Theor Appl Genet (2009) 119:65–74
DOI 10.1007/s00122-009-1017-2

ORIGINAL PAPER

Simultaneous selection of major and minor genes: use of QTL to increase selection efficiency of coleoptile length of wheat (*Triticum aestivum* L.)

Jiankang Wang · Scott C. Chapman · David G. Bonnett · Greg J. Rebetzke

Received: 1 September 2008 / Accepted: 17 March 2009 / Published online: 10 April 2009

Abstract Plant breeders simultaneously select for qualitative traits controlled by one or a small number of major genes, as well as for polygenic traits controlled by multiple genes that may be detected as quantitative trait loci (QTL). In this study, we applied computer simulation to investigate simultaneous selection for alleles at both major and minor gene (as QTL) loci in breeding populations of two wheat parental lines, HM14BS and Sunstate. Loci targeted for selection included six major genes affecting plant height, disease resistance, and grain quality, plus 6 known and 11 "unidentified" QTL affecting coleoptile length (CL). Parental line HM14BS contributed the target alleles at two of the major gene loci, while parental line Sunstate contributed target alleles at four loci. The parents have similar plant height, but HM14BS has a longer coleoptile, a desirable attribute for deep sowing in rainfed environments. Including the wild-type allele at the major reduced-height locus *Rht-D1*, HM14BS was assumed to have 13 QTL for increased CL, and Sunstate four; these assumptions being derived from mapping studies and empirical data from an actual HM14BS/Sunstate population. Simulation indicated that compared to backcross populations, a single biparental F_1 cross produced the highest frequency of target genotypes (six desired alleles at major genes plus desired QTL alleles for long CL). From 1,000 simulation runs, an average of 2.4 individuals with the target genotype were present in unselected F_1-derived doubled haploid (DH) or recombinant inbred line (RIL) populations of size 200. A selection scheme for the six major genes increased the number of target individuals to 19.1, and additional marker-assisted selection (MAS) for CL increased the number to 23.0. Phenotypic selection (PS) of CL outperformed MAS in this study due to the high heritability of CL, incompletely linked markers for known QTL, and the existence of unidentified QTL. However, a selection scheme combining MAS and PS was equally as efficient as PS and would result in net savings in production and time to delivery of long coleoptile wheats containing the six favorable alleles.

Communicated by M. Cooper.

J. Wang (✉)
Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement and CIMMYT China, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, 100081 Beijing, China
e-mail: wangjk@caas.net.cn; jkwang@cgiar.org

S. C. Chapman
CSIRO Plant Industry, 306 Carmody Rd, St Lucia, QLD 4067, Australia

D. G. Bonnett · G. J. Rebetzke
CSIRO Plant Industry, P.O. Box 1600, Canberra, ACT 2601, Australia

Present Address:
D. G. Bonnett
International Maize and Wheat Improvement Center (CIMMYT), Mexico, DF 06600, Mexico

Introduction

The availability of an ever-increasing number of useful molecular markers is allowing accurate selection at a greater number of major gene loci than has previously been possible (Paterson et al. 1991; Young 1999; Dekkers and Hospital 2002). Many breeding programs representing a range of crops are using or moving towards the use of molecular markers as diagnostic tools to screen for major

genes of interest (e.g., Eagles et al. 2001; Dubcovsky 2004). Determining the most efficient crossing and selection strategy when pyramiding multiple major genes into one genotype has been the subject of numerous research studies (e.g., Bonnett et al. 2005; Wang et al. 2007).

The availability of saturated genetic maps in many species has also led to the intensive use of QTL mapping in genetic studies for quantitative traits (Lander and Botstein 1989; Dekkers and Hospital 2002; Barton and Keightley 2002; Bernardo 2002; Bernardo and Charcosset 2006; Li et al. 2007). Plant breeding companies have apparently begun releasing maize hybrids in which QTL markers have been applied to improve the effectiveness of recurrent selection (Bernardo and Charcosset 2006; Eathington et al. 2007; Bernardo and Yu 2007). However, there are few reported examples of the implementation of QTL markers in line development of self-pollinated crops such as wheat, despite the substantial investment in QTL mapping for many traits. A major hurdle when using QTL markers in breeding is that each QTL singly accounts for a relatively small portion of genotypic variance and therefore simultaneous selection for multiple QTL is necessary to achieve useful genetic gains. In many instances, a substantial portion of the genotypic variance remains unmapped and phenotypic selection may be a more efficient means of selecting for a trait depending on heritability, levels of genotype by environment interaction (G × E), timing, and cost (Moreau et al. 2004; Davies et al. 2006). Cooper et al. (2005) developed a simulation approach to measure the effect of "unmapped variance" on the utility of mapped QTL. The method compares the relative response to phenotypic selection or MAS in a recurrent selection scenario. Inputs to the simulation are the effects of the mapped QTL for a given trait, together with an ensemble of simple to complex gene models for the "unmapped QTL". The resulting complexity–response plots are then used to establish the value of mapped QTL in terms of the expected response to selection—a more valuable measure than the "variance explained".

Major genes commonly have readily discernible effects on phenotypic traits and their inheritance can be studied by Mendelian genetics (Falconer and Mackay 1996). In contrast, the effect of a minor gene is not large enough to cause a discontinuity in phenotype and cannot be studied individually in classical quantitative genetics. The definition of QTL varies in the literature. According to Dekkers and Hospital (2002), QTL are genetic loci or chromosome regions that contribute to variability in complex quantitative traits. In this sense, a QTL can be either a major or a minor gene. If QTL are defined as identifiable through statistical analysis of complex traits rather than traditional Mendelian genetics (Barton and Keightley 2002), QTL can only be minor genes. In this study, we adopt the latter definition of QTL. Therefore, the reduced-height genes in wheat are called major genes instead of QTL, and QTL specifically indicate genes identified by QTL mapping methods such as interval mapping (Lander and Botstein 1989), and inclusive composite interval mapping (Li et al. 2007, 2008).

As breeding programs rarely focus on a single trait, use of identified QTL-marker associations to select for a trait must be balanced with selection for other traits using markers and/or phenotyping. Strategies for efficient pyramiding of diagnostic alleles at multiple loci have been reported (Bonnett et al. 2005; Wang et al. 2007), while Podlich et al. (2004) have proposed "mapping as you go" methods to utilize QTL for complex traits in recurrent selection. However, simultaneous selection for major genes and multiple QTL would be necessary to achieve useful genetic gains in species such as wheat. Coleoptile length (CL) is an important adaptive trait that contributes to improved establishment and yield in rainfed wheat-growing areas (Rebetzke et al. 1999, 2004, 2007a). Although under polygenic control and typically selected in later generations of the breeding process, CL has a high heritability and small G × E (Rebetzke et al. 2004). Several QTL have been mapped for CL across different populations evaluated in contrasting environments (Rebetzke et al. 2001, 2007b).

In this paper, we use the QuLine genetic and breeding simulation tool (Wang et al. 2003, 2004) to investigate a realistic wheat breeding scenario where MAS for major genes and QTL of smaller effects are integrated with conventional phenotypic selection to develop new lines from an elite biparental cross. Our breeding objectives were to combine six major reduced-height, disease resistance, and grain quality genes with phenotypic selection and MAS to maximize CL in target breeding lines.

Materials and methods

Six major genes and allele distribution in the two parental lines

The two parental lines used in this study were HM14BS and Sunstate. HM14BS is a long coleoptile, *Rht8* semidwarf wheat developed by CSIRO Plant Industry in Australia. It is an F_6-derived line obtained from a cross between the tall (*rht8*), long coleoptile cv. Halberd and the semidwarf (*Rht8*), short coleoptile cv. Mara. The long coleoptile of HM14BS reflects accumulation of favorable coleoptile alleles from the long coleoptile parent Halberd (Rebetzke and Richards 2000; Rebetzke et al. 2007a). Sunstate is an Australian wheat cultivar, carrying the "Green Revolution" reduced-height allele *Rht-D1b* that has a short coleoptile due to the pleiotropic effect of this

allele (Ellison et al. 1994; Rebetzke et al. 2007b). Genotype profiles were simulated to represent the relevant genetic characteristics of the two parental lines. In this particular work, we did not consider background effects other than for the genes of interest.

Table 1 shows the genotypes of the two parents at six major gene loci, including two reduced-height, two disease resistance, and two grain quality loci. Molecular markers for the six major genes are currently being deployed by breeding programs to develop new lines for commercial release. Alleles at the *Rht-D1* (syn. *Rht2*) and *Rht8* loci affect plant height (Rebetzke and Richards 2000; Ellis et al. 2002); *Sr2* is an adult-plant stem rust resistance gene; *VPM* is an *Aegilops ventricosa* chromosome translocation carrying genes for leaf (*Lr37*), stem (*Sr38*), and stripe (*Yr17*) rust resistance; and the *Glu-B1* and *Glu-A3* loci contain alleles coding for grain storage proteins (Eagles et al. 2002; Wang et al. 2005). Molecular markers are completely linked to the target alleles except *Rht8* and *Sr2* where diagnostic markers are a small chromosomal distance from the respective major genes, i.e., 0.6 cM for *gwm261* and *Rht8* (Korzun et al. 1998), and 1.1 cM for *gwm533* and *Sr2* (Spielmeyer et al. 2003). All the markers are co-dominant, except for *VPM* (Table 1) and all genes are unlinked.

Sunstate carries favorable alleles for all four of the disease and quality genes of interest and has broad adaptation to many Australian environments in the northern and eastern production regions. It is a direct descendent of the CIMMYT release Pavon 76.

Major genes and QTL for CL and distribution of alleles between the two parents

The reduced-height allele at the *Rht-D1* locus (i.e., *Rht-D1b*) has a large negative pleiotropic effect on CL (Rebetzke et al. 2007b). *Rht-D1b* decreases plant height by an average of 20 cm and coleoptile length by 19 mm (Table 2). In contrast, *Rht8* can reduce plant height by about 18 cm but has no negative effect on CL (Rebetzke and Richards 2000). Thus, the ideal genotype should have the wild-type allele, i.e., *Rht-D1a*, at locus *Rht-D1*, and the reduced-height allele at the *Rht8* locus. Double-dwarf individuals with both reduced-height genes will be too short for many environments (Richards 1992).

QTL for CL were classified as "known" or "unidentified" (Table 2), with two of the QTL (*qCL1* and *qCL4*) being on the same chromosome arm as major genes (*Glu-A3* and *Sr2*, respectively) (Table 1). The effects of the six known QTL were taken from the genetic study of the Cranbrook × Halberd DH population (Rebetzke et al. 2001, 2007b). Cranbrook and Sunstate were derived from CIMMYT wheat varieties that carry *Rht-D1b* and have a coefficient of parentage of 0.241. In this practical example, we apply knowledge about QTL from a mapping population to a breeding population. In a separate experiment, HM14BS had an average CL of 160 mm, and Sunstate had an average CL of 97 mm, a difference of 63 mm. The six known QTL and *Rht-D1a* together can increase CL by 59 mm. Therefore, QTL identified in Rebetzke et al. (2007b) were distributed between the parents in our simulation assuming that HM14BS inherited the long coleoptile allele at each known QTL (Table 2).

As it is always difficult to identify alleles with minor effects from QTL mapping studies (Bernardo 2002), we assumed there were 11 unmapped QTL of various sizes (Table 2) that appropriately explain the observed difference in CL in the two parental lines and transgressive segregation in their progenies. All QTL were assumed to be chromosomally unlinked, due to the large number of wheat chromosomes. In this genetic model, *Rht-D1* explained 42.23% of the total genetic variation, the six known QTL explained 33.90%, and the 11 unidentified QTL explained 23.87%. This is in approximate accordance with previous QTL experiments, e.g., Rebetzke et al. (2007b).

Table 1 Six major genes, their chromosomal locations, and the genotypes of the two wheat parents

Locus symbol[a]	*Rht-D1*	*Rht8*	*Sr2*	*VPM*	*Glu-B1*	*Glu-A3*
Chromosome	4DS	2DL	3BS	7DL	1BL	1AS
Marker type	Codom	Codom	Codom	Dom	Codom	Codom
Distance to the nearest marker (cM)	0.0	0.6	1.1	0.0	0.0	0.0
HM14BS	*Rht-D1a*	*Rht8*	*sr2*	*vpm*	*Glu-B1a*	*Glu-A3e*
Sunstate	*Rht-D1b*	*rht8*	*Sr2*	*VPM*	*Glu-B1i*	*Glu-A3b*
Target genotype[b]	*Rht-D1a*	*Rht8*	*Sr2*	*VPM*	*Glu-B1i*	*Glu-A3b*

[a] Alleles *Rht-D1b* and *Rht8* reduce plant height, allele *Sr2* confers resistance to stem rust, allele *VPM* confers resistance to cereal cyst nematode, and alleles *Glu-B1i* and *Glu-A3b* improve bread wheat dough quality. The six loci are located on different wheat chromosomes, as indicated

[b] Alleles in the target genotype were determined by semi-dwarfing with long CL, multiple disease resistances, and excellent grain quality. The two semi-dwarfing alleles can each produce the required plant height. However, *Rht-D1b* also reduces the CL, which is unfavorable for breeding drought-resistant wheat cultivars. *Rht8* reduces plant height without affecting CL, and is therefore the favorable dwarfing allele under drought environments. Other alleles in the target genotype are easily understood as they increase the resistance to some disease, and increase grain quality

Table 2 Additive genetic effects of CL genes used in the simulation study and genotypes of HM14BS, Sunstate and the target genotype with the longest CL

Locus	Chromosome	Distance to the nearest marker (cM)	Additive effect (mm)[a]	Additive variance explained (%)	HM14BS	Sunstate	Genotype with all increased CL alleles
Rht-D1	4DS	0.0	9.5	42.63	*Rht-D1a*	*Rht-D1b*	*Rht-D1a*
qCL1	1AS	8.1	2.9	3.97	+	−	+
qCL2	2BS	0.7	2.5	2.95	+	−	+
qCL3	2DS	1.1	4.1	7.94	+	−	+
qCL4	3BS	0.9	2.0	1.89	+	−	+
qCL5	5AL	6.2	4.9	11.34	+	−	+
qCL6	5DS	13.0	3.6	6.12	+	−	+
qCL7	Unidentified		4.0	7.56	−	+	+
qCL8	Unidentified		3.0	4.25	+	−	+
qCL9	Unidentified		3.0	4.25	−	+	+
qCL10	Unidentified		2.0	1.89	+	−	+
qCL11	Unidentified		2.0	1.89	+	−	+
qCL12	Unidentified		2.0	1.89	−	+	+
qCL13	Unidentified		1.0	0.47	+	−	+
qCL14	Unidentified		1.0	0.47	+	−	+
qCL15	Unidentified		1.0	0.47	+	−	+
qCL16	Unidentified		1.0	0.47	+	−	+
qCL17	Unidentified		1.0	0.47	−	+	+
Coleoptile length (mm)					158	97	178

[a] Additive effect is defined as half of the difference between two homozygous genotypes, population mean was 127.5 mm for CL

Distribution of the 11 unidentified QTL effects was assumed in Table 2 so that HM14BS has a CL of 158 mm, and Sunstate a CL of 97 mm, similar to the observed data, and also allowing for the observed transgressive segregation. Heritability of CL in the narrow-sense at the individual plant level was set at 0.76 (Rebetzke et al. 2004, 2007b).

Breeding objectives

The challenge facing breeders is to combine Sunstate's favorable disease resistance and grain quality alleles with alleles associated with long CL from HM14BS. We assumed that a total of 1,000 individuals can be grown and genotyped by molecular markers in the F_2 generation, and that 200 inbred lines can be generated either through DH or repeated self-pollination. Our first objective was to select inbred lines combining the six desired major genes, i.e., the target genotype in Table 1. These lines will meet requirements for plant height, disease resistance, and grain quality. Our second objective was to select inbred lines combining the six desired major genes together with long CL. In the genetic model defined in Table 2, the greatest CL is 178 mm, when all 17 increased CL genes are combined in one genotype. However, this genotype has an extremely low frequency in a limited-size breeding population, so we set a reasonable target for CL, i.e., CL $>$ 130 mm, which is 30% longer than Sunstate's. The frequency of genotypes with the target configuration alleles was used in this study to compare various crossing and selection strategies. A greater number of lines with the target genotype would provide opportunities to further select for improved yield performance and other traits.

Simulation

QuLine, an integrated genetic and breeding simulation tool based on the QU-GENE platform (Podlich and Cooper 1998), is capable of simulating most breeding methodologies for developing inbred lines (Wang et al. 2003, 2004). Each simulation was run 1,000 times, and the final number of target genotypes was recorded to estimate their frequency distributions. Initial simulations examined and compared simulated and observed population attributes. We then compared the proportion of target genotypes in fixed lines from an unselected single cross or backcross to either parent. Subsequently, we applied several marker-aided selection methods in the F_2 population to increase the frequency of target genotypes for the major genes. A range of marker and phenotypic selection methods was then

applied to select for increased CL in later generations. Finally, we evaluated the effect on MAS of several other "favorable" genetic models, where "unmapped" QTL were removed and perfect marker-QTL linkage was assumed.

Results

Validation of the CL genetic model used in this study

Breeding populations were developed at CSIRO using HM14BS and Sunstate. The CL trait was measured on 167 $F_{4:5}$ families, after enhancement in the F_2 for the six major genes. Eight individuals each from the two parents were also phenotyped (Fig. 1a). Mean CL was 160 mm for HM14BS and 97 mm for Sunstate, while the F_5 families had a mean CL of 131 mm. The breeding procedure was simulated by QuLine, and distribution of CL from one run is shown in Fig. 1b. In the simulation, mean CL was 158 mm for HM14BS, and 104 mm for Sunstate. The F_5 families had a mean CL of 133 mm. Transgressive segregation was observed in real and simulated data (Fig. 1). A qualitative comparison of the distributions indicated consistency between Fig. 1a and b, which validated the approximate genetic model of CL defined in Table 2.

Single cross versus backcross

The frequency of the target genotype (i.e., inbred lines combining the six desired major genes and with CL > 130 mm in this study) in breeding populations determines the success of further trait selection. When the six major genes are the only objective, backcrossing with Sunstate as the recurrent parent (i.e., P_2BC_1) produces the largest number of lines having the target genotype (Fig. 2a). HM14BS has two desired alleles, and Sunstate has four among the six major genes. As there is no chromosomal linkage between the six major genes, frequencies of the target genotype can be readily calculated, i.e., $\left(\frac{1}{2}\right)^6 = 0.0156$, $\left(\frac{3}{4}\right)^2 \times \left(\frac{1}{4}\right)^4 = 0.0022$, and $\left(\frac{3}{4}\right)^4 \times \left(\frac{1}{4}\right)^2 = 0.0198$ in F_1, P_1BC_1, and P_2BC_1-derived inbred lines, respectively. Thus, in a population comprising 1,000 inbred lines, P_2BC_1 will generate about 20 lines combining the six desired genes, while P_1BC_1 and F_1 will generate 2 and 16, respectively.

When a long CL is the only breeding objective, more target genotypes occur in the P_1BC_1-derived population (Fig. 2b), as there are more desired CL genes in HM14BS (Table 2). When both major genes and long CL are desired, simulation indicated an F_1-derived population had the largest number of target genotypes (Fig. 2c). When the genes of interest are unlinked, it is simple to determine whether single cross or backcross methods should be used (Wang et al. 2007). However, when linkage is present (as for two QTL and two diagnostic genes here), there can be no simple calculation of the frequency of target genotypes under complicated linkage and breeding objectives, and therefore simulation tools are required.

Increasing target genotype frequency through selection at major gene loci

The frequency of the target genotype was 2.36% in the unselected F_1-derived DH or RIL population. Hence, few target genotypes would be found in 200 DH or RIL lines, leaving no residual genetic variation for the selection of other important traits. Early generation selection (say F_2) is needed to increase target genotype frequency in the fixed line population.

For a cross between HM14BS and Sunstate, the frequency of homozygosity in the F_2 at the six major loci is $\left(\frac{1}{4}\right)^6 = 0.000244$, such that even 4,000 F_2 individuals would be insufficient to identify a single homozygous target genotype (frequency less than 1 in 4,000). Enrichment of both (++) and (+−) genotypes has been proposed when many independent genes are to be selected (Bonnett et al. 2005; Wang et al. 2007), to increase the frequency of the favorable alleles (+) in the breeding population. Here, we considered four F_2 selection schemes for major genes (Fig. 3), i.e., no targeting of homozygotes but allele enrichment at all six loci (Hom0Het6); selection of

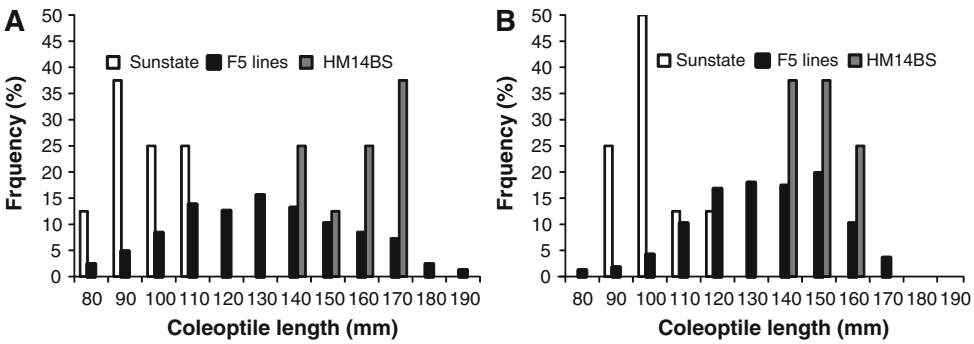

Fig. 1 Comparison of coleoptile length observed in the growth cabinet and predicted from the genetic model. **a** The observed frequency was determined from 167 $F_{4:5}$ families after enhancement in the F_2 for the six major genes; **b** the same number of F_5 families were simulated in QuLine to demonstrate the predicted frequency of CL

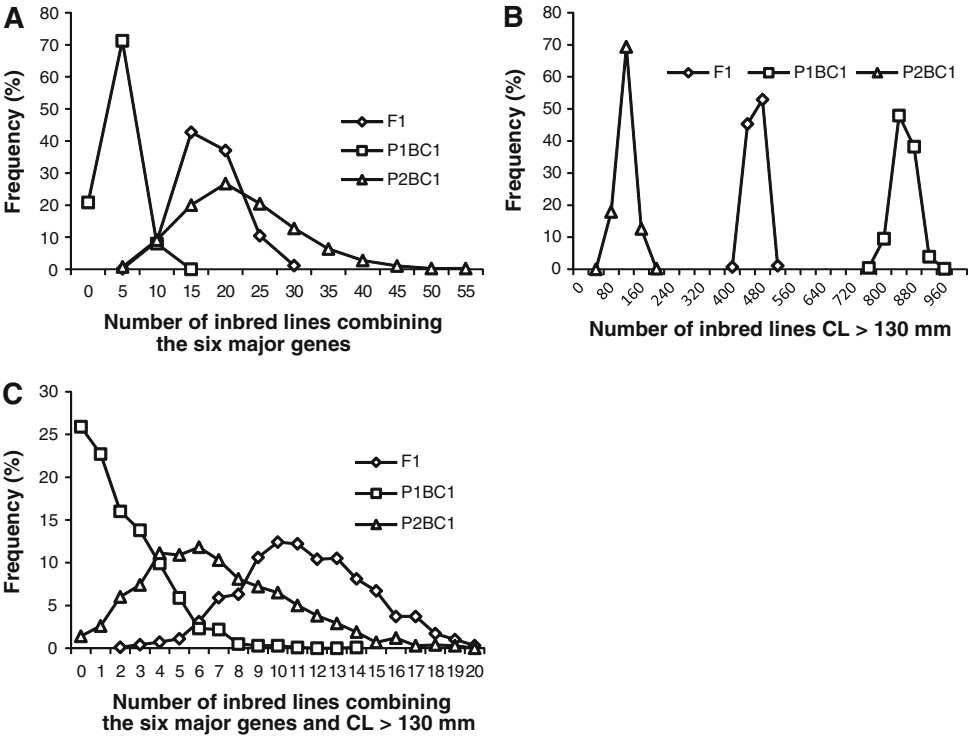

Fig. 2 Distributions of the number of target genotypes in 1,000 DH or RIL derived from F_1, P_1BC_1, and P_2BC_1. Frequency was calculated from 1,000 simulation runs. **a** Distribution of inbred lines combining the six major genes; **b** distribution of inbred lines with CL > 130; **c** distribution of inbred lines combining the six major genes and CL > 130

homozygotes at one locus and allele enrichment at five loci (Hom1Het5); selection of homozygotes at two loci and allele enrichment at four loci (Hom2Het4); selection of homozygotes at three loci and allele enrichment at the other three loci (Hom3Het3).

The Hom3Het3 scheme has the greatest selection intensity and left the fewest F_2 individuals after selection (Fig. 3a). On average, 6.4 F_2 individuals out of 1,000 were retained from Hom3Het3, 19.6 from Hom2Het4, 59.2 from Hom1Het5, and 177.7 from Hom0Het6. The small number of selected F_2 makes further selection difficult and also increases uncertainty in later generations (Fig. 3b). Therefore, we chose Hom1Het5 for further investigation. In this study, it makes no difference at which loci homozygotes are selected and at which loci allele enrichment is applied, as there is no linkage between the six major genes. If any of the loci were linked, different combinations of homozygote selection and enrichment across target loci, and their linkage phase, would produce different outcomes.

Further increasing target genotype frequency through CL selection

When Hom1Het5 was applied in F_2, an average 19.1 target genotypes (i.e., homozygous for the six major genes and CL > 130 mm) could be selected from 200 DH or RIL (Fig. 3b). When a MAS scheme based on marker score from the six known QTL was applied with a selected portion of 0.25 on selected F_2 after Hom1Het5, an average 23.9 target genotypes could be selected from 200 DH or RIL (Fig. 4), further increasing the frequency of the target genotype.

While CL can be difficult to select in the field and has a low heritability, particularly in early generations, in

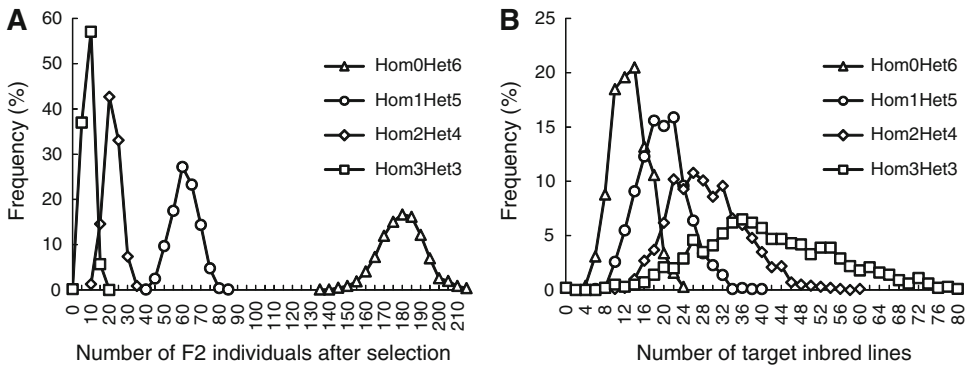

Fig. 3 Comparison of different selection schemes for the six major genes. Frequency was calculated from 1,000 simulation runs. **a** Distribution of the number of selected F_2 individuals, **b** distribution of the number of target inbred lines

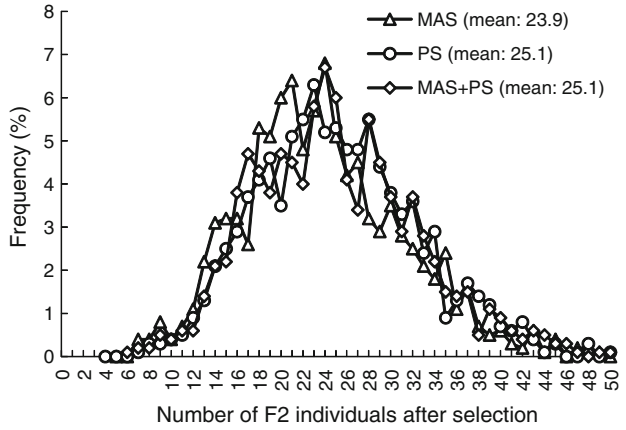

Fig. 4 Frequency distribution of target inbred lines from different selection schemes for coleoptile length. Frequency was calculated from 1,000 simulation runs. MAS = marker-assisted selection and PS = phenotypic selection. MAS + PS is a sequential selection scheme where MAS was applied first, followed by PS

advanced generations CL selection can be easily performed in the controlled environment (CE) using bulked seeds. Due to the high CL heritability in our CE screen, simulation indicated phenotypic selection will be more effective than MAS, with an average of 25.1 target genotypes selected from 200 DHs or RILs (Fig. 4) under the same selection intensity as used for MAS. However, the MAS screen can be applied to F_2 individuals before DH and RIL production, saving an entire generation (ca. 4 months) in the production of target inbred lines compared to PS.

Sequential selection of MAS and PS can be as effective as PS alone (Fig. 4). Considering the higher cost of phenotyping CL, the sequential selection of MAS and PS can greatly reduce overall breeding cost. When PS for CL was conducted, F_3 families were phenotyped. Therefore, the scheme where MAS is applied after PS was not applicable. Further, one more generation is needed when PS is applied, i.e., MAS can save one wheat growing season, compared with PS, if DH is used to derived pure lines.

In summary, when selection for target homozygotes was applied for one major gene locus and allele enrichment at the other five loci in a population of 1,000 F_2 individuals, i.e., Hom1Het5, the average number of lines with the target genotype increased to 19.1. Marker-assisted selection for a quarter of the retained F_2 individuals based on the six known QTL (i.e., Hom1Het5 + MAS) further increased the number to 23.9, while phenotypic selection for a quarter of the retained $F_{2:3}$ families (i.e., Hom1Het5 + PS) further increased the number to 25.1. At the 0.90 probability level, Hom1Het5 resulted in more than 13 target inbred lines, Hom1Het5 + MAS resulted in more than 14 target inbred lines, and Hom1Het5 + PS resulted in more than 15.

Superficial evaluation of error variance and marker-QTL linkage

Simulations make it possible to undertake comprehensive analyses of large numbers of alternative genetic models, e.g., Cooper et al. (2005). Here, we have evaluated a small number of alternative models to determine the conditions in which MAS was more efficient than PS after selection for target major genes (Table 3). Firstly, we assumed the unidentified QTL had an effect of 0, which is equivalent to the case where all QTL are known (A1). In this case, the error variance was appropriately adjusted so as to maintain the original heritability for CL. Second, we assumed that all major genes and identified QTL had completely linked molecular markers (A2). Third, we assumed that all QTL were known and had completely linked markers (A3). Fourth, we used a lower heritability (i.e., 0.10) at the individual plant level (A4). As before, each selection scheme was simulated 1,000 times.

For both selection intensities (i.e., 0.25 and 0.09) PS was more efficient than MAS even when all QTL were known, due to the high heritability of CL and the crossover between markers and known QTL (Model A1 in Table 3). When known QTL had completely linked markers, PS was more efficient (Model A2 in Table 3), due to the existence of unidentified QTL. In this case, MAS only selected for known QTL, while PS selected for both known and unidentified QTL. When all QTL were known with completely linked markers, PS was still more efficient (Model A3 in Table 3) for the less selected proportion of 0.25, due to the high heritability of CL. PS was less efficient only under the low heritability (Model A4 in Table 3). However, when combined with PS (i.e., MAS + PS, Table 3), MAS was at least as efficient as PS in most cases.

Discussion

Use of QTL mapping studies in breeding

As the number of published QTL for various traits increases, the challenge for plant breeders is to determine how best to utilize this knowledge to increase the efficiency of crop improvement and enhance genetic gain. Two types of selection involving markers can be used (Bernardo 2002). One is based on an index comprising both phenotypic value (usually for quantitative traits) and marker type (represented by marker score) (Lande and Thompson 1990; Bernardo and Charcosset 2006). The other is based on whether the marker is present or not (Young 1999; Eagles et al. 2001), and is normally used to select for important genes in crosses between largely adapted parents or to backcross specific genes into adapted backgrounds.

Table 3 Mean and standard deviation (SD) of the number of target inbred lines, and mean coleoptile length (CL) of target inbred lines from 1,000 simulation runs

Genetic model		Total selected portion 0.25			Total selected portion 0.09		
		MAS	PS	MAS + PS	MAS	PS	MAS + PS
Genetic model in Tables 1 and 2	Mean	23.91	25.14	25.15	24.97	25.31	25.38
	SD[a]	7.45	7.34	7.88	10.92	10.75	10.88
	CL	145.48	146.70	146.48	147.25	149.68	149.50
A1: All QTL were known, i.e., unidentified QTL in Table 2 had an effect of 0	Mean	25.35	25.52	25.31	25.89	26.03	25.55
	SD	7.65	7.66	7.73	11.30	11.52	10.65
	CL	144.25	144.05	144.07	146.65	146.31	146.49
A2: Both major genes and known QTL have completely linked markers	Mean	24.37	25.51	25.10	25.62	26.19	26.62
	SD	7.37	7.63	7.66	11.00	11.24	11.45
	CL	145.86	146.71	146.69	147.98	149.68	149.68
A3: All QTL were known with completely linked markers	Mean	25.71	26.16	25.69	26.27	26.18	26.55
	SD	7.63	7.56	7.47	11.26	11.43	11.88
	CL	144.78	144.12	144.21	147.50	146.49	146.42
A4: Lower heritability ($h^2 = 0.10$)	Mean	23.91	22.99	24.00	24.97	23.48	24.25
	SD	7.45	7.22	7.60	10.92	10.88	10.16
	CL	145.48	144.40	145.22	147.25	145.77	147.02

[a] Standard deviation, which is equal to the square root of variance of a distribution

Index selection is the optimal selection procedure, as it uses all of the information available about each individual (Falconer and Mackay 1996). Thus, MAS using an index calculated from both marker score and phenotype results in larger genetic gains than selection based only on marker score or phenotype (Lande and Thompson 1990; Bernardo 2002). However, index selection is more expensive compared with sequential selection (see independent culling levels in animal breeding; Falconer and Mackay 1996), where selection based on marker score is conducted first, followed by phenotypic selection of the retained individuals. Supposing there are 1,000 individuals in a breeding population and 50 individuals are retained after selection, all 1,000 individuals need to be genotyped and phenotyped if MAS is applied using an index. In practise, MAS based on marker score may be applied first—say 100 are retained after selection based on marker score—and then phenotypic selection is conducted with the retained 100 individuals. In this way, 1,000 individuals need to be genotyped and only 100 individuals phenotyped; therefore, reducing costs to the extent that the additional genotyping is cheaper than the foregone phenotyping expense.

In this study, sequential selection at major gene loci was first conducted. For example, in selection scheme Hom1-Het5, the 1,000 F_2 individuals were first screened for *Rht-D1*, and those having two *Rht-D1a* alleles were selected. The retained F_2 individuals were screened for *Rht8*, and those having at least one *Rht8* were selected, and so on. This sequential strategy greatly reduced marker screening cost (Wang et al. 2007). The score used in MAS of CL was determined from independent QTL mapping studies, so phenotypic CL was not required when applying MAS. All F_2 individuals retained after major gene selection were screened for the six markers closest to the six known QTL (Table 2). Then a portion of the F_2 was selected based on marker score. The final selected F_2 individuals will be used to derive the 200 DH or RIL.

When is MAS more efficient than conventional PS for CL?

Comparison of MAS with phenotypic selection (PS) in long-term recurrent selection has been studied through simulation by various authors (e.g., Bernardo and Charcosset 2006). Our case study was more related to the "tactical scenarios" faced by wheat breeders when developing new parental populations and/or commercial lines, i.e., the breeders have only one cycle to try to obtain the target genotypes. For the genetic and breeding model used in this study, we found PS was more efficient than MAS using identified QTL (Table 3), which is consistent with previous practical studies, e.g., Moreau et al. (2004) and Davies et al. (2006).

To superficially determine the role of experimental error and marker-trait linkage in the comparison of MAS with PS, we undertook several other simulations (Table 3). MAS was only more efficient than PS when heritability was low (Table 3), consistent with Lande and Thompson (1990). However, QTL mapping becomes more difficult for low heritability traits, and known QTL information may be

less in this case, which will, in turn, make MAS less efficient. So, there is somewhat of a dilemma in attempting to both map and utilize QTL for complex traits.

Inter-genic interactions are likely to be important in the control of complex traits, yet little is understood about how large a role of epistasis plays (Li et al. 2008). Cooper et al. (2005) demonstrated for a large number of genetic models used for recurrent selection of several complex traits, that MAS effect on selection is sensitive to the robustness of the underlying genetic model, i.e., in our case, too, MAS will be even less efficient if digenic or higher order interactions exist, since in most cases MAS was designed to select individual genes, not specific gene combinations at two or more loci. Genome-wide selection has been proposed to overcome the inaccuracies in QTL mapping, such as biased estimations of QTL positions and effects (Bernardo and Yu 2007). Under the simplest additive model, MAS can be much more efficient than PS (Bernardo and Yu 2007). However, this may not be true with more complicated genetic models, such as epistasis and QTL by environment interactions.

Conclusion

Simulations in this study indicate that PS recovers a greater number of desirable lines than MAS for selecting long CL in wheat using the same population size and our high-heritability growth cabinet screen. However, when combined with PS, MAS based on partially identified QTL can be as efficient as PS. Particularly in a case where a cheaper, higher throughput but lower heritability field CL screen is an alternative to the higher cost, lower throughput growth cabinet screen, the sequential selection of MAS followed by PS could greatly reduce overall breeding costs and time. Therefore, molecular markers, though accounting for part of the genetic variation, are still useful in breeding wheat with long CL, when combined with phenotypic selection in some common real-world breeding scenarios.

A single cross was used to investigate the efficiency of MAS based on partially known gene information. Plant breeders typically make several hundred crosses per year using many parents, making it more difficult to determine whether it will be beneficial to use linked markers for QTL that were developed in other populations. To be used, the linkage status of large numbers of parents would then need to be known, and the optimum strategy for any cross would depend on the distribution of positive and negative alleles in the parents. However, the multi-generation selection scheme and the use of MAS combined with PS provided in this study should be applicable to other crosses where several major gene loci and many QTL are segregating.

Acknowledgments The authors wish to thank two anonymous reviewers and the editor for their constructive suggestions and comments on a previous version of the manuscript. This research was funded by the National 863 Programs of China (grant No. 2006AA10Z1B1) and the Generation Challenge Programme of CGIAR (http://www.generationcp.org).

References

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. Nat Rev Genet 3:11–21

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

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

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

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breed 15:75–85

Cooper M, Podlich DW, Smith OS (2005) Gene-to-phenotype models and complex trait genetics. Aust J Agri Sci 56:895–918

Davies J, Berzonsky WA, Leach GD (2006) A comparison of marker-assisted and phenotypic selection for high grain protein content in spring wheat. Euphytica 152:117–134

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32

Dubcovsky J (2004) Marker-assisted selection in public breeding programs: the wheat experience. Crop Sci 44:1895–1898

Eagles HA, Bariana HS, Ogbonnaya FC, Rebertzke GJ, Hollamby GJ, Henry RJ, Henschke PH, Carter M (2001) Implementation of markers in Australian wheat breeding. Aust J Agric Res 52: 1349–1356

Eagles HA, Hollamby GJ, Gororo NN, Eastwood RF (2002) Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. Aust J Agric Res 53:367–377

Eathington SR, Crosbie TM, Edwards MD, Rieter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47(S3):154–163

Ellis MH, Spielmeyer W, Gale K, Rebetzke GJ, Richards RA (2002) Perfect markers for the *Rht-B1b* and *Rht-D1b* dwarfing mutations in wheat (*Triticum aestivum* L.). Theor Appl Genet 105:1038–1042

Ellison FW, Brown GN, Mares DJ, Moore SG, O'Brien L (1994) *Triticum aestivum* spp. *vulgare* (bread wheat) cv. Sunstate. Aust J Exp Agric 34:869

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex

Korzun V, Roder MS, Ganal MW, Worland AJ, Law CN (1998) Genetic analysis of the dwarfing gene (*Rht8*) in wheat. Part I. Molecular mapping of *Rht8* on the short arm of chromosome 2D of bread wheat (*Triticum aestivum*). Theor Appl Genet 96:1104–1109

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

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

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

Li H, Ribaut J-M, 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

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

Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in plant improvement. Adv Agron 46:39–90

Podlich DW, Cooper M (1998) QU-GENE: a platform for quantitative analysis of genetic models. Bioinformatics 14:632–653

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

Rebetzke GJ, Richards RA (2000) Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. Aust J Agric Res 51:235–245

Rebetzke GJ, Richards RA, Fischer VM, Mickelson BJ (1999) Breeding long coleoptile, reduced height wheats. Euphytica 106:159–168

Rebetzke GJ, Appels R, Morrison A, Richards RA, McDonald G, Ellis MH, Spielmeyer W, Bonnett DG (2001) Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (*Triticum aestivum* L.). Aust J Agric Res 52:1221–1234

Rebetzke GJ, Richards RA, Sirault XRR, Morrison AD (2004) Genetic analysis of coleoptile length and diameter of wheat. Aust J Agric Res 55:733–743

Rebetzke GJ, Richards RA, Fettell NA, Long M, Condon AG, Botwright TL (2007a) Genotypic increase in coleoptiles length improves wheat establishment, early vigour and grain yield with deep sowing. Field Crop Res 100:10–23

Rebetzke GJ, Ellis MH, Bonnett DG, Richards RA (2007b) Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 114:1173–1183

Richards RA (1992) The effect of dwarfing genes in spring wheat in dry environments. I. Agronomic characteristics. Aust J Agric Res 43:517–527

Spielmeyer W, Sharp PW, Lagudah ES (2003) Identification and validation of markers linked to broad-spectrum stem rust resistance gene *Sr2* in wheat. Crop Sci 43:333–336

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003) Comparison of two breeding strategies by computer simulation. Crop Sci 43:1764–1773

Wang J, van Ginkel M, Trethowan R, Ye G, DeLacy I, Podlich D, Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. Crop Sci 44:2006–2018

Wang J, Eagles HA, Trethowan R, van Ginkel M (2005) Using computer simulation of the selection process and known gene information to assist in parental selection in wheat quality breeding. Aust J Agric Res 56:465–473

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:580–588

Young ND (1999) A cautiously optimistic vision for marker-assisted selection. Mol Breed 5:505–510