Theor Appl Genet (2008) 116:541–553
DOI 10.1007/s00122-007-0689-8

ORIGINAL PAPER

Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments

Joseph Knoll · Gebisa Ejeta

Received: 29 September 2006 / Accepted: 28 November 2007 / Published online: 19 December 2007

Abstract Sorghum [*Sorghum bicolor* (L.) Moench] landraces from China generally exhibit excellent emergence and seedling vigor under cool conditions, and are being used as sources of genes for improvement of seedling cold tolerance in other cultivars. Marker-assisted selection (MAS) could expedite the introgression of genes from landraces into elite lines, however, only a few studies have empirically demonstrated efficacy of MAS for quantitatively inherited agronomic traits. In a preceding study we identified quantitative trait loci (QTL) for early-season performance in a recombinant inbred (RI) population, one parent of which was a cold-tolerant Chinese line, 'Shan Qui Red' (SQR). In this study, three SSR markers (*Xtxp43*, *Xtxp51*, and *Xtxp211*), each representing a QTL, were tested in two new populations: (Tx2794 × SQR F_3) and (Wheatland × SQR BC_1F_3). Individual families were genotyped, and early-season field performance was measured for two years. Statistical analyses showed that the SQR allele of *Xtxp43* had favorable effects on seedling vigor in both populations, and on emergence in the Tx2794 population. A large positive effect of the SQR allele of *Xtxp51* was observed in the Tx2794 population for vigor and emergence. Slight genotype by environment interaction was observed for *Xtxp51* in the Wheatland population. Marker *Xtxp211* had small but significant effects on seedling vigor and emergence in both populations. Various interactions between loci were also significant. This study validated QTL markers in various genetic backgrounds, and demonstrated the utility of MAS for a quantitative trait, early-season cold tolerance, evaluated in the field.

Communicated by H. H. Geiger.

J. Knoll · G. Ejeta (✉)
Department of Agronomy, Purdue University,
915 W. State Street, West Lafayette, IN 47907, USA
e-mail: gejeta@purdue.edu

Introduction

Grain sorghum [*Sorghum bicolor* (L.) Moench] is an important crop in many parts of the world, and ranks third in production among grain crops in the United States, behind maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) (FAO 2004). Sorghum originated in tropical and subtropical regions of Africa (Doggett 1988), and is therefore well adapted to warm growing conditions. Cool temperatures during the early growing season are thus a major limitation to growing sorghum in the northern United States and other temperate areas. However, we have observed that many kaoliang sorghum landraces from China exhibit higher seedling emergence and greater seedling vigor under cool conditions than most sorghum cultivars (Cisse and Ejeta 2003; Singh 1985). These landraces can be an excellent genetic resource for the improvement of cold tolerance in sorghum, though they lack other desirable agronomic characteristics. Directed introgression of seedling cold-tolerance genes from kaoliang sorghums into elite high-yielding sorghum lines, without the introduction of undesirable traits from the donor parents, could be expedited by marker-assisted selection (MAS).

Genetic markers for quantitative traits are commonly identified in mapping populations to enhance selection for cultivar improvement. However, the plant breeding community recognizes the necessity to validate these putative quantitative trait loci (QTL) across various genetic backgrounds before embarking upon marker-assisted

introgression or selection. Dudley (1993) outlined three important factors, which determine whether a QTL-marker association identified in one population will be useful for selection in a different population. These include the presence of marker-QTL segregation (i.e., polymorphism of the QTL and of the marker on which to select), the linkage phase of the marker and QTL of interest, and epistatic interactions between the QTL and various other loci in the new population. Polymorphism of the genetic markers is easy to determine, but the association of a given marker with a quantitative trait in a new population may need to be determined experimentally. Also, other loci that may interact with the introgressed QTL are usually not known, but they can affect the magnitude of expression of the QTL, possibly even obscuring it completely.

One approach to validating genetic markers across populations is to develop multiple mapping populations and perform interval mapping on each to identify QTL, which are common to two or more populations. Haussmann et al. (2002) used this approach to map QTL for stay-green in two populations of sorghum. The donor parent for the stay-green trait was the same for both populations. They identified three QTL, derived from the donor parent, that were common to both populations, and that showed similar effects across environments. Various other QTL, some derived from the non-donor sources, were also identified. Identification of these other, potentially useful, QTL is the main advantage of this approach. Because repeating an interval mapping experiment in an entirely new population to validate QTL may be time-consuming and expensive, selective genotyping (in which only the phenotypic extremes of the population are analyzed) has been shown to be nearly as effective as interval mapping for identifying QTL (Foolad et al. 2001; Ayoub and Mather 2002). Micic et al. (2005) used selective genotyping to validate two QTL for stalk rot resistance in sunflower that had been previously identified in a different population.

While it is possible to identify corresponding QTL across populations using various methods, studies that have directly tested the utility of MAS for quantitatively inherited agronomic traits across populations and/or environmental conditions have been relatively few. There are, however, several examples of such studies that are worth noting (Tanksley and Hweitt 1988; Yousef and Juvik 2002; Zhou et al. 2003). As part of their ground breaking experiments on the use of molecular markers in selection for soluble solids and pH in tomato fruit, Tanksley and Hewitt (1988) tested four markers, representing three chromosomal segments introgressed from a wild relative. They found that one of these segments, significantly associated with increased soluble solids in a previous study by Osborn et al. (1987), caused a slight increase in soluble solids across three populations, while another showed different effects in different genetic backgrounds. They also found that these introgressed segments tended to negatively affect fruit pH and other traits, further highlighting the importance of testing markers prior to their application in breeding. More recently, Zhou et al. (2003) used simple sequence repeat (SSR) markers to select for a major QTL for scab resistance in two different populations of wheat, and achieved a significant reduction in the percentage of infected spikelets in families carrying the QTL. Yousef and Juvik (2002) backcrossed three QTL markers for seedling emergence into three elite sweet corn inbreds, and then tested the effects of the markers in the resulting BC_2F_1 progeny. They found favorable increases in emergence for two of the markers in all three genetic backgrounds, while the third marker showed favorable increases in two of the populations, thus validating the QTL and the utility of MAS for emergence.

In the first part of this study (Knoll et al. 2007, this issue), genetic markers associated with seedling emergence and vigor under cold stress conditions were identified in a recombinant inbred (RI) mapping population, derived from a cross between 'Shan Qui Red' (SQR), a cold-tolerant Chinese kaoliang, and SRN39, a cold-sensitive African caudatum. Three QTL identified in the previous study were chosen for validation in two segregating populations of sorghum over two field environments. The QTL and the SSR markers used to represent them in this analysis were chosen for several reasons. For two of the QTL studied, on SBI-02 and SBI-04 [linkage group designations follow Kim et al. (2005)], the nearest flanking marker was an SSR (*Xtxp211* and *Xtxp51*, respectively). The QTL on SBI-02 was detected only by composite interval mapping (Zeng 1993, 1994), and the non-Chinese parent SRN39 contributed the favorable allele. The QTL detected on SBI-04 for early-sown seedling vigor, near *Xtxp51*, was only significant under simple interval mapping (Lander and Botstein 1989), and SQR contributed the favorable allele. Because they were only significant under one of the two mapping models used, and because their favorable alleles were from different parents, these two QTL were chosen for further validation. Both simple and composite interval mapping procedures detected at least one, and perhaps two, QTL on linkage group SBI-01 for seedling emergence and vigor, with the favorable allele(s) from SQR. SSR marker *Xtxp43* maps to this genomic region. Although several RAPD and one RFLP marker were also mapped to this region, the SSR marker was chosen due to the technical simplicity, co-dominant inheritance, and high level of repeatability of SSR markers. These particular SSR markers were also chosen because they showed distinct polymorphisms between SQR and the other two parents, Tx2794 and Wheatland.

Materials and methods

Genetic material

An introduction from China, SQR, selected for its early-season cold tolerance, was used to generate several segregating populations in cross combinations with other sorghum lines assembled in our breeding program. Two of these breeding populations were chosen for this experiment to validate QTL for early-season cold tolerance identified in a previous RI population. The first test population was created by crossing SQR with a cold-sensitive line, Tx2794. In 2003, approximately 500 random panicles of the F_2 generation were selfed to produce a population of 394 F_3 families. The second population was created by crossing SQR with 'Wheatland', a common commercial seed parent with moderate early-season vigor. The F_1 was backcrossed to 'Wheatland' to produce the BC_1F_1 generation, which was then selfed to produce the BC_1F_2. Approximately 500 random panicles of the BC_1F_2 generation were selfed in 2003 to produce a population of 390 BC_1F_3 families. Each panicle, representing one F_3 or BC_1F_3 family from these populations, was individually threshed and kept for use in this study.

DNA isolation

DNA was isolated from a total of 784 families from the two populations. A sample of approximately 15–20 seeds was taken from each of the F_3 and BC_1F_3 families for DNA analysis. The seeds were pulverized to a fine powder in a Retsch Mixer Mill MM 200 (Retsch, Haan, Germany). DNA was isolated from the pulverized seeds by a CTAB-based mini-prep protocol, similar to the method described by Stewart and Via (1993) with some modifications. One ml extraction buffer [175 mM sorbitol, 65 mM Tris, 5.5 mM EDTA, 10 mM sodium bisulfite, 210 mM NaCl, 10 g/l cetyltrimethylammonium bromide (CTAB), and 1% sarkosyl] pH 8.25 was added to 50–100 mg powdered seed tissue, was thoroughly mixed and then incubated at 65°C for 40 min with occasional gentle mixing. Hydrophobic components and solid material were removed by extraction with 800-µl chloroform:octanol (24:1), followed by centrifugation. The nucleic acid-containing aqueous phase was treated with RNase to remove RNA, and then the remaining DNA was precipitated by addition of 650 µl ice-cold isopropanol, followed by centrifugation. The supernatant was discarded and the DNA pellet was rinsed with 70% ethanol, allowed to dry, and was then resuspended in 100 µl TE buffer. Prior to PCR, the DNA was quantified by using the Hoescht dye fluorescence method in 96-well plates as described by Held (2001a, b), in a Synergy HT Multi-Detection Microplate Reader equipped with KC4 data analysis software (Bio-Tek Instruments, Inc, Winooski, VT, USA).

SSR marker genotyping

Three SSR markers (*Xtxp43*, *Xtxp51*, and *Xtxp211*), whose primers were developed at Texas A & M University by Kong et al. (2000) or Bhattramakki et al. (2000), were analyzed for each sample. Each 25-µl PCR mixture contained 15 ng template DNA, 0.4 µM each forward and reverse primer, 0.05 mM each of dATP, dCTP, dGTP, and dTTP, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris pH 8.9, 0.01% Triton X-100, 0.02% cresol red dye, 12% sucrose, and 2.5 U Taq DNA polymerase. Reactions were carried out in 96-well plates in a PTC-100 Programmable Thermal Controller (MJ Research, Inc., Watertown, MA) or in a Primus 96 thermocycler (MWG-Biotech, Inc., High Point, NC, USA). Reaction conditions were as follows: initial denaturation at 94°C for 60 s, followed by 35 cycles of 94°C (denaturation) for 10 s, 55°C (annealing) for 35 s, and 72°C (extension) for 45 s. This was followed by a final extension at 72°C for 3 min. PCR products were visualized by electrophoresis in 3% Gene-Pure HiRes agarose (ISC BioExpress, Kaysville, UT, USA) gels in TBE buffer for approximately 15 h at 50 V, followed by staining with ethidium bromide. Each gel included PCR products from both parents (SQR, and Tx2794 or Wheatland) to facilitate scoring of the markers.

Field evaluation of seedling vigor and other agronomic traits

Data on SSR markers thus generated were used to classify the F_3 and BC_1F_3 families into 27 different genotypic classes based on the three possible genotypes at each of the three marker loci. Five random entries of each genotypic class were chosen for field evaluation. Where there were fewer than five entries for a particular genotypic class, two, three, or four entries were used. In the Wheatland backcross population, one genotypic class was not represented (Table 1). A replicated field trial of progeny families representing each of the 27 classes was planted at the Purdue University Agronomy Center for Research and Education (ACRE) in West Lafayette, IN on 27 April 2004. A duplicate sample of seed was placed in cold storage, and was planted out for a second year trial on 11 April 2005 in a different field at the same facility. These planting times are regarded as early for sorghum at this location, and provided exposure to cool temperatures (Fig. 1). Each population was sown in a separate experiment, and each

Table 1 Number of entries evaluated in the field in 2004 and 2005 at West Lafayette, IN, in two populations of sorghum, for each of the 27 possible marker-genotypic classes

Genotype			Population	
Xtxp43	*Xtxp51*	*Xtxp211*	SQR × Tx2794 F_3	SQR × Wheatland BC_1F_3
s/s	s/s	s/s	4	2
		s/x	5	4
		x/x	5	0
	s/x	s/s	5	5
		s/x	5	5
		x/x	5	5
	x/x	s/s	5	5
		s/x	5	5
		x/x	5	5
s/x	s/s	s/s	5	5
		s/x	5	5
		x/x	5	5
	s/x	s/s	5	5
		s/x	5	5
		x/x	5	5
	x/x	s/s	5	5
		s/x	5	5
		x/x	5	5
x/x	s/s	s/s	3	5
		s/x	5	2
		x/x	5	4
	s/x	s/s	5	5
		s/x	5	5
		x/x	5	5
	x/x	s/s	5	5
		s/x	5	5
		x/x	4	5

[a] Where possible, five entries per genotypic class were sampled

s SQR allele, *x* Tx2794 or Wheatland allele

experiment was laid out in a randomized complete block design with three replications. The parental lines were also included as checks for phenotypic assessment and comparison of seedling vigor. A border row of a common commercial hybrid was planted between each plot, and also along all edges of the experiment.

Each 5-m experimental plot was sown with 100 seeds. Data on four early-season traits were recorded. Seedling emergence (total number of seedlings emerged in each plot) was counted twice during the early growing season. Early emergence was counted at approximately 15 days after planting in both years. A final count was taken at 23 days after planting in 2004 and 43 days after planting in 2005. Seedling vigor scores were assessed three times, at approximately 10-day intervals throughout the early season, rather than only once, to gain a better estimate of seedling vigor, which may appear to change as the seedlings grow. Vigor scores were assessed visually on a scale of 1–5 in increments of 0.5, with 1 representing very high vigor and 5 representing very poor vigor, as described by Maiti et al. (1981). Vigor scores are based on size and physical appearance of the seedlings, and are not based on seedling emergence percentages. Stand biomass was assessed by harvesting the above-ground portion of all seedlings within the first 1 m of each plot approximately 30 days after planting in 2004, and 45 days after planting in 2005, and then measuring the total dry mass. This measurement is reflective of seedling emergence, survival, and growth. The plants generally emerged and grew more slowly in 2005 due to the colder, drier conditions (Fig. 1), so they were allowed to grow longer before the final emergence was counted and biomass was harvested.

Other agronomic traits were also measured. Flowering time (when half the panicles in the plot showed dehisced anthers halfway down the panicle) was determined for each plot. Average plant height, measured to the nearest 5 cm for each plot, was assessed at maturity. Grain yield was recorded on the middle 3 m of each plot, with panicles harvested by hand and later dried at approximately 35°C with forced air circulation for one week. The dried panicles were bulk-threshed in a Low Profile Plot Thresher (ALMACO, Nevada, IA, USA), and the grain from each plot was weighed, and recorded as plot grain yield. A small subsample of grain was saved from each plot to determine 1,000-grain weight.

Data analysis

The first, second, and third vigor score readings were analyzed separately, and an average of the three was also analyzed. Despite averaging, vigor scores still have a non-linear, but ordinal, nature, which suggests that a rank-based transformation may be appropriate to improve the analysis. One such transformation is the van der Waerden transformation, which fits ordinary ranks to a normal distribution (Conover 1999). It has been suggested that the use of normal scores, instead of ordinary ranks, allows for the analysis of interactions (Conover 1999; Mansouri and Chang 1995). In SAS v. 9.1 (SAS Institute 2003), the PROC RANK procedure was applied to rank the vigor scores; tied ranks were assigned an average (midrank) value, which is the default for this procedure in SAS. The NORMAL = VW option was selected to generate van der Waerden (normal) scores from the ranks. This option fits the ranks to a standard normal distribution with mean $\mu \approx 0$ and standard deviation $\sigma \approx 1$.

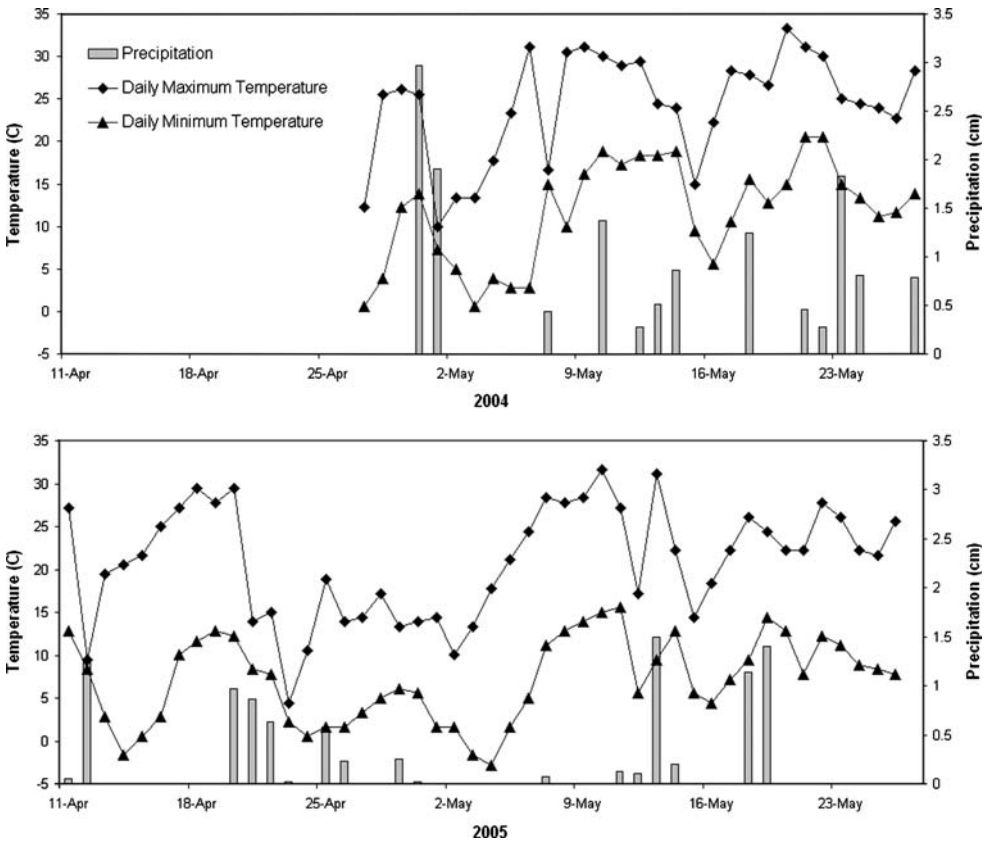

Fig. 1 Air temperatures (daily maximum and minimum) and precipitation during the early growing season, from time of planting to time of biomass harvest, at West Lafayette, IN in 2004 (*top*) and 2005 (*bottom*). Source: Indiana State Climate Office, Purdue University, West Lafayette, IN, USA

An analysis of variance (ANOVA) was applied to the early-season data to determine the effects of the markers with respect to emergence, vigor, and biomass. Using the PROC GLM procedure in SAS, the following linear model was applied in which all independent variables were treated as categorical, rather than quantitative:

$$\begin{aligned} y = {} & \mu + Y + R(Y) + M_1 + M_2 + M_3 + M_1M_2 + M_2M_3 \\ & + M_1M_3 + M_1M_2M_3 + F(M_1M_2M_3) + YM_1 + YM_2 \\ & + YM_3 + \cdots + \varepsilon \end{aligned}$$

where y is the dependent variable (emergence, normalized vigor score, or biomass), μ is the overall mean, Y indicates the year effect, $R(Y)$ is the replication nested within year, M_1 is the effect of the first marker (*Xtxp43*), M_2 is the effect of the second marker (*Xtxp51*), and M_3 is the effect of the third marker (*Xtxp211*). Each marker has three possible genotypes. $F(M_1M_2M_3)$ is the effect of the individual F_3 or BC_1F_3 families, which is nested within the three-way marker interaction term. The three-way marker interaction represents the 27 genotypic classes. The error term is represented by ε. The other higher-order interactions between multiple markers and year were also included in the model. This model is similar to that used by Yousef and Juvik (2002). Data from the parental lines were not included in these analyses. Correlations between early-season seedling traits and agronomic traits within the populations were calculated using the PROC CORR procedure in *SAS*. Correlations were calculated using untransformed data.

Results

Seedling vigor scores were assessed three times during the early season. Each was analyzed separately, and the average of the three was also analyzed. The results of the analyses of variance were very similar for each of the three readings and for the averaged scores. A high degree of correlation was found between the first, second, and third vigor scores ($R \geq 0.52$). Slightly higher R^2 values were obtained from ANOVA when the average of the three vigor scores was used. In this study the rank-based van der Waerden transformation (Conover 1999) was used to analyze vigor scores. Ranking can be performed within replicates or years, analogous to the Friedman test (Conover 1999), or over the entire experiment. In this study analysis of normal scores, whether ranked within years, replications, or over entire experiments, gave nearly identical results as analysis of untransformed vigor scores (data not shown), most likely because the distributions of the raw vigor scores were reasonably close to normal. Analysis of the normal scores, however, yielded slightly higher R^2

values than the untransformed data, indicating a slightly better fit to the model.

Tx2794 × SQR F_3 population

ANOVA showed a significant year effect for seedling biomass and 15-day emergence, but not for final emergence or seedling vigor, in the Tx2794 population (Table 2). Planting was 16 days earlier, and growing conditions were generally colder and drier, in 2005 than in 2004 (Fig. 1), resulting in an overall decrease in seedling biomass in 2005. Although there was an overall effect of year on most traits, the data from 2004 and 2005 are highly correlated with each other, as shown in Table 3. Markers *Xtxp43* and *Xtxp51* had significant effects on all four early-season traits, with *Xtxp51* having the greater effect on 15-days emergence, final emergence, and biomass. These two markers had similar effects on seedling vigor score (Table 4). SQR-derived alleles of both markers were associated with more favorable values for all four traits. Marker *Xtxp211* had a small but significant effect on seedling vigor. Lower vigor score (improved vigor) was associated with the SQR allele at this locus (Table 4). A slight effect on final seedling emergence was also noted, again favoring the SQR allele for this marker. The effect of *Xtxp211* on 15-days emergence was very small and was not significant for stand biomass (Tables 2, 4).

The interactions of *Xtxp43* and *Xtxp51* with respect to all early-season traits were highly significant (Table 2). The highest emergence counts (at 15 days or final count) were observed among F_3 families that were homozygous for the SQR allele (s/s) at both markers (43.3% in 2004 and 43.9% in 2005 at final count), or that were s/s at one marker and heterozygous (s/t) at the other (36.5–44.6% at final count, Fig. 2a). The lowest emergence counts were observed among entries lacking SQR alleles at these two markers (20.1% in 2004 and 20.7 in 2005 at final count, Fig. 2a). The lowest vigor scores were associated with F_3 families that were s/s for *Xtxp43* and s/s or s/t for *Xtxp51* (−0.907 to −0.381, expressed as normal scores, Fig. 2a). The highest vigor scores (associated with poorest seedling vigor) were observed among entries lacking SQR alleles at

Table 2 Mean squares for four seedling traits in two populations of sorghum grown at West Lafayette, IN in 2004 and 2005

Source	Tx2794 population					Wheatland population				
	df	15-days emergence	40-days emergence	Vigor[a]	Biomass	*df*	15-days emergence	40-days emergence	Vigor[a]	Biomass
Y	1	9,699.44***	214.78 NS	0.294 NS	18.431***	1	68,418.64***	10,995.07***	12.255***	63.777***
R(Y)	4	468.33	45.03 NS	0.505 NS	2.338***	4	76.53 NS	580.55***	4.048***	1.514***
M_1	2	1,649.32	2,308.07***	21.101***	1.007**	2	426.57**	455.18**	7.760***	0.238 NS
M_2	2	8,893.70	15,996.55***	51.122***	8.147***	2	625.57***	646.15***	0.761 NS	0.016 NS
M_3	2	275.32	601.46**	3.660***	0.417 NS	2	229.22*	159.44 NS	9.293***	0.750 NS
$M_1 \times M_2$	4	821.40	1,089.64***	5.588***	0.631**	4	799.21***	1,029.36***	2.287**	0.412 NS
$M_1 \times M_3$	4	697.21	562.68***	4.420***	0.803***	4	660.54***	646.37***	3.339***	0.594 NS
$M_2 \times M_3$	4	350.97	624.48***	1.656**	0.168 NS	4	218.64*	530.31***	1.590*	0.830*
$M_1 \times M_2 \times M_3$	8	298.98	395.92***	3.791***	0.468**	7	577.43***	879.13**	1.796**	0.574 NS
$F(M_1 \times M_2 \times M_3)$	104	449.55	680.90***	2.275***	0.406***	96	357.15***	425.42***	2.584***	0.729***
$Y \times M_1$	2	59.85	53.96 NS	0.509 NS	0.003 NS	2	136.95 NS	93.57 NS	0.055 NS	0.075 NS
$Y \times M_2$	2	50.39	181.82 NS	0.861 NS	0.226 NS	2	267.12*	54.08 NS	2.674**	0.080 NS
$Y \times M_3$	2	26.08	51.09 NS	0.126 NS	0.100 NS	2	34.70 NS	54.48 NS	0.199 NS	0.049 NS
$Y \times M_1 \times M_2$	4	188.70	196.62 NS	1.166*	0.371*	4	56.51 NS	56.24 NS	0.764 NS	0.230 NS
$Y \times M_1 \times M_3$	4	182.20	221.84*	0.864 NS	0.359*	4	218.65*	271.23**	1.654*	0.175 NS
$Y \times M_2 \times M_3$	4	24.05	113.96 NS	0.361 NS	0.158 NS	4	85.02 NS	93.47 NS	0.315 NS	0.359 NS
$Y \times M_1 \times M_2 \times M_3$	8	130.87	87.88 NS	0.990*	0.190 NS	7	216.23**	141.84*	1.066 NS	0.371 NS
Error	611	85.10	91.02	0.435	0.150	572[b]	67.89	65.72	0.552	0.304
R^2		0.641	0.691	0.647	0.536		0.773	0.678	0.549	0.500

Y year, *R* replicate, *M1* Xtxp43, *M2* Xtxp51, *M3* Xtxp211, *F* F2:3 or BC1F3 family, *NS* not significant

*, **, and *** indicate significance at $\alpha = 0.05$, 0.01, and 0.001, respectively

[a] Vigor scores were averaged over three readings and then transformed to normal (van der Waerden) scores

[b] Error df for biomass in Wheatland population = 555 due to missing data points

Table 3 Pearson correlation coefficients (R) of trait values between years 2004 and 2005 within the two test populations

Trait	Population	
	Tx2794 × SQR F_3	Wheatland × SQR BC_1F_3
15 days % emergence	0.679	0.565
Final % emergence	0.762	0.658
Seedling vigor score	0.695	0.576
Seedling biomass	0.565	0.411
Flowering time	0.794	0.763
Mature plant height	0.865	0.870
Grain yield	0.714	0.616
1,000-grain weight	0.799	0.883

All correlations were significant at $\alpha < 0.001$

these two loci (0.993 in 2004 and 0.728 in 2005, expressed as normal scores, Fig. 2a). The greatest biomasses were observed among F_3 families that were homozygous for the SQR alleles of both of these markers (1.17 g in 2004 and 0.74 g in 2005). In 2004, however, those families that were t/t homozygous for *Xtxp43* and heterozygous for *Xtxp51* showed statistically similar, though still lower, biomasses (0.99 g, Fig. 2a). High biomasses were also observed in the families that were s/s homozygous at one marker and heterozygous at the other (0.87–0.93 g in 2004 and 0.55–0.64 g in 2005, Fig. 2a). The interactions between single markers and years (i.e., genotype by environment interactions) were not significant for any early-season traits, including biomass, though some interactions between multiple markers and years were slightly significant at $\alpha = 0.05$ (Table 2). In general, three-way interactions tended to favor a greater number of SQR alleles at multiple loci with respect to higher emergence, lower vigor scores, and greater biomass. The lowest emergence counts, highest vigor scores, and lowest biomasses were all observed among the F_3 families that completely lacked any SQR marker alleles (Fig. 3), affirming the robustness of these markers in this population.

Wheatland × SQR BC_1F_3 population

In the Wheatland backcross population, *Xtxp43* showed a significant effect for seedling vigor with homozygosity of the SQR marker allele associated with better seedling vigor. However, it also appeared to have a slight negative effect on final seedling emergence percentage, although this did not significantly affect stand biomass (Table 4). Marker *Xtxp211* showed significant effects on seedling vigor, 15-days emergence, final emergence, and biomass,

Table 4 Means for seedling traits, based on QTL marker genotypes in two population of sorghum grown at West Lafayette, IN in 2004 and 2005

Tx2794 × SQR F_3

Marker	Genotype[a]	Emergence		Vigor	Biomass
		% at 15 days	% Final	Normal Score[b]	g
Xtxp43	t/t	23.05 a	31.16 a	0.261 a	0.609 a
	s/t	24.50 a	33.63 b	0.063 b	0.619 a
	s/s	28.03 b	36.95 c	−0.317 c	0.728 b
Xtxp51	t/t	18.61 a	24.99 a	0.496 a	0.460 a
	s/t	27.29 b	37.50 b	−0.179 b	0.695 b
	s/s	29.89 c	39.50 c	−0.327 c	0.807 c
Xtxp211	t/t	24.32 a	32.58 a	0.084 a	0.615 a
	s/t	25.68 a	34.36 ab	0.018 ab	0.667 a
	s/s	25.63 a	34.92 b	−0.106 b	0.674 a

Wheatland × SQR BC_1F_3

Marker	Genotype[a]	Emergence		Vigor	Biomass
		% at 15 days	% Final	Normal score[b]	g
Xtxp43	w/w	26.81 a	36.62 a	0.103 a	0.894 a
	s/w	25.78 a	34.24 b	0.060 a	0.899 a
	s/s	27.46 a	35.66 ab	−0.194 b	0.914 a
Xtxp51	w/w	27.79 a	35.06 a	−0.012 a	0.912 a
	s/w	24.91 b	34.18 a	0.041 a	0.890 *a*
	s/s	27.37 a	37.85 b	−0.041 a	0.902 a
Xtxp211	w/w	26.72 ab	35.39 ab	−0.004 a	0.893 ab
	s/w	28.05 a	36.91 a	−0.232 b	0.985 a
	s/s	25.12 b	34.10 b	0.228 c	0.826 b

Within markers, means with the same letter are not significantly different when tested using Tukey's Studentized Range (HSD) test at $\alpha = 0.05$

[a] *s* Shan Qui Red allele, *t* Tx2794 allele, *w* Wheatland allele

[b] Lower score indicates more vigorous seedlings

though in each case the heterozygous class appeared have the most favorable trait expression.

Significant year effects were observed for all four early-season traits in this population (Table 2). The average emergence at 15 days or final count was lower in 2005 than in 2004. Average biomass was also less and average vigor scores were higher in 2005. As in the Tx2794 population, data between years were highly correlated (Table 3), and most interactions between markers and year were small or not significant (Table 2). However, the interaction between *Xtxp51* and year was slightly significant for 15-days emergence and seedling vigor ($P = 0.020$ and $P = 0.008$, respectively). In 2004, an increase in the number of SQR alleles at this locus was associated with a slight

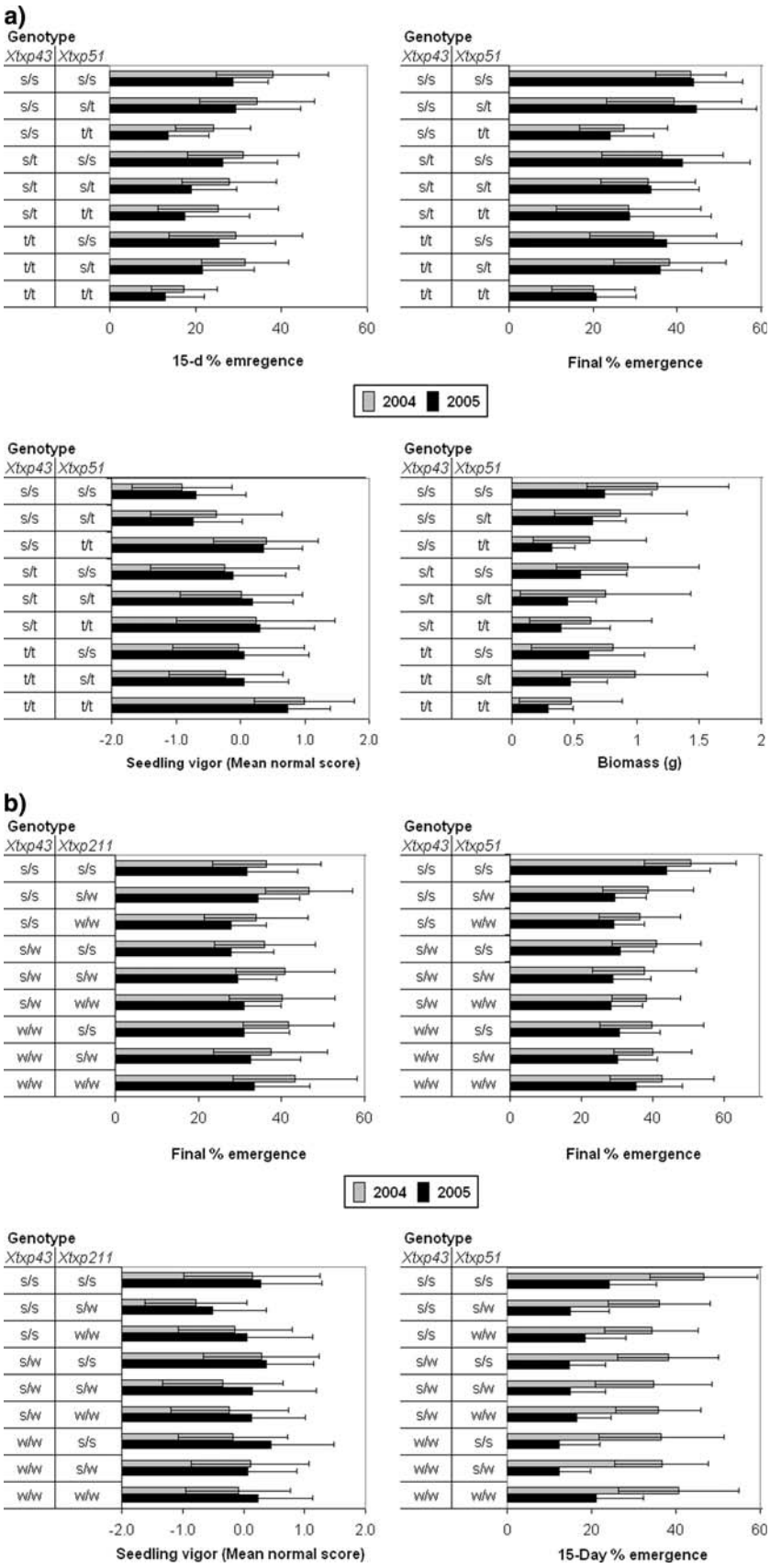

Fig. 2 Mean values for early-season seedling traits for varying genotypes at both *Xtxp43* and *Xtxp51* or *Xtxp211* marker loci in the Tx2794 × SQR F_3 (**a**) or the Wheatland × SQR BC_1F_3 (**b**) population of sorghum at West Lafayette, IN in 2004 and 2005. (*s* Shan Qui Red allele, *t* Tx2794 allele, *w* Wheatland allele.). *Error bars* represent one standard deviation of the mean

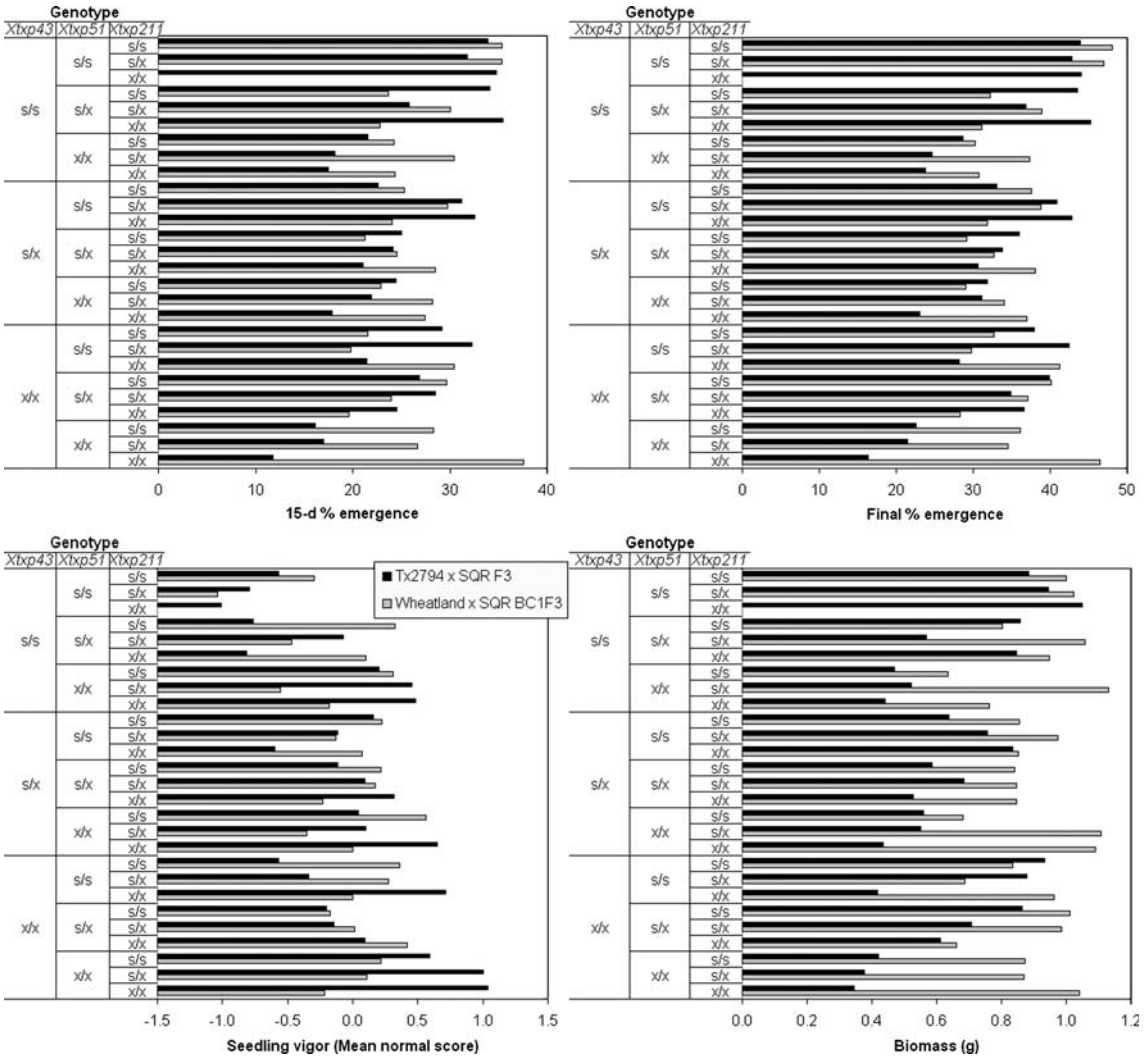

Fig. 3 Mean values for early-season seedling traits for all possible genotypes at the three QTL marker loci (*Xtxp43*, *Xtxp51*, and *Xtxp211*) in the two populations of sorghum (Tx2794 × SQR F_3 and Wheatland × SQR BC_1F_3) at West Lafayette, IN in 2004 and 2005. (*s* Shan Qui Red allele, *x* Tx2794 or Wheatland allele. Note: s/s, s/s, x/x genotype was not represented in the Wheatland population)

improvement in seedling vigor, but in 2005 improved vigor was associated with the Wheatland allele. A similar interaction was observed for 15-days emergence. In 2004 the highest emergence at 15 days was observed among the SQR homozygotes of *Xtxp51*, but in 2005, homozygotes of the Wheatland allele were favored. The main effect of *Xtxp51* taken across both years; however, was not significant for vigor, but it was significant for 15-days emergence (Tables 2, 4). Marker *Xtxp51* had a more definite effect on the final emergence count. Higher final emergence was associated with homozygosity of the SQR allele at *Xtxp51*. However, this marker had no apparent effect on seedling stand biomass.

Marker *Xtxp43* showed significant interactions with the other markers in the Wheatland backcross population, with respect to vigor and emergence. The highest emergence counts at 15 days and final count were observed among the families that were homozygous for the SQR alleles of both *Xtxp43* and *Xtxp51* (50.7% in 2004 and 43.9% in 2005 at final count, Fig. 2b), or that were homozygous for the SQR allele at *Xtxp43* and heterozygous for *Xtxp211* (46.7% in 2004 and 34.3% in 2005 at final count, Fig. 2b). Surprisingly, lowest emergence counts were observed among the families that were s/s for *Xtxp43* and w/w for *Xtxp51* (36.3% in 2004 and 29.2% in 2005 at final count, Fig. 2b). The lowest seedling vigor scores were observed in the families that were homozygous for the SQR allele of *Xtxp43* and heterozygous for *Xtxp211*, especially in 2005 (−0.788 in 2004 and −0.517 in 2005 expressed as normal scores, Fig. 2b). High vigor scores (indicative of poor seedling vigor) were observed in families that were heterozygous at *Xtxp43* and

Table 5 Pearson correlation coefficients (*R*) between various traits within years for two population of sorghum grown at West Lafayette, IN in 2004 and 2005

			15-days emergence	Final emergence	Seedling vigor	Biomass	Flowering time	Height	Yield
Tx2794 × SQR F_3	2004	Final emergence	0.955***						
		Seedling vigor	−0.846***	−0.823***					
		Biomass	0.644***	0.628***	−0.700***				
		Flowering time	−0.423***	−0.392***	0.509***	−0.337***			
		Height	0.430***	0.431***	−0.526***	0.404***	−0.235***		
		Yield	0.609***	0.632***	−0.670***	0.498***	−0.383***	0.587 ***	
		1000-grain weight	−0.265***	−0.295***	0.130**	NS	0.126*	0.244 ***	NS
	2005	Final emergence	0.807***						
		Seedling vigor	−0.707***	−0.707***					
		Biomass	0.637***	0.678***	−0.695***				
		Flowering time	−0.395***	−0.470***	0.395***	−0.347***			
		Height	0.340***	0.214***	−0.333***	0.261***	NS		
		Yield	0.442***	0.425***	−0.464***	0.315***	−0.254***	0.560***	
		1000-grain weight	−0.165**	−0.324***	NS	−0.102*	0.171***	0.294***	0.108 *
Wheatland × SQR BC_1F_3	2004	Final emergence	0.964***						
		Seedling vigor	−0.762***	−0.728***					
		Biomass	0.554***	0.541***	−0.641***				
		Flowering time	−0.388***	−0.369***	0.454***	−0.280***			
		Height	0.247***	0.238***	−0.330***	0.253***	NS		
		Yield	0.400***	0.390***	−0.378***	0.173**	−0.204***	0.368 ***	
		1000-grain weight	−0.292***	−0.315***	0.115*	NS	0.237***	0.196***	0.124*
	2005	Final emergence	0.705***						
		Seedling vigor	−0.635***	−0.592***					
		Biomass	0.493***	0.586***	−0.548***				
		Flowering time	−0.329***	−0.322***	0.303***	−0.195***			
		Height	NS	NS	−0.180***	NS	0.327***		
		Yield	0.251***	0.295***	−0.252***	0.123*	NS	0.306***	
		1,000-grain weight	−0.189***	−0.292***	NS	NS	0.286***	0.223***	0.129*

NS not significant

*, **, and *** indicate significance at $\alpha = 0.05$, 0.01, and 0.001, respectively

homozygous s/s for *Xtxp211* (0.289 in 2004 and 0.379 in 2005, expressed as normal scores, Fig. 2b). However, in 2005 the highest vigor scores were found in the families that were w/w for *Xtxp43* and s/s for *Xtxp211* (0.444, normal score, Fig. 2b). Three-way marker interactions were also significant in this population. The highest emergence percentages at 15 days or final count were observed among the families that were homozygous s/s at *Xtxp43* and *Xtxp51*. However, the families lacking any SQR alleles at any markers showed similarly high emergence. Most other families carrying only one or two SQR marker alleles had lower emergence (Fig. 3). A similar pattern was observed for seedling vigor, but the differences between the classes were not as significant, and several other classes had similarly low vigor scores (Fig. 3). The three-way marker interaction for biomass is not significant in this population (Table 2).

Correlations among traits

All early-season traits (15-days and final emergence, vigor scores, and stand biomass) were highly correlated with each other in both populations in both years. Fifteen-day and final emergence percentages were always very highly correlated ($R \geq 0.705$). Seedling vigor scores, though not based upon seedling emergence, were also highly correlated with this trait. The correlation between vigor and emergence was generally higher in the Tx2794 population than in the Wheatland population. Significant correlations

were also observed between early-season traits and other agronomic traits. Emergence, seedling vigor, and biomass were correlated with eventual grain yield, but more strongly so in the Tx2794 population. Correlations between 1,000-grain weight and other traits were generally weak or not significant (Table 5).

Discussion

All three QTL markers were shown to retain influence in different genetic backgrounds than the one in which they were initially identified. In the original QTL analysis, the SQR allele of *Xtxp43* was shown to favor seedling emergence and seedling vigor. In this study, the effect of *Xtxp43* on seedling vigor and emergence was validated in the Tx2794 background under early-season planting. This effect was further confirmed by the positive effect of *Xtxp43* on stand biomass. This marker will be useful for introgressing early-season cold tolerance from SQR into Tx2794. This marker also showed significant effects for vigor and emergence in the Wheatland background, though its effects were not as pronounced as in the Tx2794 population. There may be a slight negative effect of the SQR allele on emergence, but it has a favorable effect on seedling vigor. Also, it shows favorable interactions with the other two markers.

Marker *Xtxp51* showed the greatest effect in the Tx2794 background, where the homozygote of the SQR allele exhibited superior seedling vigor, 15-days emergence, final emergence, and stand biomass. This was unexpected because this marker was not significant for emergence in QTL analysis of the original RI population. The effect of this marker (*Xtxp51*) in the Wheatland background was, however, not as clear. It had a slight effect on emergence, with the SQR allele positively associated with final emergence. The main effect of this marker was not significant for seedling vigor, but there was a small but significant interaction between *Xtxp51* and year for seedling vigor in the Wheatland background. In 2005, which was colder, the SQR allele was associated with decreased seedling vigor, and the opposite effect was observed in 2004, which was warmer. This genotype by environment interaction obscured the main effect in the analysis of variance. In the original QTL analysis, the SQR allele of this QTL was associated with improved seedling vigor under early planting, but not for late planting. This QTL appears to be highly influenced by temperature in some genetic backgrounds, but given the magnitude of its effect in the Tx2794 background, selection for the SQR allele at this marker should result in a significant improvement in emergence, seedling vigor, and seedling biomass, at least in this background.

Marker *Xtxp211* showed relatively small, but statistically significant, effects in both genetic backgrounds. An effect on seedling vigor and final emergence in the Tx2794 background was noted, but the SQR allele was favored for both of these traits, which is contrary to previous findings in the RIL mapping population (Knoll et al. 2007, this issue), in which decreased emergence was associated with the SQR allele at this locus. However, in the Wheatland background, the SQR-homozygotes of *Xtxp211* had the highest vigor scores (indicative of poor seedling vigor), lowest emergence (15-days or final), and lowest biomasses. Curiously, though, the heterozygote was slightly favored for all these traits. Overall, the effect of *Xtxp211* is relatively small compared to the other two markers, and it is unlikely that selection at this marker will result in a significant improvement in early-season performance.

As expected, many three-way marker interactions were significant. Because there are 27 possible genotypic classes, the interpretation of these interactions is somewhat difficult to summarize. However, in the Tx2794 population, three-way interactions tended to reinforce the conclusions drawn from the two-way interactions. In general, higher emergence, lower vigor scores, and greater biomass were associated with a greater number of SQR alleles at multiple loci. This trend was similar in the Wheatland population, with multiple SQR alleles being favored. There was one major exception: the triple homozygotes of the Wheatland allele tended to perform as well as the families carrying many more SQR alleles at multiple loci. In contrast, in the Tx2794 population, those families lacking any SQR alleles consistently performed very poorly, as expected. Although the three QTL have a significant effect on seedling performance, there are probably other unidentified QTL that control a large portion of the variability between F_3 and BC_1F_3 families. Small sample sizes may have contributed to the variability seen in the three-way marker interactions. At most five families were represented in each three-way genotypic class, whereas up to 15 individual families were represented in each two-way marker class, and up to 45 families were represented in each sample when only one marker was analyzed. Thus, the performance of multiple markers in combination may require further verification with more individual families represented or in a more advanced backcross generation where the genetic backgrounds between families would otherwise be more uniform.

Our observation that some QTL-associated markers affect multiple seedling traits is supported by the fact that the early-season traits (15-days and final emergence, vigor scores, and stand biomass) were highly correlated with each other. These early-season traits were also significantly correlated with several other agronomic traits, which may

have significant implications in breeding. For example, emergence, seedling vigor, and biomass were strongly correlated with eventual grain yield. This positive correlation is most likely a reflection of early-season stand establishment, not necessarily individual plant yield potential, and it highlights the importance of improving early-season performance to enhance eventual crop productivity. Correlations between 1,000-grain weight and yield were generally weak or not significant. In this study there were many plots in which the number of plants was below the optimum density. In these plots grain yield was more dependent upon plant number (i.e., emergence) than grain size. Correlations between some traits were not completely consistent across populations and environments. For example, mature plant height was significantly associated with all early-season traits in both populations in 2004, but in 2005 the correlations between mature plant height and early-season traits were weak or not significant in the Wheatland population. The correlation between mature plant height and flowering time was also not consistent. The correlations between 1,000-grain weight and various other traits were also inconsistent, possibly due to environmental variation during the grain-filling stage. The fact that correlations are not always consistent across populations and environments is indicative of the environmental variability and genotype by environment interactions associated with these traits. This is not unexpected with quantitative agronomic traits.

Overall, the effects of the QTL-associated markers were more pronounced in the SQR × Tx2794 F_3 population than in the SQR × Wheatland BC_1F_3 population. One reason for this difference is probably due to the greater difference in cold tolerance between the parents in the first population. Tx2794 shows lower emergence and is much less vigorous in the seedling stage than Wheatland. The effects of the QTL from SQR may be obscured more in the Wheatland background due to its moderate level of cold tolerance. Also, the effects of the QTL were observed in the BC_1F_3 generation in the Wheatland background, and the $F_{2:3}$ in the Tx2794 background. However, we do not suggest any effect of these generation differences on the ability to validate markers, as the genetic backgrounds of the two populations are so different.

Nonetheless, this study has shown that the effects of QTL can be validated in F_2- or BC_1F_2-derived segregating populations. This approach is similar to that used by Yousef and Juvik (2002), who validated three QTL for seedling emergence in three different BC_2F_1 populations of sweet corn in multiple environments, though our ANOVA model was slightly different. A straightforward ANOVA model can be used to test the main effects of selected QTL-associated markers, while also identifying potentially important interactions between loci, or genotype by environment interactions when tests are conducted in multiple years or locations. This method could be extended to virtually any generation of segregating material, including advanced backcross generations, and could also be used to validate markers in selectively genotyped (Foolad et al. 2001; Ayoub and Mather 2002; Micic et al. 2005) or bulked-segregant (Michelmore et al. 1991; Darvasi and Soller 1994) populations.

Acknowledgments We are grateful for the assistance provided by several colleagues. Terry Lemming assisted in planting and grain harvesting. Matthew Erlich, Joe Johnstone, Erica Hoenie, Tina Suen, and Zenbaba Gutema assisted with emergence counts, biomass measurements, and grain harvest and threshing. Funding for this project was provided by International Sorghum and Millets (INTSORMIL)-USAID Grant # DAN254-G-00-002.

References

Ayoub M, Mather DE (2002) Effectiveness of selective genotyping for detection of quantitative trait loci: an analysis of grain and malt quality traits in three barley populations. Genome 45:1116–1124

Bhattramakki D, Dong J, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Cisse N, Ejeta G (2003) Genetic variation and relationships among seedling vigor traits in sorghum. Crop Sci 43:824–828

Conover WJ (1999) Practical nonparametric statistics, 3rd edn. Wiley, New York

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373

Doggett H (1988) Sorghum, 2nd edn. Wiley, New York

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

FAO (2004) Food and Agriculture Organization of the United Nations, Statistics Division. http://www.fao.org/es/ess/index_en.asp. (Cited 12 Sept. 2006)

Foolad MR, Zhang LP, Lin GY (2001) Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. Genome 44:444–454

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133–142

Held PG (2001a) Nucleic acid and protein quantitation in the microplate format. CHIMIA 55:40–42

Held PG (2001b) Quantification of DNA using Hoescht 33258. Bio-Tek Instruments, Inc., Winooski. http://biotek.com/products/tech_res_detail.php?id=39 (Cited Feb 2006)

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169–1173

Knoll JE, Gunaratna N, Ejeta G (2007) QTL analysis of early-season cold tolerance in sorghum. Theor Appl Genet (in review)

Kong L, Dong J, Hart G (2000) Characteristics, linkage-map positions, and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

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

Maiti RK, Raju PS, Bidinger FR (1981) Evaluation of visual scoring for seedling vigor in sorghum. Seed Sci Technol 9:613–622

Mansouri H, Chang G-H (1995) A comparative study of some rank tests for interaction. Comput Stat Data An 19:85–96

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

Micic Z, Hahn V, Bauer E, Melchinger AE, Knapp SJ, Tang S, Schön CC (2005) Identification and validation of QTL for *Sclerotinia* midstalk rot resistance in sunflower by selective genotyping. Theor Appl Genet 111:233–242

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73:350–356

SAS Institute (2003) SAS for windows, Version 9.1. SAS Inst. Inc., Cary

Singh SP (1985) Sources of cold tolerance in grain sorghum. Can J Plant Sci 65:251–257

Stewart CN, Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14(5):748–749

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato—a re-examination. Theor Appl Genet 75:811–823

Yousef GG, Juvik JA (2002) Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. Crop Sci 42:96–104

Zeng Z (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng Z (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zhou W-C, Kolb FL, Bai G-H, Domier LL, Boze LK, Smith NJ (2003) Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. Plant Breeding 122:40–46