Euphytica (2021) 217:13
https://doi.org/10.1007/s10681-020-02734-3

Detection of quantitative trait loci regulating seed yield potential in two interspecific *S. bicolor*^{2} × *S. halepense* subpopulations

Pheonah Nabukalu · Wenqian Kong · Thomas Stan Cox · Andrew H. Paterson

Received: 22 August 2020 / Accepted: 19 November 2020 / Published online: 3 January 2021

Abstract Perennial sorghum cropping offers substantial economic and ecological benefits, conserving fuel, water, and soil. To be perennial in temperate climates, a sorghum plant must over-winter and produce new growth the following spring—a trait derived from the weedy species *Sorghum halepense*. We have introduced perenniality from *S. halepense* into a *S. bicolor* background and identified QTL affecting eight seed yield-related traits and their linkage relationships. Interval mapping in this BC_1F_2 population derived from *S. bicolor* × *S. halepense* revealed a total of 80 QTL with LOD scores greater than 2.5 for the eight traits, with a range of 1 to 13 QTL per trait. Additional QTL were detected in multiple-QTL analyses. The traits mapped in this study showed diverse genetic complexity; the pattern of one major plus several minor QTL was observed for most traits, and traits varied in the number of QTL and direction of allelic effects. For four traits evaluated across locations, some QTL detected in one of the two locations had virtually no effect in the other, suggesting an environmental influence on QTL expression. The results contribute to fundamental knowledge of the genetic architecture underlying seed yield and may support development of high yielding perennial grain sorghum varieties.

Keywords QTL mapping · Rhizomes · Perennial · Wild relatives · Genotype by environment interaction

Introduction

Perennial grain crops offer ecological and economic value beyond grain production (Wagoner 1990; Cox et al. 2002, 2006; DeHaan et al. 2005; Glover et al. 2010). Perennial crops can produce more aboveground biomass and extensive root systems than their annual counterparts, making them more competitive against weeds and more effective at capturing nutrients and absorbing water. Further, perennial grain crops can be used to reduce soil erosion (Pimentel et al. 1987; Gantzer et al. 1990); minimize nutrient leaching (Randall and Mulla 2001; Dinnes et al. 2002); sequester soil carbon; and provide continuous habitat for wildlife (Entz et al. 2002).

Many beneficial traits, especially those for which little genetic variability currently exists within cultivated species, might be transferred into crops through hybridization with related wild species (Zhang et al. 2017). The genetic diversity of grain sorghum

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10681-020-02734-3) contains supplementary material, which is available to authorized users.

P. Nabukalu (✉) · T. S. Cox
The Land Institute, 2440 E Water Well Road, Salina, KS 67401, USA
e-mail: nabukalu@landinstitute.org

W. Kong · A. H. Paterson
Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA

[*Sorghum bicolor* (L.) Moench] has been augmented in recent years through introgression of genes controlling rhizome development from its wild weedy perennial relative, *Sorghum halepense* (L.) Pers. (Nabukalu and Cox 2016). *S. halepense* is both a weed and an invasive species; its rhizomes facilitate clonal propagation, stress protection, and survival through cold winters (McWhorter 1961; Holm 1991). Production of rhizomes is the sole means by which perennial sorghum plants produced by *S. bicolor* × *S. halepense* hybridization can survive multiple cropping seasons in temperate climates with cold winters (Piper and Kulakow 1994; Cox et al. 2002, 2006; Glover et al. 2010).

Perenniality in temperate regions with cold winters requires rhizome formation, but on a much more limited scale than that of *S. halepense*. The rhizomes of interspecific perennial sorghum germplasm are attenuated relative to those of *S. halepense* but still permit winter survival and perenniality. Numerous chromosomal regions affecting rhizome development and perennial growth habit have been mapped in *Sorghum* interspecific crosses; clearly, perenniality and rhizome development are highly complex traits, affected by genetic factors on almost all chromosomes (Paterson et al. 1995).

Through interspecific hybridization between *S. bicolor* and *S. halepense* and backcrossing to *S. bicolor*, we have developed perennial, grain producing sorghum populations (Nabukalu and Cox 2016). Although natural gene flow between the two species can occur where *S. halepense* grows in or near grain or forage sorghum fields (Arriola and Ellstrand 1996), obtaining large numbers of such hybrids through hand pollination of selected plants is laborious. This is mainly because *S. bicolor* is diploid (2n = 2x = 20) while *S. halepense* is tetraploid (2n = 4x = 40). Overcoming this barrier requires the use of either synthetically induced tetraploid *S. bicolor* plants as one parent (Piper and Kulakow 1994), or fertilizing male-sterile diploid *S. bicolor* plants with pollen from perennial tetraploid plants and obtaining 40-chromosome hybrids through the production of unreduced 20-chromosome gametes by the male-sterile *S. bicolor* parent (Hadley 1958). F_1 hybrids are both viable and fertile, often exhibiting heterosis, and through backcrossing to parental species, can produce later generation hybrids that facilitate the introgression of alleles between species. Consequently, to date, perennial grain sorghum germplasm has all been tetraploid. We have recently demonstrated that a high frequency of diploid hybrids can be produced through hybridization between diploid *S. bicolor* and tetraploid perennial sorghum (Cox et al. 2018), but we have not yet fully evaluated the resulting diploid progenies for perenniality.

S. halepense has many traits that are highly undesirable in grain production, including small, hulled seeds, thin culms, and excessive height and tillering – but also could be a source of valuable genes beyond those for rhizome development. For example, interspecific perennial sorghum germplasm has a very high biomass yield (Habyarimana et al. 2018), a trait of interest for forage and biofuel production. Elucidating the inheritance of quantitative traits in an interspecific hybrid with such variability in genome structure will boost our knowledge of the genetic architecture underlying grain-related traits and linkage relationships among loci. The objective of this study was to dissect the genetic architecture of the observed phenotypes, including QTL numbers, effects, and interaction with the environment, epistatic interactions, and linkage relationships among loci. Here, in a forward genetic approach, two interspecific subpopulations derived from *S. bicolor*^{2} × *S. halepense* were used for QTL mapping of eight seed (grain) yield-related traits. The implications of our results for perennial grain sorghum breeding are assessed.

Materials and methods

Parents and population development

Development of a *S. bicolor*^{2} × *S. halepense* backcross population was initiated in March 2011 in the greenhouse at The Land Institute in Saline County, Kansas, USA when plants of the colchicine-induced tetraploid (40-chromosome) *S. bicolor* inbred line BTx623 were partially emasculated by the plastic-bag method (Schertz and Clark 1967; Rooney 2004) and pollinated using one of six plants from the *S. halepense* 'Gypsum 9' population as the pollen source. Ten F_1 plants were produced. Two F_1 plants having the same male parent, Gypsum 9E, were designated H4 and H6 and propagated clonally until July 2012, when they were used to pollinate numerous plants of an induced tetraploid version of BTx623, herein referred

to as BTx623(4X), provided by Dr. Wayne Hanna (USDA-ARS and University of Georgia). Resulting BC_1F_1 plants were grown out and self-pollinated in the greenhouse in 2012-13 to produce 246 $BC_1F_{1:2}$ families, 141 from H4, and 105 from H6. These sets of lines will be referred to as the H4 and H6 subpopulations, respectively.

Field experimental design

Field evaluation of the 246 BC_1F_2 families, along with BTx623, BTx623(4X), and selfed progeny of Gypsum 9E, was carried out in two locations: The Land Institute's research farm on the south edge of Salina, KS (hereinafter, "Salina") and the University of Georgia Plant Science Farm, Watkinsville, south of Athens, GA (hereinafter, "Athens") in 2013 and 2014. The Salina site, at 38.84 N, 97.61 W, elevation 373 m, represented a typical temperate environment with cold winters. The Athens site, at 33.87 N, 83.54 W, elevation 254 m, features milder winters than Salina, but air temperatures drop well below freezing often during the winter. The Salina site, has montmorillonitic, mesic pactic Argiustoll soils (Soil Survey of Saline County, Kansas, 1992), while the Athens site has a Cecil clay loam soil (fine, kaolinitic, thermic Typic Kanhapludults).

The experiment was set up in a randomized complete block design having two replicates. The two subpopulations were randomized together along with parents in each experiment; however, because their parent plant Gypsum 9E was heterozygous at an unknown proportion of its loci, the two subpopulations were known to be derived from genetically distinct F_1 hybrids and were therefore analyzed separately. In each replicate, each family was represented by a single-row plot of 3 m containing approximately 10 plants. Rows were spaced 0.91 m apart. Three random plants from the interior of each row were marked for phenotypic evaluation.

Phenotyping

In evaluating traits, we followed published procedures (ICRISAT 1993). At harvest, panicles were separated from the rest of the plant, dried for 10 to 14 days in an electric dryer to reach a constant moisture percentage of $\sim$ 10%, and threshed with a mechanical thresher. The remaining vegetative portions of the three harvested plants in each row were cut at 15 cm above the soil surface and air-dried to constant weight to measure aboveground dry weight (g) of stalks and leaves.

Seed yield per plant was recorded as the mean weight of seed produced for the three marked plants in each row. Seed yield per panicle was the mean weight of seed harvested from the primary panicle of each marked plant in the row. 1000-seed weight was recorded as the weight of one thousand well developed whole seeds from bulk seeds obtained from each of the three marked plants in each row. Glume tenacity was visually estimated as the proportion of seeds covered by glumes after threshing. Panicle length was the distance between the base and the tip of the primary panicle. Panicle width was the average width of the primary panicle at its widest point. Panicle compactness was based on spikelet density and visually scored on a scale of 1 (most open) to 5 (most compact: Supplementary Fig. 1). Harvest index was calculated as the ratio of seed yield per plant to the weight of air-dry vegetative biomass per plant.

Seed yield per panicle, 1000-seed weight, glume tenacity, and panicle length were evaluated in Athens and Salina. Seed yield per plant, harvest index, and panicle compactness were evaluated only in Salina, while panicle width was evaluated only in Athens.

Statistical and QTL analysis

Unless otherwise indicated, all statistical analyses were done with R (version 3.4.2, R Core 2017). For each trait, we computed means and least-squares means (lsmeans) for each family in the two BC_1F_2 subpopulations, and for each of the parents BTx623, BTx623(4X), and Gypsum 9E. Pairwise contrasts between 'parents' and 'subpopulations' were carried out using the *LS means* package (Lenth 2016). In addition, descriptive statistics were calculated, and trait frequency distributions were plotted. To improve normality, data were arcsine-transformed for glume tenacity and log10-transformed for seed yield per plant, seed yield per panicle, panicle width, and harvest index.

Statistical analysis was performed using the lme4 package for mixed model analysis (Bates et al. 2015). For four traits that were evaluated in both locations—seed yield per panicle, 1000-seed weight, glume

tenacity, and panicle length—a linear mixed model was fitted

$$P_{ijk} = \mu + G_i + L_j + Y_k + B_{(ljk)} + G_i * E_i + G_i * Y_k + e_{ijkl},$$

where P_{ijk} is the phenotypic value, μ the population mean, G_i the effect of the ith genotype, L_j the effect of the jth location, Y_k the effect of the kth year, $B_{(ljk)}$ the effect of the lth replicate within the jth location and kth year, G_i*L_j the ijth effect of the genotype-by-location interaction, G_i*Y_k the ikth effect of the genotype-by-year interaction and e_{ijkl} designated the residual. The location effect was treated as fixed and all other terms as random effects.

For single location analyses—seed yield per plant, harvest index, panicle compactness, and panicle width, a reduced linear mixed model was fitted

$$P_{ik} = \mu + G_i + Y_k + B_{(lk)} + G_i * Y_k + e_{ikl},$$

where P_{ik} is the phenotypic value, μ the population mean, G_i the effect of the ith genotype, Y_k the effect of the kth year, $B_{(lk)}$ the effect of the lth replicate within the kth year, G_i*Y_k the ikth effect of the genotype-by-year interaction and e_{ikl} designated the residual. All terms were treated as random effects.

Broad sense heritability (h^2) was estimated using variance components (σ^2) calculated with multi-year and replicated data by restricted maximum likelihood (REML) method. All effects were treated as random. For traits evaluated in two environments, broad-sense heritability was estimated as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \left[\sigma_G^2 + \left(\sigma_{GXL}^2/L\right) + \left(\sigma_{GXY}^2/Y\right) + \left(\sigma_E^2/LY\right)\right]}$$

where σ^2_G is the genotypic variance, σ^2_{GxL} is the genotype × location variance, σ^2_{GxY} is the genotype × year variance, σ^2_E is the plot residual variance, and L and Y are the number of locations and years respectively. For traits evaluated only in one location, replicates were used in the heritability calculation in place of location along with the interaction between genotype and year to estimate the variance caused by genotype × environment interaction. Pearson correlations between traits were estimated based on the least-square means values over the two years using the cor () function and cor.test () function to test the significance of each correlation.

QTL analyses and detection

Genotyping and linkage map construction were reported previously (Kong et al. 2020). Briefly, genotypes were determined by single nucleotide polymorphism (SNP) 'calling' based on the reference genome of *S. bicolor* BTx623 *v1.4* (Paterson et al. 2009), using Tassel-GBS 5 (Glaubitz et al. 2014). Genetic linkage maps for the H4 and H6 subpopulations consisted of 726 and 799 SNP markers respectively, spanning 38 and 36 linkage groups distributed among all ten basic sorghum chromosomes.

QTL analysis was performed using least-squares means (lsmeans) calculated for each individual family in each location and/or across locations depending on the trait. Phenotypic and genotypic data were integrated for QTL mapping with the R/qtl package (Broman et al. 2003; Broman and Sen 2009). Initially, marker-trait associations were tested using a single marker analysis (Weller 1986). The genotypes, phenotypes, and genetic maps in the backcross format were imported with the *read.cross* function. Standard interval mapping (SIM) was performed using a single-QTL model regression method (Haley and Knott 1992). For SIM, a one-dimensional QTL genome scan was done by using the *scanone* function to determine QTL at a specified threshold and also to perform 1000 permutation tests to obtain genome-wide LOD significance thresholds (Churchill and Doerge 1994). Loci with a peak LOD score of 2.5 or above were reported as QTL, noting that genome-wide significance thresholds ($P < 0.05$) determined with 1000 permutations ranged from 3.19 to 3.48 (Supplementary Table 2).

The likelihood interval for each QTL peak was obtained via the *lodint* function, which calculates 1.5-LOD support intervals. Adjacent QTL on the same chromosome were considered different when the support intervals were not overlapping. The proportion of variance explained by QTL (R^2) was calculated as the percentage of variance explained by each QTL in proportion to the total phenotypic variance. The allele substitution effect of a putative QTL was estimated as the phenotypic difference between the heterozygous and homozygous classes. Finally, a multiple-QTL model was fitted to determine QTL for each trait using *makeqtl* and *fitqtl* functions. We developed multiple-QTL models via a penalized stepwise model selection approach (Manichaikul et al. 2009) within the software package R/qtl

(Broman et al. 2003), where terms were included at $\alpha = 0.05$. Significance was determined by 1000 permutations.

QTL were named as described (Tanksley and McCouch 1997) with some modifications. For instance, *q*PGY.H4.1F-3 corresponds to the third QTL for seed yield per plant detected on linkage group IF in the H4 subpopulation. Linkage map figures showing locations of significant QTL were constructed using Mapchart2.2 (Voorrips 2002). QTL locations were indicated with 1.5-LOD support intervals. QTL were classified as "major" if the phenotypic variance explained was larger than 10%, and "minor" when they accounted for less than 10% of the phenotypic variance (Collard et al. 2005).

Variables were scaled so that the phenotypic effect of a QTL was positive in sign if the allele from *S. bicolor* had the effect of decreasing the trait value and negative if the *S. bicolor*-derived allele increased the trait value. If the sign of the allele substitution effect was the same as the sign of the parental-mean difference Tx623(4x) – Gypsum 9E, the allelic effect was in the "expected" direction; if the signs were different, the effect was in the "opposite" direction.

QTL for different traits were declared to be co-located when their likelihood peaks were in the same marker interval. Co-location was designated as "coupling" when the allele substitution effects had the same algebraic sign and "repulsion" when they had opposite algebraic signs. A QTL was said to be constitutive or consistent when it was detected in both Athens and Salina single location analyses, while a QTL associated with a trait only at one location was designated an "adaptive" QTL (Ping et al. 2003).

QTL of target traits were compared with QTL identified in previous studies; downloaded from the Sorghum QTL Atlas on OZ Sorghum (https://aussorgm.org.au/sorghum-qtl-atlas/). The QTLs sharing similar genetic positions (1.5-LOD support interval as defined by flanking markers) were regarded as common or same QTL. When a QTL in the current study shared the same physical region as the previous QTL, it was regarded as a repeated identification of the previous QTL; otherwise, the QTL in the current study was regarded as a new one.

Results

Phenotypic variation

The complexity of seed yield reduces the power of QTL mapping; therefore, we analyzed various of its component traits: seed yield per plant, seed yield per panicle, and 1000-seed weight. We also analyzed the related traits like panicle length, panicle width, panicle compactness, and harvest index.

Analyses of variance demonstrated highly significant ($P < 0.001$) differences among genotypes for all traits (Supplementary Table 1). For seed yield per panicle, glume tenacity, panicle length, and 1000-seed weight, which were evaluated at both locations, location main effects, as well as genotype-by-location and genotype-by-year interactions, were significant (Supplementary Table 1). For seed yield per plant, panicle compactness, panicle width, and harvest index, each evaluated at a single location, genotype × year interactions were significant (Supplementary Table 1). Coefficients of genotypic variation (CVG) ranged from 6% for panicle length to 33% for harvest index (Supplementary Table 1). The estimated broad-sense heritability (h^2) values were low to moderate, ranging from 20 to 47%, with panicle width having the lowest and panicle compactness the highest heritability (Supplementary Table 1).

With the two subpopulations merged into a single population, Pearson correlations for many pairs of traits were significant ($P \leq 0.05$; Supplementary Fig. 2). As expected, correlation coefficients among seed yield per panicle, seed yield per plant, and harvest index were all positive and large, above 0.8 (Supplementary Fig. 2); harvest index was also positive and highly correlated with 1000-seed weight ($r = 0.6$, $P < 0.001$). Panicle compactness was significant and negatively correlated with panicle length ($r = -0.4$, $P < 0.001$) and width ($r = -0.3$, $P < 0.001$), whereas panicle length and width were significant and positively correlated ($r = 0.4$, $P < 0.001$).

QTL analyses and their interactions

Across all traits, standard interval mapping revealed a total of 25 and 55 significant QTL (LOD > 2.5) for the eight seed yield-related traits analyzed in the H4 and H6 subpopulations, respectively (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3).

Table 1 Quantitative trait loci (QTL) detected for seed yield related traits in H4 and H6 subpopulations derived from *S. bicolor*[2] × *S. halepense* based on single-locus interval mapping

Trait[a]	Analysis	Sub.pop	QTL name	LG[b]	Pos[c] (cM)	Interval (cM)	Flanking markers	LOD[d] score	Additive effect[e]	R^{2f}
PGY	Salina	H4	qPGY-H4-1F-1	1F	30.3	0–94.9	S1_3151759–S1_55709669	2.61	0.18	8.57
			qPGY-H4-6B-2	6B	45.6	0–74.7	S6_3622183–S6_47437944	3.9	− 0.21	12.56
		H6	qPGY-H6-2A-1	2A	69.0	56.3–83.3	S2_45313448–S2_64810952	3.28	− 0.22	13.82
PSY	Combined	H4	qPSY-H4-1F-1	1F	89.3	0–94.9	S1_3151759–S1_55709669	2.75	0.12	9.01
			qPSY-H4-6B-2	6B	51.1	0–74.7	S6_3622183–S6_47437944	2.94	− 0.14	9.60
		H6	qPSY-H6-2A-1	2A	69.0	56.3–83.3	S2_45313448–S2_64810952	5.08	− 0.21	20.55
			qPSY-H6-3D-3	3D	14.9	0–26.6	S3_16879087–S3_57435185	2.63	− 0.12	11.19
			qPSY-H6-3E-4	3E	125.9	49.8–146.8	S3_61484890–S3_10011376	2.89	0.13	12.21
			qPSY-H6-4A-5	4A	66.6	0–148.7	S4_214220–S4_65720396	2.63	0.14	11.18
			qPSY-H6-4D-6	4D	94.5	77–171.6	S4_16899809–S4_61117008	3.41	0.16	14.27
			qPSY-H6-6E-8	6E	0.0	0–25.5	S6_48323502–S6_58949495	2.66	0.15	11.32
			qPSY-H6-9B-9	9B	51.4	0-95.5	S9_55445106–S9_53579966	2.68	− 0.13	11.41
			qPSY-H6-10C-10	10C	50.0	39.3–161.6	S10_58313642–S10_59831354	2.75	0.17	11.66
	Athens	H4	qPSY-H4-6B-2	6B	8.2	0-55.2	S6_3622183–S6_45264939	3.83	− 0.17	12.32
		H6	qPSY-H6-2A-1	2A	69.0	56.3–83.3	S2_45313448–S2_64810952	4.06	− 0.23	16.76
			qPSY-H6-3C-2	3C	16.3	0–34.3	S3_11787974–S3_50452733	2.95	0.15	12.46
			qPSY-H6-3E-4	3E	125.9	49.8–175.3	S3_61484890–S3_64222330	2.73	0.16	11.61
			qPSY-H6-4A-5	4A	52.8	0–85.3	S4_214220–S4_4271211	3.01	0.17	12.72
			qPSY-H6-4D-6	4D	94.5	77–119.6	S4_16899809–S4_53405669	3.33	0.19	13.96
			qPSY-H6-5C-7	5C	72.0	55.7–94.7	S5_1520736–S5_55213444	2.81	0.16	11.90
	Salina	H4								
		H6	qPSY-H6-2A-1	2A	63.6	56.3–83.3	S2_45313448–S2_64810952	3.09	− 0.16	12.66
			qPSY-H6-4D-6	4D	111.9	77–285.5	S4_16899809–S4_63383933	2.64	0.15	10.94
TKWT	Combined	H4	qTKWT-H4-2D-1	2D	65.2	41.4–164.2	S2_18922672–S2_73158707	2.62	− 1.50	8.63
			qTKWT-H4-3B-2	3B	68.2	57.3–156.3	S3_11787974–S3_74423186	2.93	− 1.59	9.58
			qTKWT-H4-3C-3	3C	140.4	136.9–166.8	S3_60702622–S3_64894108	2.56	− 1.44	8.43
			qTKWT-H4-3D-4	3D	69.9	47.5–109.5	S3_69475087–S3_55110379	3.03	− 1.62	9.88
			qTKWT-H4-5B-5	5B	37.8	11.6–37.8	S5_17202110–S5_45709039	2.61	− 1.48	8.57
			qTKWT-H4-6B-7	6B	13.8	0.0–18.2	S6_3622183–S6_37245793	9.81	− 2.72	28.63
			qTKWT-H4-6C-8	6C	10.6	0.0–26.2	S6_42647714–S6_58127697	3.85	− 1.79	12.40
		H6	qTKWT-H6-1D-1	1D	100	80.6–114.1	S1_19182813–S1_54649751	5.01	− 2.42	20.05
			qTKWT-H6-2A-2	2A	69	56.3–83.3	S2_45313448–S2_64810952	3.06	− 1.76	12.93

Table 1 continued

Trait[a]	Analysis	Sub.pop	QTL name	LG[b]	Pos[c] (cM)	Interval (cM)	Flanking markers	LOD[d] score	Additive effect[e]	R^{2f}
			qTKWT-H6-6E-6	6E	153	120.2–166.9	S6_49815215–S6_48741391	5.49	− 1.92	21.97
	Athens	H4	qTKWT-H4-3C-3	3C	140.4	136.9–166.8	S3_60702622–S3_64894108	2.52	− 1.49	8.31
			qTKWT-H4-6B-7	6B	13.8	0.0–55.2	S6_3622183–S6_45264939	6.71	− 2.40	20.58
			qTKWT-H4-6C-8	6C	10.6	0.0–47.0	S6_42647714–S6_48867332	2.58	− 1.54	8.50
		H6	qTKWT-H6-1D-1	1D	100.4	80.6–114.1	S1_19182813–S1_54649751	3.98	− 2.61	16.46
			qTKWT-H6-2A-2	2A	69	56.3–83.3	S2_45313448–S2_64810952	2.68	− 1.99	11.39
			qTKWT-H6-4D-5	4D	94.5	39.5–327.7	S4_8741878–S4_60689590	2.66	1.83	11.31
			qTKWT-H6-6E-6	6E	152.9	120.2–166.9	S6_49815215–S6_48741391	3.45	− 1.87	14.44
			qTKWT-H6-9C-8	9C	226.7	29.8–241.9	S9_3592752–S9_53833293	2.88	− 2.14	12.19
			qTKWT-H6-10C-9	10C	69.1	44.5–83.9	S10_58643663–S10_8524545	3.35	2.16	14.05
	Salina	H4	qTKWT-H4-3B-2	3B	68.18	57.3–141.6	S3_11787974–S3_72617956	3.09	− 1.83	10.06
			qTKWT-H4-3D-4	3D	69.91	50.2–75.5	S3_62695814–S3_71175405	3.48	− 1.94	11.28
			qTKWT-H4-5B-5	5B	37.83	11.6–37.8	S5_17202110–S5_45709039	2.61	− 1.67	8.59
			qTKWT-H4-6A-6	6A	0	0.0–89.2	S6_50051710–S6_54653430	2.81	− 1.94	9.19
			qTKWT-H4-6B-7	6B	8.17	0.0–18.2	S6_3622183–S6_37245793	9.81	− 3.04	28.61
			qTKWT-H4-6C-8	6C	10.56	0.0–26.2	S6_42647714–S6_58127697	3.74	− 1.98	12.07
		H6	qTKWT-H6-1D-1	1D	100	80.6–114.1	S1_19182813–S1_54649751	3.62	− 2.23	15.02
			qTKWT-H6-3B-3	3B	108	26.6–112.4	S3_16591808–S3_65798323	3.26	− 1.71	13.69
			qTKWT-H6-4C-4	4C	124	54.9–138.7	S4_11348327–S4_60791173	2.89	1.59	12.27
			qTKWT-H6-6E-6	6E	153	120.2–166.9	S6_49815215–S6_48741391	5.41	− 2.03	21.71
			qTKWT-H6-7D-7	7D	128	40.0–162.3	S7_9984892–S7_11726102	2.72	− 1.53	11.58
GT	Combined	H4	qGT-H4-1A-1	1A	128.0	35.4–175.8	S1_2075903–S1_58063128	2.92	0.11	9.54
			qGT-H4-1F-2	1F	27.7	0–55.7	S1_1646069–S1_6979584	2.74	0.09	8.97
			qGT-H4-2C-4	2C	51.5	47.9–72.3	S2_73516056–S2_75611690	2.84	0.08	9.30
			qGT-H4-7C-6	7C	0.0	0–6.9	S7_5180810–S7_6735921	3.38	0.09	10.97
		H6	qGT-H6-4D-3	4D	274.0	262.8–300.4	S4_62176230–S4_58764349	3.73	− 0.11	15.50
			qGT-H6-6E-5	6E	186.0	166.9–207.8	S6_48741391–S6_58868853	3.69	− 0.12	15.39
			qGT-H6-7D-6	7D	0.0	0–144.4	S7_18378976–S7_37909272	3.04	0.10	12.81
			qGT-H6-8A-7	8A	269.0	247.4–269.2	S8_44723812–S8_1808533	3.26	0.10	13.70
			qGT-H6-8B-8	8B	85.7	54–131.8	S8_5265160–S8_2986253	3.29	− 0.11	13.80
	Athens	H4	qGT-H4-2A-3	2A	58.2	51.7–63.4	S2_8302422–S2_13837199	2.64	0.11	8.73
			qGT-H4-6B-5	6B	57.3	45.6–74.7	S6_40800373–S6_47437944	2.5	0.11	8.29
		H6	qGT-H6-1C-1	1C	137.2	115.9–151.9	S1_68787848–S1_71968245	2.54	− 0.09	10.93

Table 1 continued

Trait[a]	Analysis	Sub.pop	QTL name	LG[b]	Pos[c] (cM)	Interval (cM)	Flanking markers	LOD[d] score	Additive effect[e]	R^{2f}
			qGT-H6-2B-2	2B	7.2	0–93.6	S2_59828355–S2_69646013	3.37	− 0.11	14.25
			qGT-H6-4D-3	4D	285.5	262.8–300.4	S4_62176230–S4_58764349	3.27	− 0.12	13.86
			qGT-H6-6B-4	6B	70.7	60.6–104.9	S6_37245941–S6_47011071	2.88	− 0.10	12.30
			qGT-H6-6E-5	6E	185.9	166.9–207.8	S6_48741391–S6_58868853	3.12	− 0.12	13.25
			qGT-H6-7D-6	7D	11.3	0–144.4	S7_18378976–S7_37909272	2.61	0.11	11.20
			qGT-H6-8A-7	8A	269.3	247.4–269.2	S8_44723812–S8_1808533	3.29	0.12	13.91
			qGT-H6-8B-8	8B	85.7	0–131.8	S8_4800118–S8_2986253	3.92	− 0.14	16.38
	Salina	H4	qGT-H4-1A-1	1A	128.0	35.4–175.8	S1_2075903–S1_58063128	2.95	0.11	9.63
			qGT-H4-1F-2	1F	27.7	0–55.7	S1_1646069–S1_6979584	2.74	0.09	8.99
			qGT-H4-2C-3	2C	51.5	47.9–54.7	S2_73516056–S2_77730934	2.9	0.08	9.48
			qGT-H4-7C-6	7C	0.0	0–6.9	S7_5180810–S7_6735921	3.61	0.09	11.66
		H6	qGT-H6-4D-3	4D	274.0	262.8–300.4	S4_62176230–S4_58764349	3.53	− 0.11	14.74
			qGT-H6-6E-5	6E	186.0	166.9–207.8	S6_48741391–S6_58868853	3.49	− 0.12	14.65
			qGT-H6-7D-6	7D	0.0	0–144.4	S7_18378976–S7_37909272	2.88	0.10	12.21
			qGT-H6-8A-7	8A	269.0	247.4-269.2	S8_44723812–S8_1808533	3.46	0.11	14.48
			qGT-H6-8B-8	8B	85.7	54–131.8	S8_5265160–S8_2986253	3.35	− 0.11	14.04
PL	Combined	H4	qPL-H4-1A-1	1A	16.2	11.1–117.7	S1_81578–S1_12736424	3.33	1.87	10.81
			qPL-H4-1C-2	1C	52.9	2.8–62.9	S1_6958375–S1_7356002	3.88	1.81	12.47
		H6	qPL-H6-1C-2	1C	140.8	131.4–163.3	S1_66947494–S1_71777631	2.82	1.90	11.96
			qPL-H6-3E-3	3E	98.3	0–225	S3_3508310–S3_2766900	2.65	2.08	11.28
			qPL-H6-4B-4	4B	0.0	0–240.3	S4_55529640–S4_62960068	2.68	− 1.92	11.39
			qPL-H6-4D-5	4D	111.9	94.5–119.6	S4_12861174–S4_53405669	3.7	2.82	15.37
			qPL-H6-6D-6	6D	0.0	0–15.4	S6_59364592–S6_59274305	2.69	1.94	11.43
			qPL-H6-7D-8	7D	11.3	0–192.4	S7_18378976–S7_11975519	3.54	− 2.53	14.77
			qPL-H6-10C-10	10C	50.0	44.5–83.9	S10_58643663–S10_8524545	4.07	3.11	16.79
PL	Athens	H4	qPL-H4-1A-1	1A	112.9	13.6–117.7	S1_1481073–S1_12736424	2.6	2.11	8.54
			qPL-H4-1C-2	1C	52.9	0–74.8	S1_6744852–S1_13139592	2.75	1.82	9.03
		H6	qPL-H6-4B-4	4B	228.0	0–240.3	S4_55529640–S4_62960068	2.72	− 2.26	11.54
			qPL-H6-4D-5	4D	112.0	94.5–119.6	S4_12861174–S4_53405669	3.87	3.08	15.92
			qPL-H6-10C-10	10C	50.0	39.3–83.9	S10_58313642–S10_8524545	3.68	3.16	15.32
	Salina	H4	qPL-H4-1A-1	1A	20.0	0–117.7	S1_1090512–S1_12736424	2.76	1.83	8.68
			qPL-H4-1C-2	1C	21.7	2.8–61	S1_6958375–S1_5681094	3.68	1.80	11.39
		H6	qPL-H6-1B-1	1B	69.7	16.7–76.7	S1_58839105–S1_61912352	3.4	2.59	13.85

Table 1 continued

Trait[a]	Analysis	Sub.pop	QTL name	LG[b]	Pos[c] (cM)	Interval (cM)	Flanking markers	LOD[d] score	Additive effect[e]	R^{2f}
			qPL-H6-1C-2	1C	140.8	115.9–163.3	S1_68787848–S1_71777631	2.88	2.23	11.88
			qPL-H6-4D-5	4D	107.4	9.4-285.5	S4_3167797–S4_63383933	2.71	2.78	11.22
			qPL-H6-6E-7	6E	14.6	0–218	S6_48323502–S6_58729665	2.83	2.63	11.69
			qPL-H6-7D-8	7D	0.0	0–192.4	S7_18378976–S7_11975519	4.11	− 2.90	16.50
			qPL-H6-8B-9	8B	246.5	153.4–299	S8_47509071–S8_47878656	3.04	2.36	12.49
			qPL-H6-10C-10	10C	59.4	19.4–83.9	S10_2406347–S10_8524545	2.83	2.43	11.67
PW	Athens	H4	qPW-H4-4D-1	4D	12.9	0–39.2	S4_1398138–S4_65814835	2.51	-0.06	8.28
		H6	qPW-H6-2C-1	2C	0.0	0–35.4	S2_77242837–S2_68670090	2.67	0.07	11.35
			qPW-H6-3E-2	3E	23.1	0–101	S3_3508310–S3_71280511	2.54	0.10	10.84
			qPW-H6-5C-3	5C	17.0	0–135.6	S5_2552542–S5_62185637	2.51	0.08	10.71
PC	Salina	H4	qPC-H4-1A-1	1A	175.8	86.5–187.1	S1_8669629–S1_60795489	2.89	− 0.45	9.46
			qPC-H4-1E-2	1E	10.8	0–10.8	S1_21379189–S1_28000818	2.64	− 0.30	8.67
			qPC-H4-6A-3	6A	57.3	7.8–64.5	S6_49815215–S6_52584907	5.05	− 0.49	15.93
		H6	qPC-H6-1A-1	1A	104.8	92.6–112.8	S1_4336161–S1_2075903	4.03	− 0.43	16.64
			qPC-H6-1D-2	1D	144.5	136.6–166.6	S1_4464073–S1_51100819	3.37	0.41	14.11
			qPC-H6-2B-3	2B	93.6	0–93.6	S2_59828355–S2_69646013	3.77	− 0.37	15.67
			qPC-H6-3A-4	3A	156.5	137.2–171.8	S3_52466330–S3_66256314	2.89	− 0.44	12.23
			qPC-H6-3B-5	3B	83.8	26.6–92.9	S3_16591808–S3_60878509	3.37	− 0.41	14.10
			qPC-H6-4B-6	4B	240.3	228–259.6	S4_61407366–S4_64876657	2.97	0.34	12.56
			qPC-H6-4C-7	4C	111.4	13–123.7	S4_59931550–S4_65038519	2.52	0.31	10.77
			qPC-H6-4D-8	4D	136.4	119.6–217.8	S4_53405669–S4_66856106	4.07	0.41	16.79
			qPC-H6-5A-9	5A	41.7	31.8–59.6	S5_60011353–S5_61417202	2.59	− 0.42	11.05
			qPC-H6-6A-10	6A	94.0	68.5–180.2	S6_54242767–S6_57499373	3.06	− 0.36	12.90
			qPC-H6-6D-11	6D	26.6	0–26.6	S6_59364592–S6_59261261	2.61	− 0.30	11.12
			qPC-H6-9A-12	9A	168.8	156.4–195.1	S9_49472992–S9_52885753	3.88	− 0.45	16.06
			qPC-H6-9C-13	9C	146.5	143.7-153.1	S9_55848834–S9_9922630	3.79	0.45	15.71

Table 1 continued

Trait[a]	Analysis	Sub.pop	QTL name	LG[b]	Pos[c] (cM)	Interval (cM)	Flanking markers	LOD[d] score	Additive effect[e]	R^{2f}
HI	Salina	H4	qHI-H4-6B-1	6B	43.2	39.4–51.1	S6_1970777–S6_43160977	9.7	− 0.37	28.36
		H6	qHI-H6-4A-1	4A	66.6	60.3–76.5	S4_4321566–S4_5066063	5.25	0.25	21.09

[a]Trait abbreviations: Seed yield per plant (PGY), Seed yield per panicle (PSY), 1000-seed weight (TKWT), Glume tenacity (GT), Panicle length (PL), Panicle compactness (PC), Panicle width (PW), Harvest index (HI)

[b]LG is linkage group

[c]Position is expressed in centimorgans

[d]LOD is logarithm of odds

[e]Allele substitution effect = $(\bar{x}_{AB} - \bar{x}_{BB})$; given as (*S. halepense* – *S.bicolor*) values: positive values indicate that higher value alleles are from *S. halepense* and negative values indicate that higher value alleles are from *S. bicolor*

[f]R^2 % is percentage of phenotypic variation explained by individual QTL

The number of QTL detected per trait ranged from 1 to 8 and 1 to 13 in H4 and H6, respectively (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3). The largest number of QTL were located on chromosomes 1 and 6 in H4 and on chromosomes 1, 2, 4, 6, 7, and 9 in H6 (Fig. 1 and Supplementary Fig. 3). While QTL affecting the eight traits were distributed across all ten chromosomes in H6, chromosomes 8, 9, and 10 were conspicuous in the H4 subpopulation for the absence of detected QTL for any of the eight traits (Fig. 1 and Supplementary Fig. 3). Additional QTL were detected in the multiple-QTL analyses presented in Table 2 (with details in Supplementary Table 3).

Seed yield per plant (PGY)

Two significant QTL were associated with seed yield per plant in the H4 subpopulation, on linkage groups 1F and 6B (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3a). Their contributions to the total phenotypic variation were 8.6 and 12.6%, respectively (Table v; Supplementary Table 2). At the locus on 1F, the *S. halepense* allele increased seed yield per plant (Table 1; Supplementary Table 2); the effect was in the expected direction because the *S. halepense* parent Gypsum 9E had higher seed yield per plant than BTx623(4X) (Supplementary Table 4, Supplementary Fig. 4). At the 6B locus, the *S. bicolor* allele increased seed yield per plant, opposite of the difference between parents (Table 1; Supplementary Table 2). In the multiple-QTL analysis, three significant QTL, mapping to linkage groups 1F, 6B (67.0), and 6B (81.0), were associated with seed yield per plant in H4 (Supplementary Table 3). The QTL on 1F and 6B (67.0) mapped in the same marker intervals as loci detected in the single-QTL analysis. The three additively interacting QTL jointly accounted for 33.5% of the total phenotypic variation (Supplementary Table 3).

In H6, only one major QTL was associated with seed yield per plant (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b). Mapping to linkage group 2A, this QTL accounted for 13.8% of the total phenotypic variance; the *S. bicolor* allele increased the trait value (Table 1; Supplementary Table 2), even though the *S. halepense* parent had the higher value. A QTL with an overlapping support interval was detected in multiple-QTL analysis,

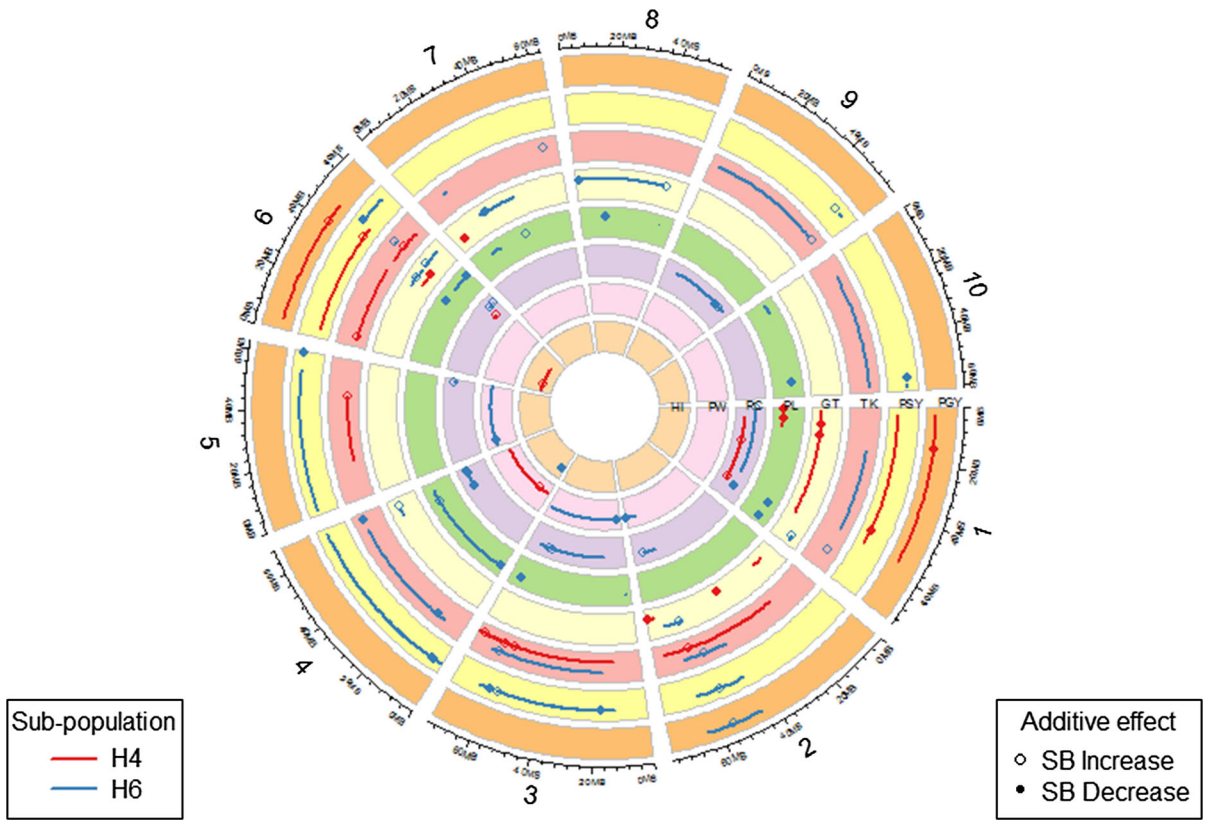

Fig. 1 Genome-wide distribution of QTL for eight traits: Seed yield per plant (PGY), Seed yield per panicle (PSY), 1000-seed weight (TK), Glume tenacity (GT), Panicle length (PL), Panicle compactness (PC), Panicle width (PW), Harvest index (HI) identified in the H4 and H6 subpopulations derived from *S. bicolor*^{2} × *S. halepense*

contributing 14.8% to the total phenotypic variance (Supplementary Table 3).

Seed yield per panicle (PSY)

Two significant QTL were associated with seed yield per panicle in the H4 subpopulation (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3a). The two loci—located on linkage groups 1F and 6B —were identified using multi-location data; only one QTL, mapping to linkage group 6B, was detected in Athens; and no significant QTL were detected in Salina. At the QTL mapped to linkage group 1F, the *S. halepense* allele exerted a positive allele substitution effect (Table 1; Supplementary Table 2; Fig. 1), which was opposite to expectations, because the *S. halepense* parent had the lower seed yield per panicle (Supplementary Table 4, Supplementary Fig. 4). From the multiple-QTL analysis across locations (Supplementary Table 3), two additive QTL were associated with seed yield per panicle, and they mapped to linkage groups 1F and 6B. The three peaks detected on 1F (19.0 cM, 87.0 cM, 88.0 cM) lie in the same support interval as the QTL detected in the single-QTL analysis and are likely to represent one QTL. All QTL additively had a joint LOD score of 13.4 and accounted for 36.9% of the total phenotypic variation (Supplementary Table 3). In Athens, there were significant interactions among QTL on linkage groups; 1F, 5C, 6B, and 10B (Supplementary Table 3). These interacting QTL had a joint LOD score of 15.6 and accounted for 41.5% of the total phenotypic variation (Supplementary Table 3).

Ten significant QTL were associated with seed yield per panicle in H6 (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b). QTL mapping to linkage groups 2A, 3D, 3E, 4A, 4D, 6E, 9B, and 10C were identified in the combined analysis across locations; other QTL mapping to linkage

groups 3C and 5C were identified only in Athens, and two QTL mapping on linkage group 2A and 4D were detected in all three analyses. LOD scores ranged from 2.6 to 5.1, and the percentage of variance explained by a single QTL ranged from 11.2 to 20.6%. For five QTL mapping to linkage groups 1C, 2A, 3D, 7B, and 9B, the *S bicolor* allele increased the trait value; therefore, the allele effect was in the expected direction (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b). In the multiple-QTL analysis across locations (Supplementary Table 3), there was an interaction between QTL mapping to linkage groups 2A and 8B. The QTL on linkage group 2A had been identified in the single-QTL analysis, while QTL on linkage group 8B had not. The two QTL had a combined LOD score of 9.4 and accounted for 34.7% of the total phenotypic variation (Supplementary Table 3). In the single-location analyses, a significant interaction was detected between QTL mapping to 2A and 4A in Athens; both mapped close to genomic regions detected in the single-locus analyses. The two QTL had a combined LOD score of 6.9 and accounted for 26.6% of the total phenotypic variation (Supplementary Table 3).

1000-seed weight (TKWT)

Eight significant QTL were associated with 1000-seed weight in the H4 subpopulation, mapping to linkage groups 2D, 3B, 3C, 3D, 5B, 6A, 6B and 6C (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3a). QTL on linkage group 6B and 6C were detected in all analyses within and across locations (Table 1; Supplementary Table 2). Individual QTL had LOD scores ranging from 2.5 to 9.8, and percentages of variance were between 8.3 and 28.6% (Table 1; Supplementary Table 2). For all eight QTL, the allele from the larger-seeded *S. bicolor* parent increased the trait value (Table 1; Supplementary Table 2). In the multiple-QTL analysis across locations, three interacting QTL were in linkage groups 4C (141.0 cM), 4C (155.0 cM), and 6B (11.0 cM), with a joint LOD score of 17.2 and percentage of the phenotypic variance of 44.5% (Supplementary Table 3). The QTL on 6B had been detected in the single-locus analysis, but the two QTL on 4C had not. In Salina, there were interactions among QTL on linkage groups 3B, 4C and 6B (Supplementary Table 3), with a joint LOD score of 16.6 and percentage of the phenotypic variance of 43.5% (Supplementary Table 3).

Nine QTL, mapping to linkage groups 1D, 2A, 3B, 4C, 4D, 6E, 7D, 9C, and 10C, were associated with 1000-seed weight in H6 (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b). Those mapping to 1D and 6E were detected in all analyses. LOD scores ranged from 2.7 to 5.1, and the percentage of variance explained by a single QTL ranged from 11.3 to 21.7% (Table 1; Supplementary Table 2). For QTL mapping to 1D, 2A, 3B, 6E, 7D, and 9C, the *S. bicolor* allele increased 1000-seed weight as expected (Table 1; Supplementary Table 2). From the multiple-QTL analysis across locations, four interacting QTL were associated with 1000-seed weight in the H6 subpopulation, on linkage groups 1D, 2A, 3D, and 5A (Supplementary Table 3). Only the QTL on 1D had been detected in the single-locus analysis. The interaction accounted for 50.9% of the phenotypic variance, with a joint LOD score of 15.7. A second interaction occurred between QTL on 1D and 5A.

Glume tenacity (GT)

Six QTL were associated with glume tenacity in the H4 subpopulation (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3a). Four QTL were mapped to linkage groups 1A, 1F, 2C and 7C in across-location analysis, while two others were mapped to 2A and 6B in the Athens analysis. LOD scores for individual QTL ranged from 2.5 to 3.6, and the percentage of variance explained varied from 8.3 to 11.7% (Table 1; Supplementary Table 2). The *S. halepense* allele, as expected, increased glume tenacity at all loci detected (Table 1; Supplementary Table 2). In the multiple-QTL analysis across locations (Supplementary Table 3), three additive QTL mapping to linkage groups 1A, 2C, and 7C were associated with glume tenacity in H4, with a combined LOD score of 12.9 and contributing 35.8% to the total phenotypic variance. All three loci were in support intervals that had been identified in the single–QTL analysis. In the Athens analysis, there was an additive interaction between QTL mapping to linkage groups 1A and 7C (Supplementary Table 3). These two loci had a joint LOD score of 5.7 and accounted for 17.8% of the total phenotypic variance. One significant three-locus additive interaction occurred in Salina, among loci mapping to linkage groups 1A, 2C, and 7C, with a

combined LOD score of 13.3 and accounting for 36.6% of the total phenotypic variance. All three loci were in the same support interval as those detected in the single–locus analysis.

Of the eight QTL associated with glume tenacity in H6 (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b), five mapped to linkage groups 4D, 6E, 7D, 8A, and 8B in the across-location analysis, while three others mapped to linkage groups 1C, 2B, and 6B in Athens. Five QTL mapped to 4D, 6E, 7D, 8A, and 8B in all three analyses (Table 1; Supplementary Table 2). LOD scores of individual QTL varied from 2.5 to 3.9, and the percentage of variance ranged from 10.9 to 16.4 (Table 1; Supplementary Table 2). At six of eight loci, the *S. bicolor* allele increased glume tenacity (Table 1; Supplementary Table 2). A QTL associated with glume tenacity in the across-location and Salina multilocus analyses, and also detected in the single-locus analysis, mapped to group 4D (Supplementary Table 3), was initially detected in the single-locus analysis. A trio of additively interacting QTL detected in Athens mapped to groups 1A, 8B and 10A. The interaction had a combined LOD score of 9.8 and contributed 35.9% to the total phenotypic variance (Supplementary Table 3).

Panicle shape

Panicle length (PL) was measured in both locations, while compactness (PC) and width (PW) were rated in Salina and Athens, respectively. Significant QTL were detected for all panicle shape traits in both subpopulations (Table 1; Supplementary Table 2, Fig. 1).

In all, six distinct QTL were found for panicle shape in the H4 subpopulation: two loci for panicle length, one for panicle width, and three for panicle compactness (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3a). LOD scores varied from 2.5 to 5.1, and the percentage of variance explained by a QTL varied from 8.3 to 15.9.

In H6, a total of 26 distinct QTL were associated with panicle shape: 10 for panicle length, three for width, and 10 for compactness (Table 1, Supplementary Table 2, Fig. 1 and Supplementary Fig. 3b). LOD scores of individual QTL ranged from 2.5 to 4.1 and share of phenotypic variance explained ranged from 10.7 to 16.5%. In both subpopulations, allele substitution effects of *S. halepense* alleles were positive at some loci and negative at others. No significant QTL affected more than one panicle-shape trait. Most of the QTL associated with panicle length, width, and compactness in the single-locus analyses were also detected in multiple-QTL analyses (Supplementary Table 3).

Harvest index (HI) was evaluated only in Salina. With single-locus mapping, one significant QTL was identified in each subpopulation, mapping to linkage group 6B (43.17 cM) in H4 and 4A (66.60 cM) in H6 (Table 1 Supplementary Table 2, Fig. 1 and Supplementary Fig. 3). With multiple-QTL analysis in H4, the QTL mapping to group 6B that was detected in the single locus analysis interacted with other QTL mapping to another region of 6B (8.2 cM) and 10B (21.0 cM) (Table 2). This three-way interaction contributed 43% to the total phenotypic variance with a combined LOD score of 16.4. In H6, the QTL mapping to 4A in the single locus analysis interacted with four QTL that mapped to 2A, 5B, 5C, and 6A (Supplementary Table 3). This interaction accounted for 58.2% of the total phenotypic variance with a combined LOD score of 19.3.

Tight linkage or pleiotropy

QTL for different traits were declared to be co-located when their peaks were in the same marker interval. Co-location was designated "coupling" when the allele substitution effects had the same algebraic sign and "repulsion" when they had opposite algebraic signs. In the H4 subpopulation, we found genomic regions on linkage groups 1A, 1F, and 6B where QTL for several traits co-localized (Supplementary Fig. 5). Most notably in H4, in a genomic region of 6B between 0 and 74.7 cM, we found QTL associated with seed yield per plant (*qPGY.H4.6B-2)*, seed yield per panicle (*qPSY.H4.6B-1)*, 1000-seed weight (*qTKWT.H4.6B-7)*, glume tenacity (*qGT.H4-6B-5*), and harvest index (*qHI.H4.6B-1*). Except for glume tenacity, the *S. bicolor* allele increased trait values, as indicated by negative allele substitution effects in Supplementary Fig. 5.

Conversely in H6, we found 14 genomic regions where QTL for several traits co-localized on linkage groups 1C, 2A, 2B, 3B, 3E, 4B, 4C, 4D, 5C, 6D, 6E, 7D, 9C and 10C (Fig. 2 and Supplementary Fig. 5). The smallest interval with overlapping QTL was on linkage group 6D between 0 and 26.6 cM in which we

found QTL for panicle length (*qPL-H6-6D-6*) and panicle compactness (*qPC-H6-6D-11*; Supplementary Fig. 5).

Overlap of seed yield per plant, seed yield per panicle, and 1000-seed weight QTL to previously published QTL of other studies

Except *qPSY-H6-10C* and *qTKWT-H4-3C-3*, all QTL identified for three target traits: seed yield per plant, seed yield per panicle, and 1000-seed weight overlapped with previously published QTL (Supplementary Table 5). The two new QTL (*qPSY-H6-10C* and *qTKWT-H4-3C-3*) could be a result of a unique genetic background (*S. bicolor*^{2} × *S. halepense* subpopulations) used in this study. Despite the complex genetic architecture of grain yield and related traits, this comparison provided evidence of stable QTL across contrasting environments and genetic backgrounds.

Discussion

This *S. bicolor*^{2} × *S. halepense* population was useful for examining the complex genetic architecture underlying several seed-yield-related traits. We also examined genotype × environment interactions, which are very important in the expression of QTL (Paterson et al. 1991; Xu and Crouch 2008; Bernardo 2008; MacKay et al. 2009). In this population, genotype-by-environment interactions estimated from phenotypic data on four of the traits across locations—seed yield per panicle, 1000-seed weight, glume tenacity, and panicle length) —were highly significant. This is consistent with the result that some QTL for these traits exhibited environmental specificity while others did not.

Whereas some of the seed yield per panicle, 1000-seed weight, glume tenacity, and panicle length QTL were classified as "constitutive" because they were consistent across analyses, others had effects that varied across locations and were classified as "adaptive". Of the 18 QTL involved in the above four traits in H4, 15 were detected both in the combined analysis and in at least one single-location analysis, indicating some degree of stability of expression. The other three QTL were detected only in a single location, indicating environmental specificity. A much larger number of QTL affecting the same four traits were identified in H6. Of 37 QTL associated with the above four traits, 23 were detected in the combined analysis and at least one single-location analysis, while 13 were detected only in one location (Supplementary Fig. 6b). QTL detected in both combined and single-location analyses are more likely to be constitutive than those detected in only one location (Ping et al. 2003).

Numerous studies have examined effects of individual QTL in multiple environments, typically finding that environment influences QTL expression to varying degrees. QTL with large, constitutive effects, such as those detected in this study for seed yield per panicle (*qPSY-H6-2A-1*, *qPSY-H6-4D-6*), 1000-seed weight (*qTKWT-H4-6B-7*, *qTKWT-H6-1D-1)*, glume tenacity (*qGT-H6-4D-3*, *qGT-H4-7C-6*), and panicle length (*qPL-H4-1A-1*, *qPL-H4-1C-2*, *qPL-H6-4D-5*) are the ones most useful in breeding programs that employ marker-assisted selection (MAS) and target large geographical regions (Ping et al. 2003; Bauer et al. 2009; Almeida et al. 2013). Environment-specific QTL may also be used to improve productivity in specific target locations (Paterson et al. 1991). For instance, the adaptive QTL on linkage groups 3C (*qPSY-H6-3C-2*) and 5C (*qPSY-H6-5C-7*) that were associated with seed yield per panicle in the H6 subpopulation and detected only in Athens might be found after further study to be useful in selecting for higher seed yield in the southeastern United States. Alternatively, by combining several adaptive QTL into a single genetic background, we might develop genotypes that are buffered from unpredictable or extreme environments (Haggard et al. 2013).

For several traits, alleles from the wild parent (*S halepense*) had a positive effect on the trait mean. For example, of the 10 QTL associated with seed yield per panicle in H6, the *S halepense* allele increased the trait value of five, *qPSY-H6-3E-4*, *qPSY-H6-4A-5*, *qPSY-H6-4D-6*, *qPSY-H6-6E-8*, and *qPSY-H6-10C-10*. The detection of trait-enhancing alleles from agronomically unfavourable wild parents has been widely reported (Tanksley and McCouch 1997; Xiao et al. 1998; Yoon et al. 2006), providing evidence that the phenotype of a germplasm source is a far from perfect predictor of its value as a parent. A number of loci are assumed to underlie any given trait, of which, perhaps, only a subset can be detected. A parent having a high mean value for a certain character does not necessarily have the positive allele at all loci influencing this trait. This suggests the possibility of an allelic series in

which wild relatives may harbor different alleles at given loci, some being superior and others inferior to those found segregating in the cultivated gene pool.

Co-location of QTL controlling multiple seed-yield-related traits, such as we found in several cases, may be attributable either to pleiotropy (Ritter et al. 2008) or tight linkage (Chen and Lübberstedt 2010). In H4, we found several traits co-localized in three genomic regions on linkage groups 1A, 1F, and 6B (Supplementary Fig. 5). Regions mapping to 1F and 6B were associated with both seed yield per plant and seed yield per panicle. These traits, with one being a component of the other, were strongly correlated phenotypically ($r = 0.8, P < 0.001$), and it is expected that some QTL were involved in both traits. On the other hand, pleiotropic effects or tight linkage might exist among some QTL for seed yield per panicle and panicle compactness, although these two traits were not correlated (Supplementary Fig. 2). Both pleiotropy and tight linkage can be either beneficial or detrimental in the context of plant breeding (Brown 2002; Boerma and Walker 2005), and it is useful to understand the underlying genetic basis of correlation among multiple traits. For example, late-maturing, blast-resistant *indica* rice cultivars had been used as donors of resistance genes into susceptible, early-maturing *japonica* varieties; however, even with extensive backcrossing, poor grain quality and late-flowering were introduced along with the target genes (Zhao et al. 2010). This sort of phenomenon often assumed to result from "linkage drag," has been reported for numerous economically important species, including bean (Miklas 2007), canola (Cao et al. 2010), potato (Collins et al. 1999), tobacco (Lewis and Rose 2010), and tomato (Frary et al. 2004). For some traits, even after many backcross generations and selection, selected genes may be accompanied by linked genomic segments large enough to carry hundreds of undesirable genes (Newbury 2003; Collard and Mackill 2008; Xu 2010; Singh and Singh 2015).

In this study, we were able to detect statistically significant digenic epistatic interactions between unlinked QTL, with either positive or negative effects. For genes affecting quantitative traits, epistasis has been defined as a deviation from the sum of independent effects of individual genes (Falconer 1989). In H4, we found evidence of positive epistasis for seed yield per plant and seed yield per panicle (Supplementary Table 3). In H6, there was evidence of epistasis in seed yield per panicle (positive), 1000-seed weight (positive), and harvest index (negative) (Supplementary Table 3). In most cases, the proportion of phenotypic variation explained by these digenic epistatic interactions was small, implying that genetic variation was mainly influenced by either single-locus and/or higher-order epistatic interactions—involving alleles at three or more loci (Taylor and Ehrenreich 2015). Conventional genetic mapping methods have generally low statistical power to identify higher-order epistatic interactions (Carlborg et al. 2006; Cordell 2009). Our power to detect even simple digenic epistatic effects was further limited because our population was derived from two non-identical F1 plants; therefore, we had to analyze the two smaller subpopulations separately. For this reason, we acknowledge that a significant proportion of digenic interactions may have gone undetected.

This study had some limitations. Because of the difficulty in producing large numbers of cross-pollinated seed with this material, we managed to obtain only 246 families. Population sizes of less than 500, as in this study, can inflate the mean magnitude of QTL effects – increasing the number of individuals and/or markers is expected to resolve trait variation into more QTL with smaller effects (Beavis 1997; Xu 2003). It is likely that additional QTL exist for the traits presented here, but that the effects were too small to be detected with the population size available. Further, our tests for QTL × environment interaction by comparing QTL detected in different environments have some limitations. Individual QTL × environment interaction effects are difficult to measure, largely because of a lack of appropriate analytical methods. Furthermore, only a single threshold is used. It remains unknown whether inconsistency of QTL detection across environments is due to type-II error arising from the use of single thresholds, or to true differential trait expression across environments. Multi-environment joint analysis methods can avoid such problems.

Phenotypic correlations between traits in the overall BC_1F_2 population between traits that are anatomically, physiologically, and/or mathematically interrelated are to be expected, and were found for interrelated trait pairs such as seed yield per plant and seed yield per panicle ($r = 0.8, P < 0.001$); seed yield per panicle and 1000-seed weight ($r = 0.7$, $P < 0.001$), seed yield per plant and 1000-seed weight

($r = 0.5$, $P < 0.001$), seed yield per panicle and harvest index ($r = 0.8$, $P < 0.001$), seed yield per plant and harvest index ($r = 0.8$, $P < 0.001$) and 1000-seed weight and harvest index ($r = 0.6$, $P < 0.001$). Consistent with that, we found some QTL that affected more than one trait.

In conclusion, genetic improvement of perennial sorghum can be achieved by either directly selecting for grain yield or indirectly through selecting for secondary traits related to higher grain yield potential. Identifying stably expressed QTL leading to higher grain yield is an important target for the genetic improvement of perennial sorghum. QTL associated with grain yield and related traits are useful for marker-assisted selection of high-yielding genotypes.

Author contributions All authors contributed to and approved the final manuscript. SC conceived the study, generated the population, and contributed to experimental setup and phenotyping, manuscript review; PN contributed to phenotyping, data analyses, and manuscript write-up, WK contributed to phenotyping, mapping, marker identification, and data analyses; AP co-conceived the study, contributed to phenotyping, marker identification, data interpretation, and manuscript review.

Funding This work was funded in whole or part by the United States Agency for International Development (USAID) Bureau for Resilience and Food Security under Agreement # AID-OAAA-13-00044 as part of Feed the Future Innovation Lab for Climate Resilient Sorghum. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

Availability of data and material The datasets generated during and/or analysed during the current study are available from the corresponding author on request.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publication Not applicable.

References

Arriola PE, Ellstrand NC (1996) Crop-to-weed gene flow in the genus Sorghum (Poaceae): spontaneous interspecific hybridization between johnsongrass, *Sorghum halepense*, and crop sorghum, *S. bicolor*. Am J Bot 83:1153–1159

Almeida GD, Makumbi D, Magorokosho C et al (2013) QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theor Appl Genet 126:583–600. https://doi.org/10.1007/s00122-012-2003-7

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Statist Softw 67(1):1–48

Bauer AM, Hoti F, von Korff M et al (2009) Advanced back-cross-QTL analysis in spring barley (*H. vulgare* ssp. *spontaneum*) comparing a REML versus a Bayesian model in multi-environmental field trials. Theor Appl Genet 119:105–123. https://doi.org/10.1007/s00122-009-1021-6

Beavis W (1997) QTL analyses: power, precision and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, New York

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Boerma HR, Walker DR (2005) Discovery and utilization of QTLs for insect resistance in soybean. Genetica 123:181–189

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinform Appl Note 19:889–890. https://doi.org/10.1093/bioinformatics/btg112

Broman KW, Sen S (2009) Guide to QTL mapping with R/qtl. Springer-Verlag, New York, NY, pp 1–396

Brown JK (2002) Yield penalties of disease resistance in crops. Curr Opin Plant Biol 5:339–344. https://doi.org/10.1016/S1369-5266(02)00270-4

Cao Z, Tian F, Wang N et al (2010) Analysis of QTLs for erucic acid and oil content in seeds on A8 chromosome and the linkage drag between the alleles for the two traits in *Brassica napus*. J Genet Genom 37:231–240. https://doi.org/10.1016/S1673-8527(09)60041-2

Carlborg O, Jacobsson L, Ahgren P et al (2006) Epistasis and the release of genetic variation during long-term selection. Nat Genet 38:418–420. https://doi.org/http://www.nature.com/ng/journal/v38/n4/suppinfo/ng1761_S1.html

Chen Y, Lübberstedt T (2010) Molecular basis of trait correlations. Trends Plant Sci 15:454–461. https://doi.org/10.1016/j.tplants.2010.05.004

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196. https://doi.org/10.1007/s10681-005-1681-5

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363:557–572

Collins A, Milbourne D, Ramsay L et al (1999) QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. Mol Breed 5:387–398. https://doi.org/10.1023/A:1009601427062

Cordell HJ (2009) Detecting gene–gene interactions that underlie human diseases. Nat Rev Genet 10:392–404

Cox S, Bender M, Picone C et al (2002) Breeding perennial grain crops. CRC Crit Rev Plant Sci 21:59–91

Cox S, Glover JD, Van Tassel DL et al (2006) Prospects for developing perennial grain crops. Bioscience 56:649–659. https://doi.org/10.1641/0006-3568(2006)56[649:pfdpgc]2.0.co;2

Cox TS, Nabukalu P, Paterson AH et al (2018) Development of perennial grain sorghum. Sustain 10:. https://doi.org/10.3390/su10010172

DeHaan LR, Van Tassel DL, Cox TS (2005) Perennial grain crops: a synthesis of ecology and plant breeding. Renew Agric Food Syst 20:5–14. https://doi.org/10.1079/RAF200496

Dinnes DL, Karlen DL, Jaynes DB et al (2002) Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. Agron J 94:153–171

Entz MH, Baron VS, Carr PM et al (2002) Potential of forages to diversify cropping systems in the Northern Great Plains. Agron J 94:240–250

Falconer DS, Douglas S (1989) Introduction to quantitative genetics, 3rd ed. Longman, Scientific & Technical, Burnt Mill

Frary A, Fulton TM, Zamir D, Tanksley SD (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theor Appl Genet 108:485–496. https://doi.org/10.1007/s00122-003-1422-x

Gantzer CJ, Anderson SH, Thompson AL, Brown JR (1990) Estimating soil erosion after 100 years of cropping on Sanborn Field. J Soil Water Conserv 45(6):641–644

Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipelineTASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. https://doi.org/10.1371/journal.pone.0090346

Glover JD, Reganold JP, Bell LW et al (2010) Increased food and ecosystem security via perennial grains. Science 328:1638–1639

Habyarimana E, Lorenzoni C, Redaelli R et al (2018) Towards a perennial biomass sorghum crop: a comparative investigation of biomass yields and overwintering of *Sorghum bicolor* × *S. halepense* lines relative to long term *S. bicolor* trials in northern Italy. Biomass Bioenerg 111:187–195. https://doi.org/10.1016/j.biombioe.2017.03.004

Hadley HH (1958) Chromosome numbers, fertility and rhizome expression of hybrids between grain Sorghum and Johnsongrass1. Agron J 50:278. https://doi.org/10.2134/agronj1958.00021962005000050015x

Haggard JE, Johnson EB, St. Clair DA (2013) Linkage relationships among multiple QTL for horticultural traits and late blight (*P. infestans*) resistance on chromosome 5 introgressed from wild tomato *Solanum habrochaites*. G3 Genes, Genomes, Genet 3:2131–2146. https://doi.org/10.1534/g3.113.007195

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Holm LRG (1991) The world's worst weeds: distribution and biology. Krieger Publishing Company, Malabar

ICRISAT (1993) Descriptors for Sorghum: *Sorghum bicolor* (L.) Moench = Descripteurs du sorgho: *Sorghum bicolor* (L.) Moench. IBPGR/ICRISAT

Kong W, Nabukalu P, Cox TS et al (2020) Transmission Genetics of a *Sorghum bicolor* × *S. halepense* Backcross Populations. Front Plant Sci 11:467. https://doi.org/10.3389/fpls.2020.00467

Lenth RM (2016) Least-squares means: The R package lsmeans. J Statist Softw 69:1–33

Lewis RS, Rose C (2010) Agronomic performance of tobacco mosaic virus-resistant tobacco lines and hybrids possessing the resistance gene N introgressed on different chromosomes. Crop Sci 50:1339–1347. https://doi.org/10.2135/cropsci2009.10.0615

MacKay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577

Manichaikul A, Moon JY, Sen Ś et al (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics 181:1077–1086. https://doi.org/10.1534/genetics.108.094565

McWhorter CG (1961) Morphology and development of johnsongrass plants from seeds and rhizomes. Weeds 558–562

Miklas PN (2007) Marker-assisted backcrossing QTL for partial resistance to sclerotinia white mold in dry beanMarker-assisted backcrossing QTL for partial resistance to sclerotinia white mold in dry bean. Crop Sci 47:935–942. https://doi.org/10.2135/cropsci2006.08.0525

Nabukalu P, Cox TS (2016) Response to selection in the initial stages of a perennial sorghum breeding program. Euphytica. https://doi.org/10.1007/s10681-016-1639-9

Newbury HJ (2003) Plant molecular breeding. Blackwell, Oxford

Paterson A, Damon S, Hewitt JD et al (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127(1):181–197

Paterson A, Schertz KF, Lin YR et al (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. Proc Natl Acad Sci USA 92:6127–6131. https://doi.org/10.1073/pnas.92.13.6127

Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. https://doi.org/10.1038/nature07723

Pimentel D, Allen J, Beers A et al (1987) World agriculture and soil erosion. Bioscience 37:277–283

Ping LI, LiKhuang ZHU, YingKguo ZHU (2003) Characterization of QTLs for harvest index and source-sink characters in a DH population of rice (*Oryza sativa* L.). Acta Genet Sin 30(12):1118–1126

Piper JK, Kulakow PA (1994) Seed yield and biomass allocation in *Sorghum bicolor* and F_1 and backcross generations of *S. bicolor* × *S. halepense* hybrids. Can J Bot 72:468–474. https://doi.org/10.1139/b94-062

Randall GW, Mulla DJ (2001) Nitrate nitrogen in surface waters as influenced by climatic conditions and agricultural practices. J Environ Qual 30:337–344. https://doi.org/10.2134/jeq2001.302337x

Ritter KB, Jordan DR, Chapman SC et al (2008) Identification of QTL for sugar-related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Mol Breed 22:367–384. https://doi.org/10.1007/s11032-008-9182-6

Rooney W (2004) Sorghum improvement–integrating traditional and new technology to produce improved genotypes. Adv Agron 83:37–109

Schertz KF, Clark LE (1967) Controlling dehiscence with plastic bags for hand crosses in Sorghum 1. Crop Sci 7:540–542. https://doi.org/10.2135/cropsci1967.0011183x000700050040x

Singh BD, Singh AK (2015) Marker-assisted plant breeding: principles and practices. Springer, New Delhi

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066. https://doi.org/10.1126/science.277.5329.1063

Taylor MB, Ehrenreich IM (2015) Higher-order genetic interactions and their contribution to complex traits. Trends Genet 31:34–40. https://doi.org/10.1016/j.tig.2014.09.001

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78. https://doi.org/10.1093/jhered/93.1.77

Wagoner P (1990) Perennial grain development: past efforts and potential for the future. CRC Crit Rev Plant Sci 9:381–408. https://doi.org/10.1080/07352689009382298

Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627–640

Xiao J, Li J, Grandillo S et al (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899

Xu Y (2010) Molecular plant breeding. CABI, Wallingford

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

Xu S (2003) Theoretical basis of the beavis effect. Genetics 165(4):2259–2268

Yoon DB, Kang KH, Kim HJ et al (2006) Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between *Oryza grandiglumis* and the *O. sativa* japonica cultivar Hwaseongbyeo. Theor Appl Genet 112:1052–1062

Zhang H, Mittal N, Leamy LJ et al (2017) Back into the wild-apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl 10:5–24. https://doi.org/10.1111/eva.12434

Zhao K, Wright M, Kimball J et al (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One. https://doi.org/10.1371/journal.pone.0010780