotyping for rough mapping and high-resolution mapping

The effect of marker density on the power of QTL detection was studied by comparing the outcomes of selective genotyping for two very different mapping goals (Fig. 2). In the first instance, conventional selective genotyping was used for rough mapping (Fig. 2, Strategy A), where relatively small entire population ($n = 200$) and tail sizes ($n = 10$) were used with a low density of marker coverage (one marker every15 cM). In this case, selective genotyping resulted in the detection of only one marker in the target region (with an average LOD score just above 2.5). Moreover, the power of QTL detection of 48% observed in this situation cannot distinguish a false positive without validation through genotyping the entire population. In contrast, where large entire population ($n = 500$) and tail sizes ($n = 25$) were used for high-resolution mapping (Strategy B, Fig. 2) along with a high density of marker coverage (one marker per cM), selective genotyping resulted in the detection of multiple markers around the target region with the highest having a LOD score of 9.82 and a power of detection of 99%. Although the region with a LOD score >6 spans more than 10 cM (Fig. 2, Strategy B), there is a sharp peak for LOD scores within a 3-cM region that directly brackets the target QTL. This suggests that selective genotyping can be used for fine mapping when a high-density marker map is available. The accumulative probability of finding false positives decreases proportionally with an increase in the number of markers that simultaneously show significant association. Thus, there is no need to confirm associations identified by high-resolution mapping through genotyping the entire population, which has been the routine procedure for putative markers identified through bulked segregant analysis.

Factors that influence the QTL detection power of selective genotyping

The power of detecting QTL with various genetic effects (listed in Table 1) under different scenarios for the proportion of the entire population selected for each tail population (SP) and different marker densities is shown in Fig. 3, for entire population sizes of 250 and 500. Differences in power of detection can be seen at four levels of marker density (one marker ever 1, 2, 5 or 15 cM) for IQ1 which is the QTL with the smallest effect. When a marker

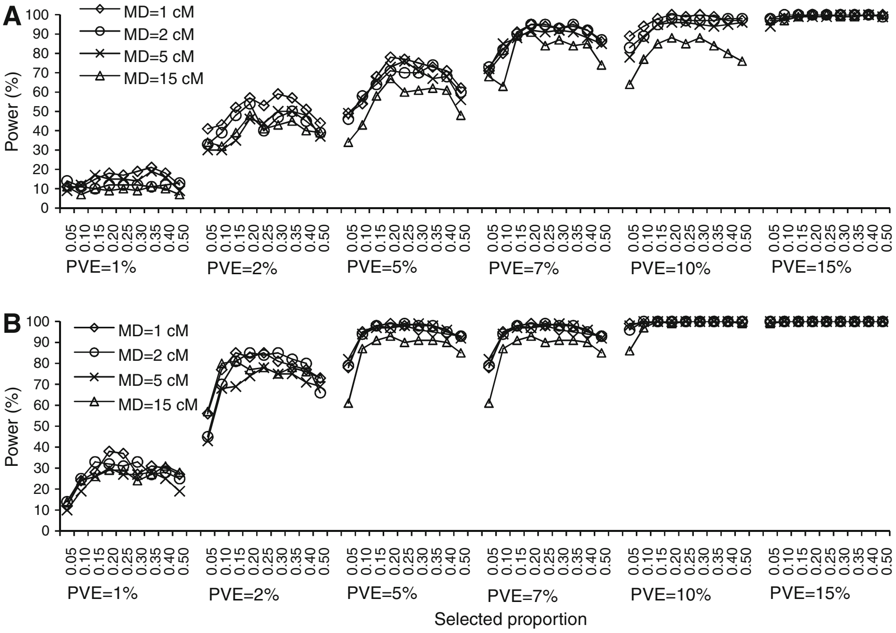

Fig. 3 Power of bidirectional selective genotyping at four levels of marker densities (marker densities = 1, 2, 5 and 15 cM) and nine levels of proportion of the entire population selected for each tail population = 5, 10, 15, 20, 25, 30, 35, 40 and 50%) calculated from 100 simulation runs. **A** Population size = 250; **B** population size = 500. *MD* marker density, *PVE* phenotypic variation explained

density of 15 cM is used, there is a low power of detection for IQ3 and IQ5, even though these two QTL are fairly well separated (as defined in Table 1). Darvasi et al. (1993) reported that a QTL located at the mid-point between flanking markers is the most difficult scenario for QTL detection. This is confirmed in our simulation for PVE = 10% at a marker density of 15 cM where the distance between the QTL and markers is the largest (7 cM) (Table 1). For QTL with other effects at various marker densities, the detection powers are very similar, providing the proportion of tail population to entire population is maintained the same (Fig. 3). When the population size is 250, the power of detection for all methods was over 90% for QTL with PVE = 15% (Fig. 3A). When the population size is 500, however, the power of detection for most methods was around 90% for much smaller QTL with PVE = 5%. Only 5% of the plants from an entire population of 500 need to be included in each tail for the power of detection of selective genotyping to reach 100% for a QTL with PVE = 15%, irrespective of marker density (within the range tested; 1–15 cM) (Fig. 3B).

As expected, power of detection increases with increasing QTL effect (higher PVE; Fig. 3) and population size (Fig. 4). For a population size over 400 and a marker density of 5 cM with PVE = 5% and SP = 5%, power of detection is over 80% (Fig. 4). It can be seen from Figs. 3 and 4 that for most independent QTL, the detection power is maximized with SP = 25%. Selecting more than 25% of the entire population for each tail population does not significantly improve the detection power (Figs. 3 and 4). In some cases, the detection power may even be reduced at higher SP levels (Fig. 4), due to confounding effects of mixed genotypes within the same tail particularly when the population size is too small.

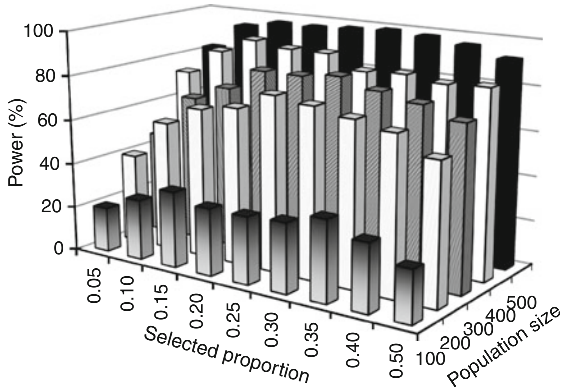

Fig. 4 Power of selective genotyping for various population sizes and selected proportion with phenotypic variation explained = 5% and marker density = 5 cM

For convenience, the SP level that maximizes the detection power of selective genotyping at a marker

Table 5 The proportion of the entire population selected for each tail population where the maximum detection power was achieved at the marker density of 5 cM in bidirectional selective genotyping

PS	PVE = 1%		PVE = 3%		PVE = 5%		PVE = 7%		PVE = 10%		PVE = 15%	
	SP	Power (%)	SP	Power (%)	SP	Power (%)	SP	Power (%)	SP	Power (%)	SP	Power (%)
100	0.30	8	0.35	22	0.35	36	0.20	46	0.35	66	0.35	84
150	0.20	9	0.25	33	0.30	58	0.25	59	0.25	83	0.15	96
200	0.10	13	0.25	39	0.25	78	0.20	80	0.25	92	0.10	95
250	0.35	19	0.30	50	0.25	76	0.20	92	0.15	95	0.10	99
300	0.15	15	0.25	57	0.30	85	0.25	93	0.20	98	0.10	99
350	0.25	27	0.25	55	0.25	86	0.15	95	0.10	97	0.05	98
400	0.25	33	0.30	72	0.15	92	0.15	96	0.05	97	0.05	97
500	0.20	30	0.25	78	0.15	97	0.10	98	0.05	100	0.05	100
600	0.30	35	0.20	90	0.10	97	0.05	95	0.05	100	0.05	100
700	0.25	46	0.20	96	0.10	98	0.05	100	0.05	100	0.05	100
800	0.25	56	0.25	93	0.05	95	0.05	98	0.05	100	0.05	100
1,000	0.25	60	0.15	98	0.05	97	0.05	100	0.05	100	0.05	100
1,500	0.20	83	0.05	99	0.05	100	0.05	100	0.05	100	0.05	100
2,000	0.25	92	0.05	100	0.05	100	0.05	100	0.05	100	0.05	100
2,500	0.15	95	0.05	100	0.05	100	0.05	100	0.05	100	0.05	100
3,000	0.10	95	0.05	100	0.05	100	0.05	100	0.05	100	0.05	100

PVE phenotypic variation explained, *PS* population size, *SP* selected proportion

density of 5 cM has been summarized in Table 5. In large mapping populations (up to 3,000), the QTL detection power can reach 100%, even for QTL explaining 1–3% of the phenotypic variation. For example for IQ2 ($a = 0.1732$) (PVE = 3%) as shown in Table 1, the detection power is only 90% when for a population size of 600 with SP = 20%, but the detection power reaches 100% for a population size of 2,000 with SP = 5% even though the absolute number of individuals genotyped is lower in the latter scenario (Table 5). For QTL with larger effect, say IQ5 ($a = 0.3162$) and IQ6 ($a = 0.3873$), a much smaller population is required to reach the power of 100% (Table 5). Thus, as expected, it is much easier for selective genotyping to detect major QTL, which is also the case for all other mapping methods. When PVE is 5% or larger, similar power of detection can be obtained for some QTL with a large population but a small SP, or a small population with a large SP. Taking IQ5 (PVE = 10%) as an example, for a population size of 250 the detection power of 95% is achieved when SP = 15%, but for a population size of 400 a higher detection power of 97% is achieved at a lower SP = 5% (Table 5). Thus, across most scenarios, genotyping costs can be reduced while at the same time increasing QTL detection powers by increasing the entire population size and decreasing the SP.

Since increasing the SP above 25% does not improve detection power in moderate to large population sizes (Figs. 3, 4; Table 5), in the following sections, we have used a population size of 500 and a marker density of 5 cM with SP = 25% for most comparisons of selective genotyping mapping (SGM) with other available methods, i.e. simple interval mapping (SIM) and inclusive composite interval mapping (ICIM).

Comparison of selective genotyping with other methods for mapping independent QTL

For QTL with PVE = 5%, detection powers of SGM, SIM, and ICIM were simulated by the QTL IciMapping software, as shown in Fig. 5. For population sizes of 500 or larger, all methods have a detection power of 100% for a QTL with PVE = 5% (Fig. 5). For smaller population sizes, ICIM has the highest detection power, which reflects the gains from using model selection in controlling background genetic variation (Li et al. 2007; Zhang et al. 2008). Thus

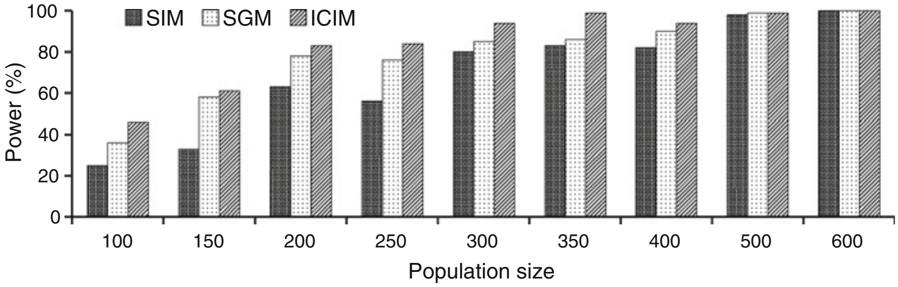

Fig. 5 Comparison of QTL detection power in 100 simulation runs of a QTL with PVE = 5% when using ICIM, SIM and SGM (selective genotyping under bidirectional selection) across a range of population sizes by fixing marker density at 5 cM. For SGM the proportion of the entire population selected for each tail population is where the maximum power can be achieved. *ICIM* inclusive composite interval mapping, *SIM* simple interval mapping, *SGM* selective genotyping mapping

where it is possible to use moderate to large mapping populations there is no loss of power for detected unlinked QTL from the application of selective genotyping.

Comparison of selective genotyping with other methods for mapping linked QTL

For linked QTL (Table 2), LOD scores and detection powers of SGM, SIM, and ICIM were simulated. The average LOD scores are shown in Fig. 6A, C, E, G, and detection powers are shown in Fig. 6B, D, F, H, for the chromosome where two linked QTL are located. Two clear peaks can be seen on the LOD profile from ICIM for the two linkage distances and two linkage phases, indicating the high power of ICIM to dissect linked QTL. For both SGM and SIM, the distinction of two peaks at the true QTL positions was obscure in the LOD profile when the linkage distance was 20 cM (Fig. 6A, E). When the linkage distance was 50 cM, two peaks around the true QTL positions can be seen, but there is still a substantial overlap (Fig. 6C, G), which will lead to a much wider confidence interval for these QTL in the resultant map.

Marker interval-based power analysis (Li et al. 2007) confirms the property of LOD profile of each method. ICIM has high powers of detection around the two QTL positions for the two linkage distances and two linkage phases, but rather low in those marker intervals where no QTL were located (Fig. 6B, D, F, H). This means that ICIM can have a high detection power combined with a low rate of false positives. For the linkage phase of coupling, SGM and SIM have rather higher detection powers than ICIM across the marker intervals where QTL were located but the rate of false positives is also high (Fig. 6B, D). For the linkage phase of repulsion, the detection powers of SGM and SIM are lower than ICIM (Fig. 6F, H). Therefore, the use of SGM may not be recommended when dissecting linked QTL, unless a large population size with a large tail size is used with a high density of markers.

Genetic separation of two tightly linked QTL has been a challenge for most, if not all, statistical methods even when using the entire population genotyping approach. Using population sizes much larger than those used in Fig. 5, with marker densities from 1 to 15 cM, and four different SP levels, Fig. 7 provides the result for two QTL linked at a distance of 20 cM. At a marker density of 1 cM, the two target regions (spanning 4 cM) are associated with two peaks for both detection power and mean LOD score, although there are LOD scores above the 2.5 threshold across the entire region containing the two QTL. Among the five scenarios studied through simulation analysis (Fig. 7A), those based on larger entire population sizes and high SP values show not only stronger association but also more distinguishable peaks for two linked QTL, compared to results from smaller entire population sizes and lower SP values. However, at a marker density of 15 cM, the two linked QTL could not be separated at all (Fig. 7B). These results indicate that when the population size is larger than 500 and marker density is 1 cM, two linked QTL 20 cM apart could be separated by selective genotyping. In contrast, none of the methods tested could separate two QTL tightly linked (5 cM apart), irrespective of the population size or marker density tested (data not shown).

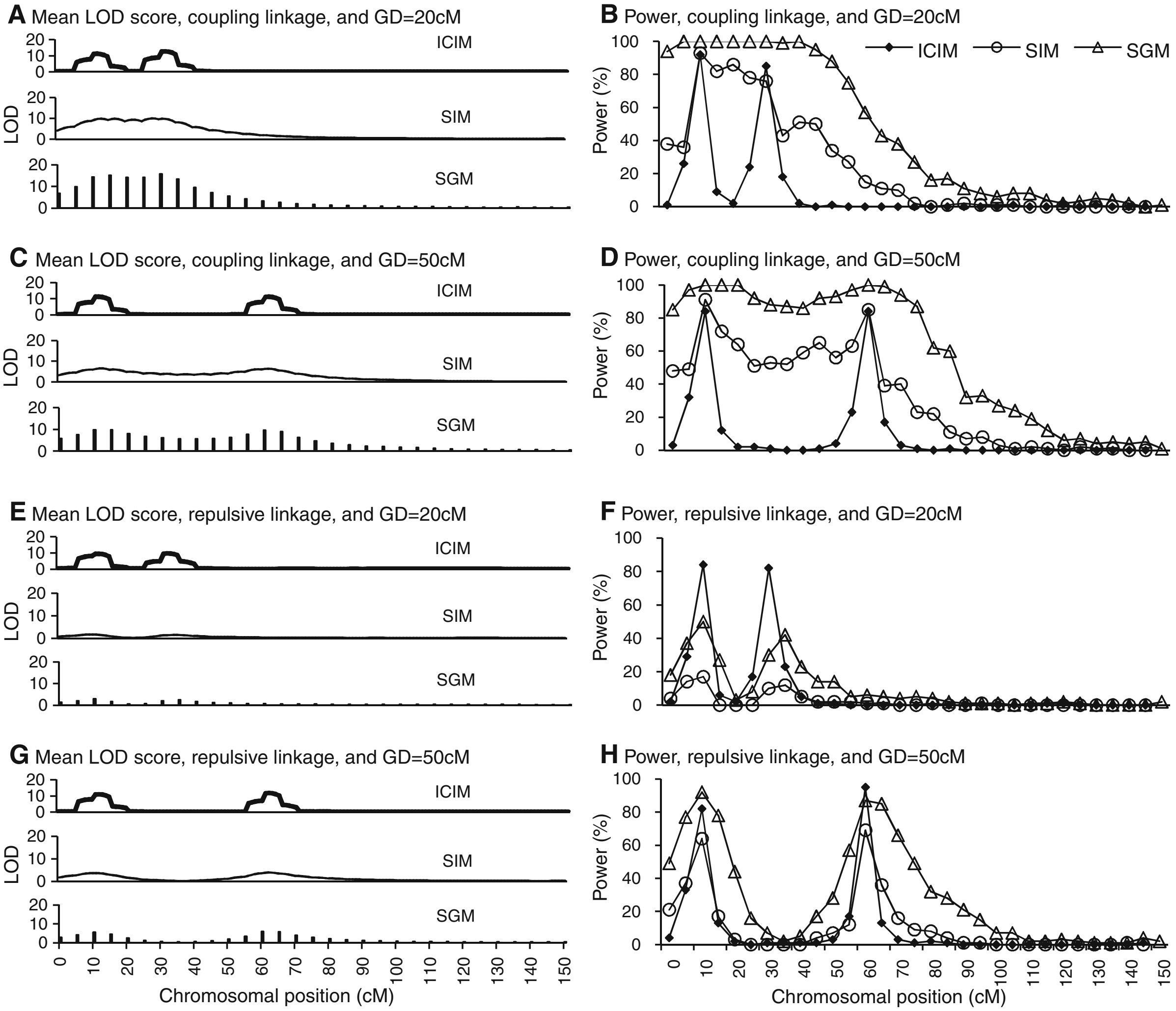

Fig. 6 Mean LOD score (**A**, **C**, **E**, **G**) and power of QTL detection (**B**, **D**, **F**, **H**) from ICIM, SIM and SGM (selective genotyping under bidirectional selection) for two linked QTL at two genetic distances (GD = 20 and 50 cM). Each of the two QTL explains 5% of phenotypic variance. The population size is 500 and proportion of the entire population selected for each tail population is 25% for bidirectional selective genotyping. *ICIM* inclusive composite interval mapping, *SIM* simple interval mapping, *SGM* selective genotyping mapping, *GD* genetic distance

Comparison of selective genotyping with other methods for mapping epistatic QTL

For epistatic QTL (Table 3), the average LOD scores are shown in Fig. 8A, C, E, and the detection powers are shown in Fig. 8B, D, F, for the two chromosomes where the two epistatic QTL are located. Two clear peaks can be seen on the LOD profile generated by each method, when both QTL have additive effects (Fig. 8A). In contrast, no peaks were observed when additive effects were absent (Fig. 8B). When one of the two additive effects was present, only one peak around the position of the additive QTL was observed (Fig. 8C). Interestingly, SGM has the highest power to detect epistatic QTL as shown by its higher LOD score (Fig. 8A, E). As with the results on independent and linked QTL, ICIM generates much sharper peaks, indicating the narrower confidence interval of ICIM.

Comparative analysis of the three methods using empirical data from a barley DH population

In a barley DH population of 145 individuals, genotyped by 127 markers, ICIM identified nine

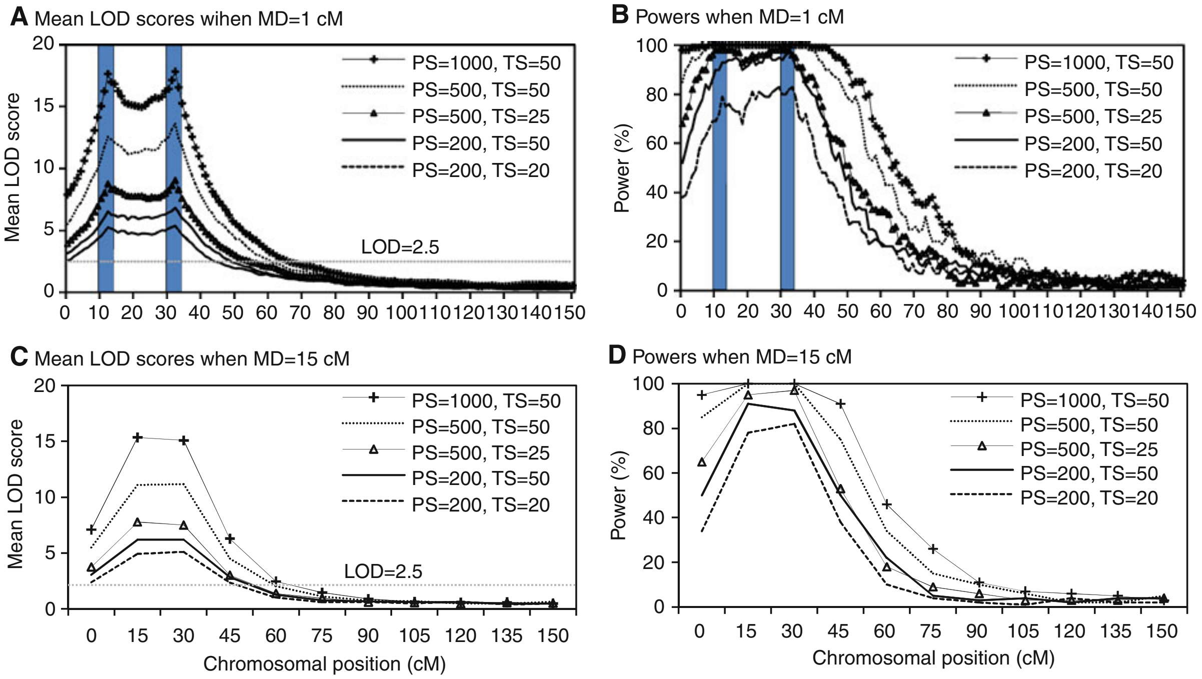

Fig. 7 Mean LOD scores of selective genotyping for two linked QTL (located at 12 and 32 cM on the same chromosome, each explaining 5% of phenotypic variation). Five scenarios of different PS and TS are compared. The *grey areas* indicate two target QTL regions, each spanning 4 cM. *MD* marker density, *PS* population size, *TS* tail size

additive QTL contributing to kernel weight that were distributed across five of the seven barley chromosomes (although all were under the LOD threshold of 2.5) (Fig. 9; Table 6). The largest two were qKWT5H (PVE = 38.27%) located at 5.0 cM on chromosome 5H, and qKWT7H (PVE = 17.51%) located at 95.0 cM on chromosome 7H. The nine QTL collectively explained a total of 80.76% of the phenotypic variation, indicating additive effects are the major source of genetic variation in this population. In contrast, SIM was only able to identify the three largest QTL identified by ICIM: qKWT5H, qKWT7H-1, and qKWT7H-2 (Table 6).

SGM identified six significant markers around the nine QTL positions identified by ICIM (Table 6). The two independent QTL (qKWT4H and qKWT5H) were identified by SGM. Though on the same chromosome, qKWT7H-1 and qKWT7H-2 have a linkage distance of about 90 cM (Table 6), which less affects the QTL detection by SIM and SGM. qKWT2H-1, -2 and -3 are linked in the repulsion phase on chromosome 2H. The linkage distances are 54 cM between qKWT2H-1, and -2, and 62 cM between qKWT2H-2, and -3. The repulsive linkage reduces the power of detection for SIM and SGM and therefore, only qKWT2H-3 was detected by SGM. Moreover, by chance in this empirical map there was a marker closely linked to qKWT2H-3 (at 201.7 cM) which will have further aided SGM to identify qKWT2H-3. qKWT3H-1 and qKWT3H-2 are also linked in the repulsion phase at a genetic distance of 22 cM. SIM failed to identify either of these QTL while SGM only identified qKWT3H-2 due to its relatively large effect.

Discussion

Power of QTL mapping based on selective genotyping

In this report, we have revealed that QTL mapping based on differences in allelic frequency between high- and low-tails of the phenotypic distribution of an RIL mapping population had a lower QTL detection power than ICIM but higher than SIM.

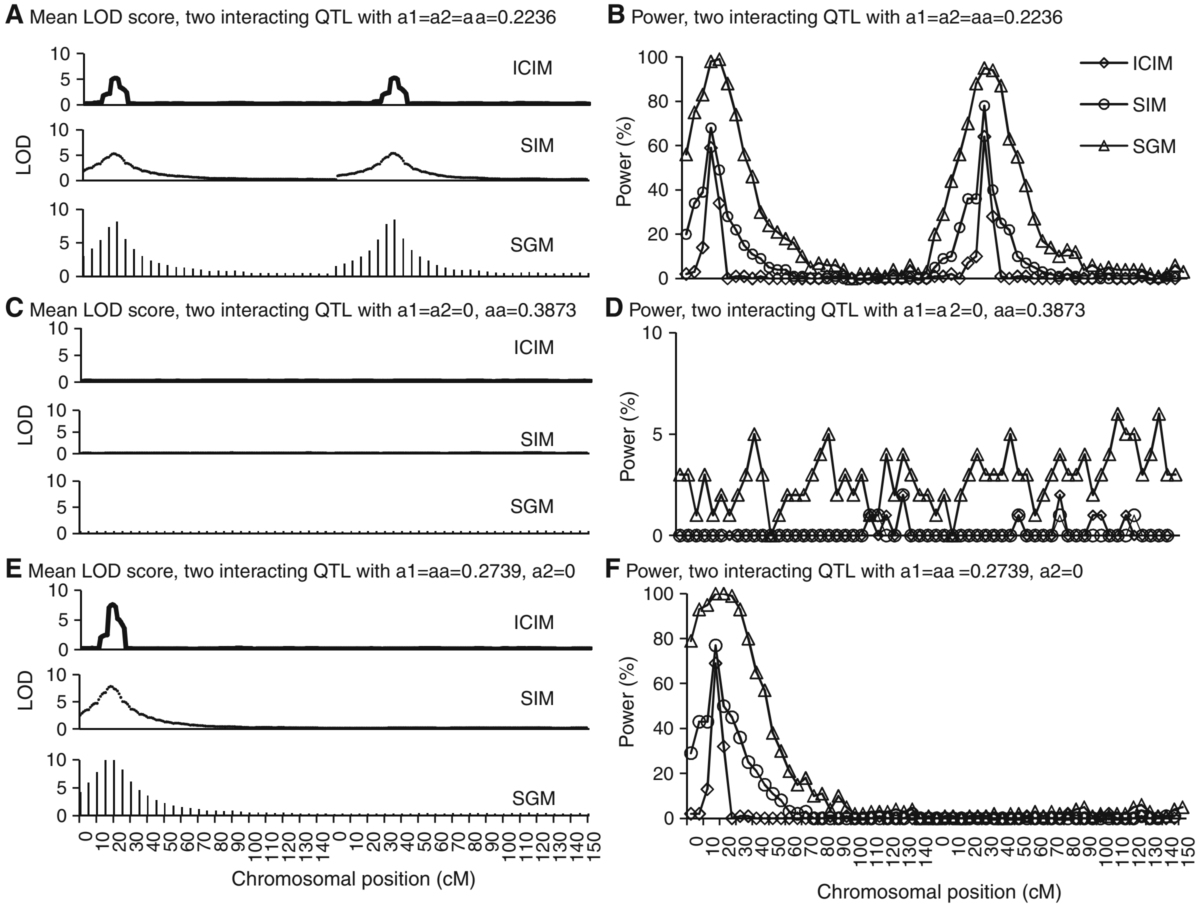

Fig. 8 Mean LOD score (**A**, **C**, **E**) and power (**B**, **D**, **F**) from ICIM, SIM and SGM (selective genotyping under bidirectional selection) for the two interacting QTL defined in Table 3. Two chromosomes are shown, where the two interacting QTL are located. The population size is 500, marker density is 5 cM, and the proportion of the entire population selected for each tail population is 25% for bidirectional selective genotyping. *ICIM* inclusive composite interval mapping, *SIM* simple interval mapping, *SGM* selective genotyping mapping

We have confirmed that SGM has been very successful in mapping QTL with large effect (PVE = 15% or higher) by using relatively small population sizes (150–250), low proportions in each tail (SP = 15–20%), and low marker density (15–25 cM). However, our simulation results indicate that the efficiency of SGM has been far from fully exploited in terms of the QTL detection power that can be achieved when the critical factors are addressed. By improving population size, proportion of the population selected for each tail and marker density, the QTL detection power can be substantially increased and the rate of false positives can be significantly reduced.

SGM can be effectively used for almost all genetic models we have simulated including QTL with linkage, epistasis and various levels of PVEs. First, SGM can be used to simultaneously map as many QTL as there are chromosomes if each chromosome has only one QTL. QTL with very low effects (PVE = 1%) can be detected if the population size is increased to 3,000 with a suitable tail size ($n > 100$) and marker density (<5 cM). Second, two linked QTL can be readily separated if they are more than 20 cM apart. Third, interaction between two QTL can be detected if at least one of the QTL has significant additive effects and population size is relatively large (500 or larger). Our results support the conclusions of

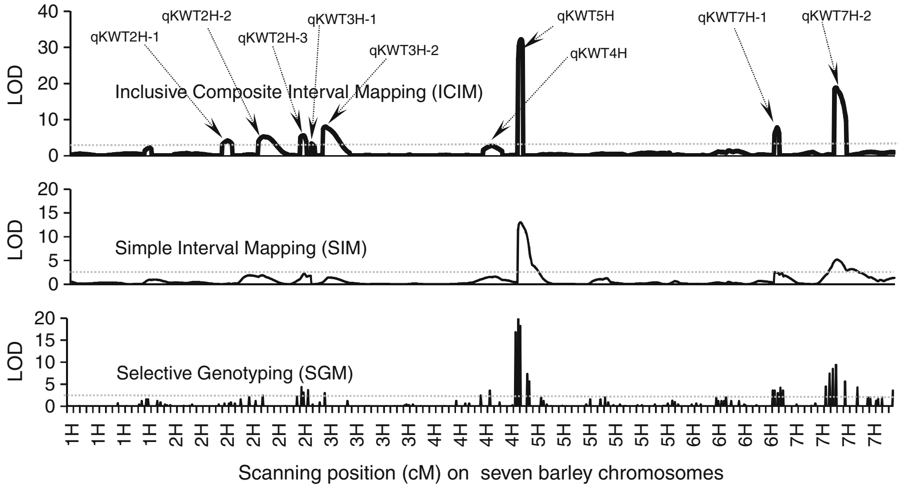

Fig. 9 Mapping results from SGM, ICIM and SIM for kernel weight in the barley DH population. Proportion of the entire population selected for each tail population is 25% for bidirectional selective genotyping. *ICIM* inclusive composite interval mapping, *SIM* simple interval mapping, *SGM* selective genotyping mapping. *Dashed lines* represent the LOD threshold of 2.5

Table 6 Mapping of kernel weight in the barley DH population with ICIM, SIM, and SGM

QTL	ICIM (PIN = 0.01, and POUT = 0.02)				SIM	SGM
	Position (cM)	Additive	LOD	PVE (%)	LOD	LOD
qKWT2H-1	83.00	0.39	4.16	3.04	0.35	0.53
qKWT2H-2	139.00	−0.46	5.28	4.23	1.99	2.44
qKWT2H-3	201.00	0.45	5.60	4.20	2.21	**4.36**
qKWT3H-1	0.00	−0.35	3.35	2.40	0.05	0.02
qKWT3H-2	22.00	0.57	8.00	6.50	1.11	**2.97**
qKWT4H	125.00	−0.31	2.73	1.95	1.50	**3.55**
qKWT5H	5.00	−1.38	32.19	38.27	**13.05**	**19.82**
qKWT7H-1	4.00	−0.56	7.81	6.38	**2.55**	**3.55**
qKWT7H-2	95.00	−0.94	18.86	17.51	**5.36**	**9.41**

Notes: The LOD peak from SIM is at the same position of the corresponding QTL identified by ICIM; the LOD peak of SGM is at the nearest marker of the corresponding QTL identified by ICIM with SP = 25% for each tail. Bold values indicate LOD > 2.5

SGM selective genotyping mapping under bidirectional selection, *ICIM* inclusive composite interval mapping, *SIM* simple interval mapping, *SGM* selective genotyping mapping, *PVE* phenotypic variation explained

two previous studies combining analysis of marker frequencies with selective genotyping and comparative ANOVA and allele frequency analysis for independent QTL (Gallais et al. 2007; Navabi et al. 2009). Our studies simulated the factors affecting the mapping power much beyond the range of conventional selective genotyping, with population sizes up to 3,000, PVE down to 1%, and marker density up to 1 cM using genetic models involving linkage and epistasis. Genotyping costs would not change very much for SGM as entire population sizes are increased, as the larger the entire population size the smaller the proportion of individuals required in each tail in order to maintain the same QTL detection power. Comparative analysis with the empirical dataset from barley confirmed almost all of the trends observed in the simulation results, indicating that SGM is more powerful than SIM but less powerful than ICIM in detecting most QTL. As concluded from the simulation results, the QTL detection power should be improved by using a larger population size and a higher density marker.

Replacement of entire population genotyping with selective genotyping

Selective genotyping can be used to replace entire population genotyping in almost all cases we have simulated, without loss of QTL detection power if the entire and tail population sizes are large enough and a high density of markers is used. In addition, there is no need to eliminate false positive markers through validation screening of the entire population, as the power of QTL detection under these conditions is extremely high and the accumulative probability of false positives within a specific chromosome region decreases significantly with the increase of marker density.

Selective genotyping can be used for genetic mapping of QTL with relatively small effects as well as for epistatic QTL with additive effects and two linked QTL that are not too close to each other. In addition, selective genotyping can be used for fine mapping to narrow down associated genetic regions to less than 1 cM or even a few candidate genes. Our recommendation for selective genotyping for QTL of large effects (PVE = 10–15% or larger) would be: minimum 20 individuals (or more than SP = 10%) in each tail from an entire population of around 200 individuals. Conversely, for QTL of medium effect (PVE = 3–10%) we would recommend around 50 individuals (SP = 5–10%) in each tail from an entire population of 500–1,000 individuals. Finally, for QTL of small effect (PVE = 0.2–3%) we would recommend around 100 individuals (SP = less than 5%) in each tail from an entire population of 3,000–5,000 individuals. The need for populations of this large size to detect small QTL was recently demonstrated in mapping QTL for flowering time in maize using nested mapping populations totaling over 5,000 lines (Buckler et al. 2009). The proportions of the entire population recommended above for each tail population (SP), are significantly lower than the optimum SP (20–30%) proposed by previous studies (Darvasi and Soller 1992; Gallais et al. 2007; Navabi et al. 2009). By using large entire population sizes combined with a low SP, the absolute amount of genotyping will remain highly cost effective. However, the initial phenotyping of the entire population may become rate limiting. The overall cost of the experiment may not be reduced (or may be increased) due to the increased cost of phenotyping a larger entire population—unless quicker, easier and cheaper methods for accurately identifying extreme phenotypes (for the target trait) can be devised. When the cost ratio of genotyping to phenotyping was higher than 1, the optimal SP appeared to be between 10 and 20% for each tail (Gallais et al. 2007). As the number of QTL and their effects are unknown in most cases, the entire and tail population sizes required for a specific experiment will depend on the objectives of the study. The empirical barley dataset used in this study provides a practical demonstration of SGM, highlighting which factors such as those simulated in this study should be improved when replacing entire population-based QTL mapping. However, we would like to indicate some limitations of selective genotyping, including limited usefulness for multiple traits, difficulty to estimate QTL genetic effects, and reduced suitability for non-additive epistatic QTL and QTL by environment interaction.

Using selective genotyping for "All-in-one plate" mapping of all target traits in one step

A large number of highly homozygous trait-specific materials have been developed for genetic analysis and breeding in many crops. These include inbred lines and cultivars with extreme phenotypes, eternal/fixed segregating populations (such as RILs, doubled haploids, near isogenic lines and introgression lines), genetic stocks (e.g. single segment substitution lines) and mutant libraries. These are all valuable directly for the purpose they were developed but also offer a novel resource for genetic mapping and gene discovery when used collectively. These materials have often been phenotyped in multiple environments due to their permanently fixed genetic composition. By collecting phenotypic extremes from currently available genetic and breeding materials, and utilizing selective genotyping and pooled DNA analysis, it is theoretically possible that one 384-well plate could be designed to cover the mapping of almost all major gene/QTL controlled agronomic traits of importance in a crop species. Progress towards testing this approach is already underway in maize at CIMMYT (Xu et al. 2009).

Genomewide linkage and linkage disequilibrium mapping

Recent developments in SNP genotyping technologies and application methodologies have enabled cost effective genomewide linkage disequilibrium (LD)-based association mapping in humans using selective genotyping, pooled DNA analysis and microarray-based SNP genotyping with 10,000–1,000,000 markers (Sham et al. 2002; Meaburn et al. 2006; Yang et al. 2006; Wilkening et al. 2007; Docherty et al.

2007; Kruglyak 2008; McCarthy et al. 2008). This system has the power to estimate allele frequencies and identify unique alleles from a pooled DNA sample of several hundreds of individuals. If this approach is successfully translated to plants it will resolve many of the constraints of pooled DNA analysis and can be used for both linkage mapping and LD mapping.

In a segregating population of plant species, SNP markers that are tightly linked to the target trait or within the gene, particularly for major-gene controlled traits, will clearly show one of the two polymorphic nucleotides, the segregation of which are completely correlated with the two contrasting phenotype pools. Thus, this situation is similar to genotyping two contrasting homozygotes. Any markers with less clear segregation would be automatically rejected on the assumption of less than optimum linkage to the target locus (Xu et al. 2009). As a result, it is not a prerequisite to develop a set of SNPs optimized for pooled DNA analysis as done in human genomics in order to make the pooled DNA analysis practical in linkage-based genetic mapping in plant species. This type of pooled DNA analysis strategy has been successfully used with seed DNA-based genotyping developed by Gao et al. (2008) for linkage mapping of genes affecting quality protein maize (QPM), where a SNP chip containing 1536 SNP markers was screened across seven F_2 populations derived from QPM × QPM crosses that segregated for kernel hardness (Xu et al. 2009). With the much higher-density of markers recently developed in maize, individual results from pooled DNA-based genetic mapping can be confirmed by multiple linked markers within the same experiment and thereby the power of QTL detection can be significantly increased as indicated by Fig. 2.

Genomewide association (linkage disequilibrium) mapping may provide a shortcut to discovering functional alleles and allelic variations that are associated with agronomic traits of interest. Selective genotyping, along with pooled DNA analysis, can be extended to using inbred lines with extreme phenotypes selected from various collections of germplasm. This is in principal similar to LD-based association mapping but using selected phenotypic extremes. For association mapping of quantitative traits governed by a large number of minor genes which interact with each other and with the environment, selective genotyping will face the same challenges as experienced with linkage-based QTL mapping using entire population genotyping.

Combining selective genotyping with selective phenotyping

The selective phenotyping method involves selecting individuals that maximize genotypic dissimilarity. Selective phenotyping is most effective when prior knowledge of genetic architecture allows focus on specific genetic regions (Jannink 2005; Jin et al. 2004) and specific allele combinations. Gallais et al. (2007) analyzed the cost ratio of genotyping to phenotyping when an optimal selected proportion of genotypes was determined. As genotyping becomes cheaper, it may be more efficient to first carry out low density genotyping of the whole population in order to identify the most informative subset of individuals in terms of minimum level of relatedness plus optimum subpopulation structure and allele representativeness. Precision phenotyping using physiological component and surrogate traits can then be carried out on this subset to enable further selection followed by whole genome genotyping. In this way, the optimum number of individuals (from a genetic and cost perspective) can be phenotyped and genotyped to maximize the power of the QTL detection while minimizing the overall cost of the experiment. For some target traits, phenotypic extremes can be easily identified by using a simple screening method, for example by using a strong abiotic stress screen to identify the most and least tolerant lines for genotyping and eliminate a large proportion of the rest of the population (Lebowitz et al. 1987). In these cases, selective genotyping can be highly optimized for maximum power of QTL detection and minimum cost of the overall experiment. High-density planting and selection at early stages of plant development, combined with selective phenotyping and genotyping should also be investigated as a potential option for some traits in order to allow one to work with more plants/families at the same cost (Xu and Crouch 2008). Where the target trait is influenced by planting density or strong selection pressure this will clearly confound the ability to make genetic gain. However, many major-gene controlled traits can be investigated in this way without much disturbance.

Using selective genotyping to develop a breeding-to-genetics approach

Traditionally molecular biologists start with genetic mapping of target genes in a specifically designed population, validate candidate markers in a representative target population and then select for those mapped genes in plant breeding through marker-assisted selection (Table 7). For many complex traits, favorable alleles at the genetic loci contributing to a specific trait are usually dispersed through a range of genetic materials. As a result, several mapping populations have to be developed plus phenotyped and genotyped. Moreover, combining multiple favorable alleles from different genetic loci in different sources requires multiple cycles of intermating and selection. We believe that the process could work better in reverse, starting with identification of extreme phenotypes from segregating populations involving multiple parental lines that are being used in breeding programs. The selected extremes are assumed to host diverse favorable alleles and loci from different sources but have been brought together into a single population by intermating and selection. The selected extremes are then used for rapid discovery of individual genes/alleles and their combined effects (Table 7). This approach would be particularly powerful (in terms of speed and cost) when combined with selection under appropriate target biotic or abiotic stresses where a large number of plants can be selected for extreme phenotypes. Compared with the normal genetics-to-breeding approach, this reversed approach can save 3–4 crop seasons in each cycle and can be fully integrated with ongoing breeding programs. Several long-term selection programs in maize and rice (Dudley 2004; Xu et al. 1998), indicate that favorable alleles for a complex trait from different genetic loci can be combined through multiple cycles of selection for extreme phenotypes.

Table 7 Numbers of crop seasons required for two contrasting genetics-and-breeding strategies for pyramiding favorable alleles of multiple sources

Procedure	Number of crop seasons required
Genetics-to-breeding approach	
Parental survey	1
Development of mapping populations	3
Phenotyping and genetic mapping using multiple populations	2
Intermating lines with different favorable alleles	1
Accumulating effect study	2
Release of materials for breeding	1
Total	10
Breeding-to-genetics approach	
Parental survey	1
Intermating lines with different favorable alleles	2–3
Phenotyping and accumulating effect study	2
Release of materials for breeding	1
Total	6–7

Acknowledgments This work was supported by the National 863 Program of China (No. 2006AA10Z1B1). Genomics and molecular breeding research at CIMMYT has been supported by the Rockefeller Foundation, Bill and Melinda Gates Foundation, European Community, Generation Challenge Program and HarvestPlus Challenge Program, and through other attributed or unrestricted funds provided by the members of the CGIAR and national governments of USA, Japan and UK. ICIM was implemented by the QTL mapping software package called QTL IciMapping, available from http://www.isbreeding.net.

References

Barua UM, Chalmers KJ, Hackett CA, Thomas WT, Powell W, Waugh R (1993) Identification of RAPD markers linked to a *Rhynchosporium secalis* resistance locus in barley using near-isogenic lines and bulked segregant analysis. Heredity 71:177–184

Brohede J, Dunne R, Mckay JD, Hannan GN (2005) PPC: an algorithm for accurate estimation of SNP allele frequencies in small equimolar pools of DNA using data from high density microarrays. Nucleic Acids Res 33:e142

Buckler SE, Holland JB, McMullen MM, Kresovich S, Acharya C, Bradbury P, Brown P, Browne C, Eller M, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz J, Goodman M, Harjes C, Hutchins K, Kroon D, Larsson S, Lepak N, Li H, Mitchell S, Pressoir G, Peiffer J, Rosas MO, Rocheford T, Romay C, Romero S, Salvo S, Sanchez Villeda H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z (2009) The genetic architecture of maize flowering time. Science 325:714–718

Charcosset A, Gallais A (1996) Estimation of the contribution of quantitative trait loci (QTL) to the variance of quantitative trait by means of genetic markers. Theor Appl Genet 93:1193–1201

Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. Theor Appl Genet 112:1205–1220

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. Genetics 138:1365–1373

Darvasi A, Weinreb A, Minke V, Wellert JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R (2007) Applicability of DNA pools on 500K SNP microarrays for cost-effective initial screens in genome wide association studies. BMC Genomics 8:214

Dudley JW (2004) 100 generations of selection for oil and protein in corn. Plant Breed Rev 24(Part 1):79–110

Dunninnton EA, Haberefeld A, Stallard LG, Siegel PB, Hillel J (1992) Deoxyribonucleic-acid fingerprint bands linked to loci coding for quantitative traits in chicken. Poult Sci 71:1251–1258

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Foolad MR, Jones RA (1993) Mapping salt-tolerance genes in tomato (*Lycopersicon esculentum*) using trait-based marker analysis. Theor Appl Genet 87:184–192

Gallais A, Moreau L, Charcosset A (2007) Detection of marker–QTL associations by studying change in marker frequencies with selection. Theor Appl Genet 114:669–681

Gao S, Martinez C, Skinner DJ, Krivanek AF, Crouch JH, Xu Y (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol Breed 22:477–494

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal interval using DNA pools from existing mapping populations. Nucleic Acid Res 19:6553–6558

Hillel J, Avner R, Baxter-Jones C, Dunnington EA, Cahaner A, Siegel PB (1990) DNA fingerprints from blood mixes in chickens and turkeys. Anim Biotech 2:201–204

Hormaza JI, Dollo L, Polito VS (1994) Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis. Theor Appl Genet 89:9–13

Jannink JL (2005) Selective phenotyping to accurately mapping quantitative trait loci. Crop Sci 45:901–908

Jin C, Lan H, Attie AD, Churchill GA, Bulutuglo D, Yandell BY (2004) Selective phenotyping for increased efficiency in genetic mapping study. Genetics 168:2285–2293

Knight J, Sham P (2006) Design and analysis of association studies using pooled DNA from large twin samples. Behav Genet 36:665–677

Korol A, Frenkel Z, Cohen L, Lipkin E, Soller M (2007) Fractioned DNA pooling: a new cost-effective strategy for fine mapping of quantitative trait loci. Genetics 176:2611–2623

Kruglyak L (2008) The road to genome-wide association studies. Nat Rev Genet 9:314–318

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

Lebowitz RL, Soller M, Beckmann JS (1987) Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in cross between inbred lines. Theor Appl Genet 73:556–562

Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Li H, Ribaut JM, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor Appl Genet 116:243–260

Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM (2008) Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res 36(6):e35

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–369

Meaburn E, Butcher LM, Schalkwyk LC, Plomin R (2006) Genotyping pooled DNA using 100 K SNP microarrays: a step towards genomoewide association scans. Nucleic Acids Res 34(4):e28

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance gene by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. Euphytica 137:111–118

Navabi A, Mather DE, Bernier J, Spaner DM, Atlin AN (2009) QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analyses. Theor Appl Genet 118:347–358

Plotsky Y, Cahaner A, Haberfeld A, Lavi U, Lamont SJ, Hillel J (1993) DNA fingerprint bands applied to linkage analysis with quantitative trait loci in chickens. Anim Genet 24:105–110

Quarrie SA, Lazić-Jančić V, Kovačević D, Steed A, Pekić S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J Exp Bot 50:1299–1306

Sham P, Bader JS, Craig I, O'Donovan M, Owen M (2002) DNA pooling: a tool for large-scale association studies. Nat Rev Genet 3:862–871

Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A (2008) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 4(2):e28

Soller M, Beckmann JS (1990) Marker-based mapping of quantitative trait loci using replicated progenies. Theor Appl Genet 80:205–208

Stuber CW, Moll RH, Goodman MM, Schaffer HE, Weir BS (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays*). Genetics 95:225–336

Stuber CW, Goodman MM, Moll RH (1982) Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci 22:737–740

Tinker NA, Mather DE, Rossnagel BG, Kasha KJ, Kleinhofs A, Hayes PM, Falk DE, Ferguson T, Shugar LP, Legge WG, Irvine RB, Choo TM, Briggs KG, Ullrich SE, Franckowiak JD, Blake TK, Graf RJ, Dofing SM, Saghai Maroof MA, Scoles GJ, Hoffman D, Dahleen LS, Kilian A, Chen F, Biyashev RM, Kudrna DA, Steffenson BJ (1996) Regions of the genome that affect agronomic performance in two-row barley. Crop Sci 36:1053–1062

van Treuren R (2001) Efficiency of reduced primer selectivity and bulked DNA analysis for the rapid detection of AFLP polymorphisms in a range of crop species. Euphytica 117:27–37

Villar M, Lefevre F, Bradshaw HD Jr, du-Cros ET (1996) Molecular genetics of rust resistance in Poplars (*Melampsora larici-populina* Kleb/*Populus* sp.) by bulked segregant analysis in a 2 × 2 factorial mating design. Genetics 143: 531–536

Wang J (2009) Inclusive composite interval mapping of quantitative trait genes. Acta Agronomica Sinica 35(2): 239–245

Wilkening S, Chen B, Wirtenberger M, Burwinkel B, Försti A, Hemminki K, Canzian F (2007) Allelotyping of pooled DNA with 250 K SNP microarrays. BMC Genomics 8:77

Wingbermuehle WJ, Gustus C, Smith KP (2004) Exploiting selective genotyping to study genetic diversity of resistance to Fusarium head blight in barley. Theor Appl Genet 109:1160–1168

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48: 391–407

Xu Y, McCouch SR, Shen Z (1998) Transgressive segregation of tiller angle in rice caused by complementary action of genes. Crop Sci 38:12–19

Xu Y, Babu R, Hao Z, Lu Y, Gao S, Yan J, Zhang S, Li J, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Taba S, Palacios N, Pixley K, Guimarães CT, Araus J, Crouch JH (2009) SNP-chip based genomewide scan for germplasm evaluation and marker-trait association analysis and development of a molecular breeding platform. In: Proceedings of 14th Australasian Plant Breeding & 11th Society for the Advancement in Breeding Research in Asia & Oceania Conference, 10–14 August 2009, Cairns, Tropical North Queensland, Australia

Yang HC, Liang YJ, Huang MC, Li LH, Lin CH, Wu JY, Chen YT, Fann CSJ (2006) A genome-wide study of preferential amplification/hybridization in microarray-based pooled DNA experiments. Nucleic Acids Res 34(15):e106

Zhang J, Xu Y, Wu X, Zhu L (2002) A bentazon and sulfonylurea sensitive mutant: breeding, genetics and potential application in seed production of hybrid rice. Theor Appl Genet 105:16–22

Zhang LP, Lin GY, Niño-Liu 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

Zhang L, Li H, Li Z, Wang J (2008) Interactions between markers can be caused by the dominance effect of QTL. Genetics 180:1177–1190