ORIGINAL RESEARCH

Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm

Richard E. Boyles, Elizabeth A. Cooper, Matthew T. Myers, Zachary Brenton, Bradley L. Rauh, Geoffrey P. Morris, and Stephen Kresovich*

Abstract

Grain yield and its primary determinants, grain number and weight, are important traits in cereal crops that have been well studied; however, the genetic basis of and interactions between these traits remain poorly understood. Characterization of grain yield per primary panicle (YPP), grain number per primary panicle (GNP), and 1000-grain weight (TGW) in sorghum [*Sorghum bicolor* (L.) Moench], a hardy C_4 cereal with a genome size of ~730 Mb, was implemented in a diversity panel containing 390 accessions. These accessions were genotyped to obtain 268,830 single-nucleotide polymorphisms (SNPs). Genome-wide association studies (GWAS) were performed to identify loci associated with each grain yield component and understand the genetic interactions between these traits. Genome-wide association studies identified associations across the genome with YPP, GNP, and TGW that were located within previously mapped sorghum QTL for panicle weight, grain yield, and seed size, respectively. There were no significant associations between GNP and TGW that were within 100 kb, much greater than the average linkage disequilibrium (LD) in sorghum. The identification of nonoverlapping loci for grain number and weight suggests these traits may be manipulated independently to increase the grain yield of sorghum. Following GWAS, genomic regions surrounding each associated SNP were mined for candidate genes. Previously published expression data indicated several TGW candidate genes, including an ethylene receptor homolog, were primarily expressed within developing seed tissues to support GWAS. Furthermore, maize (*Zea mays* L.) homologs of identified TGW candidates were differentially expressed within the seed between small- and large-kernel lines from a segregating maize population.

Core Ideas

- Association mapping elucidated the genetic basis of sorghum grain yield components.
- GWAS results suggest yield component traits may be manipulated independently.
- The yield component loci identified may be targeted for grain sorghum improvement.

A GROWTH IN INTEREST in sorghum production for use as a food, feed, and biofuel feedstock has occurred as a result of the crop's diversity, non-GMO status, and ability to thrive in harsh environments (Dahlberg et al., 2011). As further new end-use commodities, such as gluten-free foods and industrial products, create additional markets, it seems the demand for sorghum grain will continue to rise. This increase in demand places an importance on grain yield where progress in sorghum has been slower in comparison to other cereals including maize and rice (*Oryza sativa* L.) (Mason et al., 2008; FAO, 2015). Whole-genome sequencing in sorghum has revealed genetic diversity that has yet to be fully

R.E. Boyles, Z. Brenton, S. Kresovich, Dep. of Genetics and Biochemistry, Clemson University, Clemson, South Carolina 29634; E.A. Cooper, M.T. Myers, B.L. Rauh, S. Kresovich, Advanced Plant Technology Program, Clemson University, Clemson, South Carolina 29634; G.P. Morris, Dep. of Agronomy, Kansas State University, Manhattan, Kansas 66506. *Corresponding author (skresov@clemson.edu). Accepted 30 Nov. 2015. Received 15 June 2015.

Abbreviations: BSLMM, Bayesian sparse linear mixed model; DBW, dry vegetative biomass weight; DPW, dry panicle weight; DTA, days to anthesis; DTM, days to maturity; FPKM, fragments per kilobase of exon per million reads mapped; GNP, grain number per primary panicle; GWAS, genome-wide association studies; LD, linkage disequilibrium; MLM, mixed linear model; QTL, quantitative trait loci; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism; TGW, 1000-grain weight; YPP, grain yield per primary panicle.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.09.0091

exploited for crop improvement (Paterson et al., 2009; Mace et al., 2013). Finding genetic controls of YPP and its components, grain number and weight, would allow breeding efforts to manipulate these traits and potentially elevate the current yield ceiling, making the crop more efficient and desirable to the grower.

Grain yield in cereal crops is determined by four primary components: number of plants per area (plant density) (Sukumaran et al., 2015b), number of panicles per plant, number of grains per panicle, and grain weight. This study of sorghum focuses on the latter two components with respect to the primary panicle, the panicle of the main culm, to characterize traits and identify candidate genes important to grain yield. Maximizing YPP could be beneficial by reducing the C investment of the plant required to generate additional stalks, that is, tillers, and devoting more energy into filling grain. Grain number and weight are complex, quantitative traits that are influenced both genetically and environmentally (Austin and Lee, 1998). The trade-off observed between these two traits has been reviewed (Sadras, 2007), but recent evidence from studies in sorghum (Gambín and Borrás, 2012), wheat (*Triticum* spp.) (Griffiths et al., 2015), and *Arabidopsis thaliana* (L.) Heynh. (Gnan et al., 2014) indicate increasing one yield component without reducing the other is plausible. The critical periods for determination of grain number and weight are also generally considered separated by the developmental stage of anthesis (flowering), although Ugarte et al. (2007) found that grain weight was affected by preanthesis environmental conditions in other cereals including wheat. A strong presence of abiotic stress to limit photosynthesis during the reproductive stage may enhance the trade-off between grain number and weight because of finite C availability; however, sorghum typically has excess assimilates available in the stalk during grain development and at physiological maturity when grown under favorable conditions (unpublished data, 2013), suggesting assimilate production is not the sole limiting factor underlying the trade-off between these traits. Murray (2012) reports that the majority of studies on the source–sink relationship in sorghum provide evidence toward sink limitations hindering grain yield, although there is controversy in the scientific literature.

Like that of maize and wheat, final GNP in sorghum is determined by the total number of fertile and unaborted florets produced by the inflorescence (van Oosterom and Hammer, 2008). A previous study in sorghum found grain number to be the primary determinant of grain yield, accounting for 52% (Borrell et al., 1999). The crop growth period and rate from floral meristem initiation until anthesis are critical for a plant's response in determining floret number (Ritchie et al., 1998). In maize, there is high variation of total kernel number among individuals when water and nutrient limitations are not present, which indicates number of seeds produced in cereals is partially driven by genetic variation (Amelong et al., 2015). The genotype × environment interaction for grain number is likely strong in sorghum based on previous studies on yield components in other cereal grains (Sadras and Richards, 2014; Xu et al., 2014; Amelong et al., 2015). Sadras (2007) reviewed how grain number has maintained higher phenotypic plasticity throughout domestication events when compared with grain weight, which enables sorghum to effectively respond to resource availability during early reproductive stages. Thus, finding significant associations for grain number across multiple environments may be challenging, although identifying environment-specific markers for yield-related traits is still beneficial.

Comparing yield components under investigation, grain weight has been studied more extensively in sorghum and cereal crops in general. Grain weight variation in sorghum can be attributed to preanthesis ovary size, endosperm cell number, and size of cells within the endosperm, the major storage compartment in the seed (Yang et al., 2009). Although grain weight in maize has considerably less variation than grain number, there was still nearly a threefold range observed by Severini et al. (2011). Zhang et al. (2015) found the mass of individual grains from diverse sorghum genotypes ranged from 11 to 60 mg. Cisse and Ejeta (2003) found grain weight to be highly heritable in sorghum, and Yang et al. (2009) revealed that much of this genetic control of grain weight is likely additive, making genome-wide association mapping in diverse germplasm an excellent tool to discover multiple loci that contribute to the overall variance in grain weight. The recent study of Zhang et al. (2015) also compared GWAS results of different seed size related traits with mapped grain weight quantitative trait loci (QTL) from previous studies and found several significant associations located within QTL intervals.

To explore candidate genes underlying yield-related traits, GWAS were conducted to identify underlying loci for each phenotype. Association mapping has been used to successfully discover significant marker–trait associations in cereal crops including maize (Remington et al., 2001; Thornsberry et al., 2001), rice (Huang et al., 2010), barley (*Hordeum vulgare* L.) (Cockram et al., 2010), wheat (Neumann et al., 2011; Sukumaran et al., 2015a), and sorghum (Sukumaran et al., 2012; Morris et al., 2013a). Shehzad et al. (2009) conducted association analyses for yield components using simple-sequence repeats on 107 diverse sorghum accessions but found few significant marker–trait associations. The increased number of individuals evaluated in this study and greater genome coverage with high-throughput genotyping will improve power in an effort to better detect significant associations. The primary advantage of GWAS is the potential to capture greater diversity by including a large number of unrelated individuals with distinct genetic backgrounds, while limitations include a potential lack of power to detect minor effect loci and susceptibility to identify spurious associations (Morris et al., 2013b). In an attempt to minimize spurious associations, two distinct linear models have been implemented in this study

to identify overlapping genomic regions associated with yield-related traits. Successful identification of true positives is a critical step in ultimately finding genes containing casual allelic variants.

In this study, the genetic basis of grain number and weight was characterized in a diverse sorghum panel to reveal favorable alleles that could be introgressed and stacked into elite germplasm as a strategy to raise the achievable grain yield of sorghum. Genome-wide association studies of YPP, GNP, and TGW identified loci associated with each trait and highlighted genomic regions for targeted resequencing to determine causal allelic variants and support marker development. Tissue- and time-specific gene expression using publically available sorghum RNA sequencing (RNA-seq) data was also investigated to provide further support for selected candidate genes. Understanding the genetic basis underlying the natural variation of yield-related traits is important to aid crop improvement and increase grain yield in sorghum, a crop that has beneficial agronomic characteristics but a lower average grain yield than other cultivated cereal species (FAO, 2015).

MATERIALS AND METHODS

Field Design

A total of 390 diverse accessions were planted in 2013 and 2014 in Florence, SC. Of the 390 accessions, 332 were from the original US sorghum association panel developed by Casa et al. (2008). The remaining 58 accessions were included for their unique phenotype, diverse origin, or elite grain classification. Seed was originally obtained from the Germplasm Resources Information Network (GRIN) (USDA–ARS–National Genetic Resources Program, 2014). Accessions were grown out near San José del Valle, Nayarit, Mexico, and panicles of each genotype were bagged before anthesis for self-pollination to increase seed numbers required for planting. The experiment was planted 15 May 2013 and 7 May 2014 in Florence, SC, in a completely randomized design and replicated twice in each year. Each plot consisted of two rows, 6.1 m in length, and a row spacing of 0.762 m. Calculated from seeding rate and using a plant establishment of 75% (El Naim et al., 2012), plant density was ~130,000 plants ha^{-1}. Abiotic stress was minimal, as appropriate water and nutrients were applied to observe the trait relationships in favorable growing conditions. Fields were irrigated on an as-needed basis to prevent confounding effects on yield as a result of maturity effects and varying degrees of drought tolerance across genotypes. Variable-rate fertilizer (N, P, and K) before planting and 125 kg ha^{-1} of lay-by N ~30 d after planting were applied both years. A preemergence herbicide, Bicep II Magnum (*S*-metolachlor + atrazine; Syngenta), was applied at 3.5 L ha^{-1} before planting. A postemergent application of atrazine at a rate of 4.7 L ha^{-1} was administered each year when average plant height reached 40 cm. In 2013, Tracer 4E (spinosad; Dow AgroSciences) was applied at 0.15 L ha^{-1} 80 d after planting to control for corn earworms (*Helicoverpa zea*) and minimize reductions in grain yield. No insecticides were required throughout the 2014 field season.

Phenotype Collection

The primary panicle, the panicle radiating from the main culm, from three random plants per plot was covered with a mesh bag at date of anthesis to prevent data bias as a result of grain losses from bird and insect predation. The first and last plant in each row was not considered to eliminate confounding results caused by border effect. Data from plots with low plant stands were excluded from analyses to minimize the effect of plant density on GNP and TGW (Sukumaran et al., 2015b). Harvest of secondary panicles from tillers in sorghum was avoided to prevent strongly confounding grain yield and number with flowering time and biasing TGW as a result of harvest of immature panicles. Delaying harvest until all tillers matured would have likely altered grain weight from unnecessary grain weathering and further biased results. Number of days to anthesis (DTA) was recorded at bagging date and was measured as days after planting to when 50% of the plants in the plot were at midbloom. Number of days to maturity (DTM) was the number of days from planting to physiological maturity (denoted by presence of black layer on the basal grains of the panicle). Grain fill duration was measured as the difference between DTM and DTA in days. Height (cm) was taken from ground to apex of main panicle at physiological maturity. The three plants that contained bagged panicles were harvested at the plant base with a machete at physiological maturity. Mean harvest dates were 29 Aug. 2013 and 3 Sept. 2014.

Panicles were separated from the rest of the plant, dried for 10 to 14 d in an electric dryer to reach a constant weight (~10% moisture content) and threshed manually by hand to avoid grain loss and contamination caused by mechanical threshers. Before threshing, all three panicles were weighed to get an average dry panicle weight (DPW). The aboveground vegetative tissues (stalk and leaves) of the three plants were also dried to constant weight to collect dry vegetative biomass weight (DBW) and calculate harvest index: harvest index = YPP/(DBW + DPW). The grain obtained after threshing was first cleaned with an air aspirator (AT Ferrell Company, Inc.) and then a wheat dehuller (Precision Machine Co., Inc.) to remove glumes still attached to the seed. The proportion of grains with glumes still attached after aspiration was used to measure glume tenacity. Glume tenacity, along with head mold and head smut, was ranked on a nominal scale from 1 to 5 (1 = no occurrence and 5 = severe occurrence). The cleaned grain was run through seed counters (Old Mill Model 900-2) to record GNP. Total YPP was measured with a Discovery series scale (Ohaus), and TGW was calculated by YPP and GNP: TGW (g) = (YPP/GNP) × 1000.

Genotype-by-Sequencing

Raw sequencing reads previously obtained as described in Morris et al. (2013a) were combined with additional

sequencing data for 25 individuals to increase coverage. Resequencing of the 25 individuals was performed at the Clemson University Genomics and Computational Laboratory following the same protocols as used in the original data set (Morris et al., 2013a). All raw data were aligned to the most recent version of the sorghum reference genome (v2.1; www.phytozome.net) and filtered with the TASSEL 5.0 GBS pipeline (Glaubitz et al., 2014), resulting in a total of 268,830 SNPs. Average intermarker distance using this density of SNPs was 2.7 kb. Missing genotypes were imputed in fastPHASE (Scheet and Stephens, 2006) with 20 independent starts of the expectation–maximization algorithm. Following imputation, individuals with more than 30% missing data remaining were removed, which resulted in 375 and 378 individuals included in GWAS with phenotypic data in 2013 and 2014, respectively.

Phenotypic Analysis

The simple mean from three data values (one value for each panicle) for YPP, GNP, and TGW was calculated within each replicate in both years. The trait correlation matrix was generated by the Pearson method with the cor() function in R software (R Core Development Team, 2013). The cor.test() function in R was used to determine significance for each correlation. The chart.Correlation() function within the PerformanceAnalytics package was used to generate scatter plots and histograms (Peterson et al., 2014). Variance components (σ^2) using multiyear and replicated data were calculated with the lme4 package in R (Bates et al., 2015). All effects were treated as random. The lmer() function within this package optimized the linear mixed model using restricted maximum likelihood and was implemented to determine variance components for each random effect. Because there was only one location in each year, replicates were used in the heritability calculation in place of location along with interaction between genotype and year to estimate the variance caused by genotype × environment interaction. Broad-sense heritability (H^2) using the calculated variance components was estimated as:

$$H^2 = \sigma^2_G/[\sigma^2_G + (\sigma^2_{G\times R}/R) + (\sigma^2_{G\times Y}/Y) + (\sigma^2_E/RY)]$$

where G is genotype, R is replication, Y is year, and E is error.

Genome-Wide Association Studies

Two linear models were used to perform association analyses: a mixed linear model (MLM) and Bayesian sparse linear mixed model (BSLMM). The MLM was implemented using the Genome Association and Prediction Integrated Tool (GAPIT) in R (Lipka et al., 2012) to determine significant associations among yield-related traits and SNPs across individuals (Sukumaran et al., 2012; Morris et al., 2013b). Quantile–quantile plots of the association results from the MLM suggest the model effectively controlled for false positives (Supplemental Fig. S1). To perform the MLM, population structure (Q), as estimated by the Bayesian Markov chain Monte Carlo program STRUCTURE (Pritchard et al., 2000), was incorporated into the model as a covariate. We initially selected a subset of 64,019 SNPs with coverage >80% and minor allele frequencies >10% in TASSEL (Glaubitz et al., 2014) and thinned to 12,200 SNPs by removing loci that were <20 kb apart using VCFtools (Danecek et al., 2011). A priori values of $k = 1$ through $k = 12$ were used for individual model runs with three replicates averaged for each k-value. Each run consisted of 2×10^4 burn-in steps followed by 10^5 sampling iterations. The optimal value of k was determined based on the estimated logarithm likelihood of the data, which increased exponentially until $k = 4$ (Supplemental Fig. S2), and thus this was the primary Q matrix chosen. This optimal number of populations is consistent with a previous analysis of the US sorghum association panel (Adeyanju et al., 2015). Kinship (K) was internally calculated within GAPIT using the default VanRaden method to estimate relatedness (VanRaden, 2008). Permutation tests as described in Zhang et al. (2015) were conducted for each trait in each year to determine the empirical significance threshold of $p = 10^{-5}$. The proportion of sampled permutations across traits where the p-value was smaller than $p = 10^{-5}$ ranged from 2.60×10^{-6} to 7.44×10^{-6}, indicating that this p-value was an adequate significance cutoff to control for Type I errors.

Polygenic modeling with the BSLMM was also implemented using Genome-Wide Efficient Mixed Model Association (GEMMA) software (Zhou et al., 2013), which takes into account multiple SNPs to fit the model. Results from 10 separate runs using 5×10^6 and 20×10^6 burn-in iterations and sampling steps, respectively, were averaged together for each trait. For this multilocus model, a posterior inclusion probability >0.01 ($p = 1.79 \times 10^{-5}$) was considered significant based on the null distribution of posterior inclusion probability values from 10 simulated data sets (Supplemental Table S1) and the significance thresholds used in other studies (W. Bridges, personal communication, 2015; Comeault et al., 2014). For the 10 simulated data sets, 268,830 genotypes (the same size as the original dataset) were simulated for all individuals as Bernoulli random variables so that no loci were expected to be associated with the phenotype. Grain number per primary panicle from 2013 was used as the phenotype in the simulated data set because this trait contained the highest variation.

For all GWAS, regardless of model, the phenotypic data from two replicates within each year were averaged using least squares. The CSGRqtl database (Zhang et al., 2013) was used to identify significant SNPs associated with yield-related traits located within existing QTL intervals. Following GWAS, genes within 20 kb of a SNP that was significantly associated with a yield-related trait were extracted from the Sorghum v2.1 reference genome in Phytozome (www.phytozome.net) and studied for functional relevance in regard to each trait. The 20-kb window was used based on average LD previously identified in sorghum (Hamblin et al., 2004; Bouchet et al.,

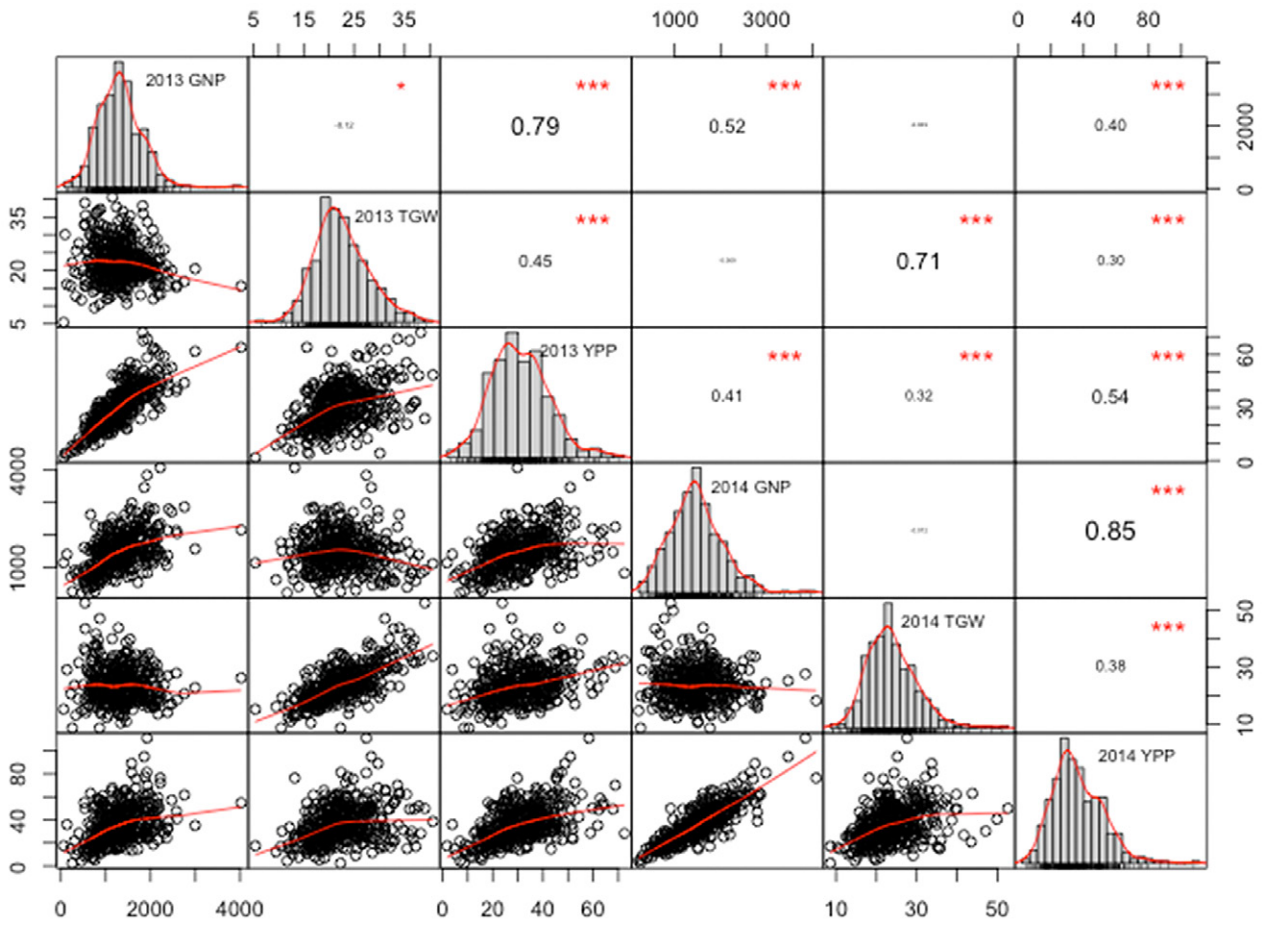

Fig. 1. Variation and Pearson pairwise correlations among yield-related traits. Histograms for grain number per primary panicle (GNP), 1000-grain weight (TGW, g), and grain yield per primary panicle (YPP, g) are displayed along the diagonal. To the left and below the diagonal are scatter plots containing measured individuals from the diverse panel of 390 accessions. The red line through the scatter plot represents the line of best fit. Pearson correlation coefficients between yield-related traits are shown above and to the right of the diagonal. The correlation significance levels are: *p = 0.05, **p = 0.01, and ***p = 0.001, and the size of the coefficient values are proportional to the strength of the correlation.

2012; Mace et al., 2013). The putative functions of candidate genes were determined based on their homology to functionally characterized genes in other plant species. Finally, the effects of SNPs within or near candidate genes were predicted using SnpEff (Cingolani et al., 2012).

Comparing Expression Data

Publically available RNA-seq data for sorghum BTx623 (Dugas et al., 2011; Davidson et al., 2012; Makita et al., 2015) and maize 'Krug' recombinant inbred lines (Sekhon et al., 2014) were mined for potential expression changes across all candidate genes in LD with an associated SNP. These data comprised multiple tissues and developmental stages. Expression of candidate genes across various sorghum tissues and developmental stages were investigated to (i) provide corroborating evidence to substantiate candidates and (ii) target specific candidate genes for future sequencing and validation. Gene expression was measured in units of fragments per kilobase of exon per million reads mapped (FPKM) (Trapnell et al., 2010).

RESULTS

Trait Characterization and Heritability

Values for YPP, GNP, and TGW across the 390 accessions included in this study are listed in Supplementary Table S2. Overall, the sorghum diversity panel exhibited extensive trait variation across years (Fig. 1; Supplemental Table S3). Mean number of DTA was much later in 2014 (80 d) than 2013 (68 d); however, mean height decreased from 149 cm in 2013 to 114 cm in 2014. These changes observed in 2014 did not appear to have an adverse effect on yield components given YPP, GNP, and TGW all had higher minimum, maximum, and mean values in 2014 than 2013. Thousand-grain weight displayed more consistent variation between years (2013, 7.5-fold; 2014, 6.1-fold) than GNP and YPP. Grain number per primary panicle had greater variation in 2013 (63.9-fold), while YPP was more variable in 2014 (50.6-fold).

Grain yield components were also observed in context of the five primary botanical races (bicolor, caudatum, durra, guinea, and kafir) and their intermediates. Race of the 390 accessions was either based on classification reported in Casa et al. (2008) or the GRIN database

Table 1. Pearson correlation coefficients for 13 traits calculated for 2013 and 2014. Correlations from 2013 data are represented on the upper right of the diagonal. Correlations from 2014 data are shown below and to the left of the diagonal.

		Trait‡												
	H^2†	DTA	DTM	GFD§	Height	DPW¶	DBW#	HI††	GT‡‡	Head mold	Head smut	GNP	TGW	YPP
DTA	0.9	–	0.68***	0.06	0.37***	0.09	0.57***	–0.36***	0	–0.09	–0.15**	0.1	–0.19***	0.02
DTM	0.83	0.87***	–	0.61***	0.16**	0.16**	0.55***	–0.3***	–0.05	0.1	0.03	0.11*	–0.04	0.11*
GFD	0.42	0.09	0.03	–	0.08	0.08	0.2***	–0.06	0	0.23***	0.18***	–0.04	0.24***	0.07
Height	0.95	–0.01	–0.05	–0.01	–	0.07	0.43***	–0.28***	0.12*	–0.2***	–0.2***	–0.02	0.05	0.04
DPW	0.63	0.26***	0.25***	–0.03	–0.04	–	0.29***	0.34***	–0.32***	0.05	0.01	0.66***	0.43***	0.87***
DBW	0.78	0.47***	0.45***	0.04	–0.05	0.25***	–	–0.5***	–0.12*	0.01	–0.04	0.13*	0.17**	0.24
HI	0.68	–0.2***	–0.16**	–0.08	0.06	0.33***	–0.52***	–	–0.22***	0.04	0.05	0.49***	0.15**	0.5***
GT	0.8	0	–0.03	0.02	–0.05	–0.11*	–0.06	–0.18***	–	0.05	0.07	–0.26***	–0.22***	–0.36***
Head mold	0.44	0.23***	0.31***	–0.11*	–0.08	0.06	0.02	–0.01	–0.02	–	0.64***	0	0.11*	0.02
Head smut	0.84	–0.02	0.09	–0.03	0.03	0.05	–0.04	0.09	–0.08	0.44***	–	–0.06	0.22***	0.02
GNP	0.68	0.36***	0.3***	0.05	–0.04	0.76***	0.26***	0.37***	–0.14**	–0.01	–0.08	–	–0.12*	0.79***
TGW	0.83	–0.06	0.07	–0.12*	–0.06	0.32***	0.04	0.29***	–0.16**	0.17**	0.3***	–0.08	–	0.45***
YPP	0.68	0.35***	0.35***	–0.01	–0.04	0.88***	0.27***	0.44***	–0.19***	0.07	0.06	0.85***	0.37***	–

* Significance at the 0.05 probability level.

**Significance at the 0.01 probability level.

***Significance at the 0.001 probability level.

† H^2, broad-sense heritability.

‡ DTA, days to anthesis; DTM, days to maturity; GFD, grain fill duration; DPW, dry primary panicle weight (g per panicle); DBW, dry vegetative biomass weight (g per plant); HI, harvest index; GT, glume tenacity; GNP, grain number per primary panicle; TGW, 1000-grain weight (g); YPP, grain yield per primary panicle.

§ GFD = DTM – DTA.

¶ DPW, total dry weight of panicle containing grain.

DBW, total dry weight of stalk and leaves.

†† HI = YPP/(DBW + DPW).

‡‡ GT, head mold, and head smut were observational measurements ranked on a nominal scale from 1 to 5.

(USDA–ARS–National Genetic Resources Program, 2014). Although the diversity panel contained genotypes from all 10 intermediate races, only four intermediates (durra–bicolor, durra–caudatum, guinea–caudatum, and kafir–caudatum) contained greater than five accessions and were included in the analysis. In general, there was a large amount of variation for yield components within each individual race, and several races contained outliers (Supplemental Fig. S3). The bicolor race had some of the lowest overall mean values for the three yield components, especially TGW. Of the primary races, caudatum and kafir had higher mean GNP, TGW, and overall YPP. This was not surprising considering caudatum and kafir are the most common races used to develop commercial grain hybrids (House, 1985).

Overall, a large portion of phenotypic variance for grain yield components could be attributed to genotypic effects, indicated by broad-sense heritability estimates (Table 1). Heritability for TGW (H^2 = 0.83) was higher than GNP and YPP (both H^2 = 0.68), and the heritability difference observed between TGW and GNP supported recent findings in wheat (Reynolds et al., 2015). The high heritability for TGW was also consistent with findings within different sorghum recombinant inbred populations (Cisse and Ejeta, 2003; Murray et al., 2008). Height (H^2 = 0.95) as well as DTA (H^2 = 0.90) displayed higher broad-sense heritability than grain yield components. Trait heritability for all 13 measured phenotypes can be found in Table 1.

Trait Correlations

There were only a few statistical year-to-year differences in Pearson pairwise correlations among traits, including yield components (Table 1). Grain yield per primary panicle was significantly correlated with DTM, DPW, DBW, harvest index, glume tenacity, GNP, and TGW across both years. The relationships between YPP and other phenotypes were consistent; given that DTA was the only significant pairwise correlation with YPP in 1 yr and not the other (2013, r = 0.02; 2014, r = 0.37). While plant height was not significantly correlated with grain-yield-related traits, which is consistent with results from previous studies (Ritter et al., 2008; Gambín and Borrás, 2012), DBW was positively correlated with YPP, GNP, and TGW. In general, maturity traits DTA and DTM were positively correlated with GNP and YPP, although much stronger positive correlations were found in 2014 (Table 1), which could be attributed to an earlier planting date. The positive relationships of DTA with GNP and YPP are consistent with previous studies in sorghum when grown in favorable conditions (Dalton, 1967; Hassan and Mohammed, 2015). Head mold and smut occurrence on the grain were positively correlated with TGW, while glume tenacity was negatively correlated with all three yield-related traits,

Table 2. Pearson correlations between grain number per primary panicle and 1000-grain weight using subsets of the top 100, middle 200, and bottom 100 accessions for grain yield per primary panicle in 2013 and 2014.

	Mean YPP†		Correlation	
	2013	2014	2013	2014
	— g per panicle —			
Top 100 grain yield lines	79.9	97.1	–0.66***	–0.62***
Middle 200 grain yield lines	52.1	59.4	–0.69***	–0.54***
Bottom 100 grain yield lines	28.9	31.6	–0.37***	–0.5***

***Significance at the 0.001 probability level.

† YPP, grain yield per primary panicle.

which is likely a consequence of panicle architecture among botanical races (Harlan and de Wet, 1972).

Grain yield per primary panicle had a much stronger positive correlation with GNP than TGW, which is consistent with previous findings in sorghum (Lothrop et al., 1985) and rice (Begum et al., 2015). The overall average correlation of $r = -0.1$ observed across years in sorghum between GNP and TGW implied there is only a minor trade-off that exists between grain number and weight. However, because the distributions for yield components were skewed right (Fig. 1), we also compared Pearson pairwise correlations between GNP and TGW using the top 100, middle 200, and bottom 100 accessions for YPP in each year (Table 2). The trade-off between GNP and TGW was much stronger within all three subsets when compared with the correlations from the entire data sets containing the 390 accessions. In general, the higher grain-yielding accessions displayed a stronger negative association between GNP and TGW, with the one exception being in 2013 where the middle 200 yielding accessions ($r = -0.69$) had a slightly lower correlation than the top 100 accessions ($r = -0.66$).

Association Mapping of Grain Yield Components

Two linear model approaches were used to predict significant marker–trait associations for the grain-yield-related traits. The MLM executed in GAPIT tests association significance for every single SNP across the genome independently with the phenotype. Therefore, the MLM does not take into account significance of additional SNPs when determining single SNP association. The polygenic BSLMM implemented in GEMMA considers both large (sparse) and small effects of all SNPs while also controlling for population structure to identify all loci associated with the phenotype (Zhou et al., 2013). When comparing associations across the two models previously described, many association peaks that were identified with the BSLMM were ranked highly in the MLM results (Table 3). Manhattan plots displaying 2014 YPP results from the MLM and BSLMM association scans were superimposed to highlight the consistency across models (Fig. 2; Supplemental Table S4). However, the majority of MLM associations fell below the determined significance threshold ($p = 10^{-5}$) for the complex yield-related traits. This potential overcorrection with the MLM is likely because botanical race is strongly correlated with both population structure and panicle architecture, resulting in true associations falling below the significance cutoff (Morris et al., 2013b). The multilocus approach of the BSLMM has been previously shown to have increased statistical power required for complex traits, such as grain yield and its components, than single SNP testing (Zhou et al., 2013; Moser et al., 2015). Identifying loci that were associated across single SNP and polygenic models provided corroborating evidence to reduce the discovery of false positive associations.

Pearson correlations using SNP association scores from GWAS were examined to determine the consistency between linear models, reproducibility of GWAS across years, and genetic correlation between traits (Supplemental Table S4). The correlation of SNP associations for each yield component between years was $r = 0.17$ for YPP, $r = 0.16$ for GNP, and $r = 0.19$ for TGW. The greater prevalence of SNPs to have similar association scores for TGW across the 2 yr than GNP is consistent with the higher heritability of TGW that was observed both within and outside this study (Cisse and Ejeta, 2003; Murray et al., 2008). Using 100 SNPs with the lowest *p*-values from the MLM and thus highest significance for each trait, the number of communal SNPs shared across phenotypes was also evaluated to determine degree of colocalization. First, to examine potential confounding phenotypes with the yield-related traits, the top 100 SNPs associated with DTM, DBW, and plant height were each compared with the top 100 SNPs for YPP, GNP, and TGW. There were no shared loci between any of these agronomic traits except in 2013, where two plant height SNPs (S9_57055052 and S9_57064440) and two biomass SNPs (S9_57055197 and S9_57838309) were shared with TGW. These SNPs are near the *Sb-HT9.1* (*Dw1*) locus on chromosome 9 (Brown et al., 2006). Although these SNPs were ranked in the top 100, they were not above the significance threshold in the BSLMM or MLM. We next observed the top 100 ranked SNPs for each yield-related trait to find colocalization. For TGW and GNP, there were one and 59 shared SNPs with YPP in 2013, respectively. Results from the 2014 MLM association scan generated zero SNPs that were shared between TGW and YPP, while there were 18 communal SNPs between GNP and YPP. This higher degree of colocalization between GNP and YPP was expected since these two phenotypes had a much stronger positive correlation (2013, $r = 0.79$; 2014, $r = 0.85$) than what was found between TGW and YPP (2013, $r = 0.41$; 2014, $r = 0.35$). In both 2013 and 2014, there were no shared SNPs between GNP and TGW, revealing no statistically significant colocalization of loci between these two primary grain yield components.

Genome-wide association studies, using multiple years of grain yield component data, revealed 36 statistically significant SNPs associated with YPP, which was between the number of significant associations for GNP ($n = 53$) and TGW ($n = 19$) (Fig. 3). Few SNPs associated

Table 3. List of functional candidate genes in linkage disequilibrium with a single-nucleotide polymorphism (SNP) significantly associated in at least one linear model with a yield-related trait.

Model†	Trait	Year	SNP	Alleles‡	MAF§	Allelic effect¶	BSLMM PIP#	MLM *P*-value††	BSLMM rank	MLM rank	Gene ID	Putative gene description	Start position	End position
													bp	
BSLMM, MLM	GNP	2013	S1_31332309	C/T	0.07	236.99	0.027	4.30×10^{-6}	3	1	Sobic.001G259700	Jasmonate-zim-domain protein 11	31316851	31317973
BSLMM	GNP	2013	S10_57503212	A/C	0.29	–173.67	0.028	2.92×10^{-5}	2	15	Sobic.010G234500	O-methyltransferase family protein	57514467	57516335
BSLMM	GNP	2013	S10_57503212	A/C	0.29	–173.67	0.028	2.92×10^{-5}	2	15	Sobic.010G234400	O-methyltransferase family protein	57499499	57501273
BSLMM	GNP	2013	S3_62444198	A/G	0.49	115.68	0.013	1.15×10^{-4}	15	81	Sobic.003G291800	Hexokinase 1	62453853	62460717
BSLMM	GNP	2013	S6_43346779	T/C	0.13	191.36	0.013	2.03×10^{-5}	18	12	Sobic.006G070000	Alpha/beta-Hydrolases superfamily protein	43327286	43330548
BSLMM	GNP	2013	S6_58207863	T/G	0.13	178.17	0.012	3.22×10^{-5}	20	17	Sobic.006G228000	Cold shock domain protein 1	58189046	58190350
BSLMM	GNP	2013	S6_58207863	T/G	0.13	178.17	0.012	3.22×10^{-5}	20	17	Sobic.006G228100	Xyloglucan endotransglucosylase/hydrolase 9	58198167	58200279
BSLMM	GNP	2014	S1_17488332	G/C	0.42	153.1	0.027	3.81×10^{-5}	5	20	Sobic.001G195100	Cytochrome P450 superfamily protein	17469073	17471980
BSLMM	GNP	2014	S1_17488332	G/C	0.42	153.1	0.027	3.81×10^{-5}	5	20	Sobic.001G195200	Cytochrome P450 superfamily protein	17486672	17488683
BSLMM	GNP	2014	S1_17488332	G/C	0.42	153.1	0.027	3.81×10^{-5}	5	20	Sobic.001G195300	Cytochrome P450, family 89, subfamily A, polypeptide 2	17503931	17508689
BSLMM, MLM	GNP	2014	S3_58618053	C/A	0.48	152.13	0.076	6.01×10^{-6}	1	1	Sobic.003G247000	Glycosyl hydrolase superfamily protein	58614522	58618226
BSLMM	GNP	2014	S7_53393137	C/G	0.32	–132.57	0.016	7.86×10^{-5}	10	48	Sobic.007G130500	Cation efflux family protein	53401780	53404173
BSLMM	TGW	2013	S4_8949701	C/A	0.46	1.03	0.013	2.27×10^{-4}	4	25	Sobic.004G099900	Early nodulin-related	8954244	8958232
BSLMM	TGW	2013	S9_57231144	A/T	0.32	1.26	0.027	4.74×10^{-5}	1	3	Sobic.009G232700	Zinc finger C-x8-C-x5-C-x3-H type family protein	57247619	57250270
BSLMM	TGW	2014	S1_51671938	T/G	0.16	–1.98	0.011	5.84×10^{-5}	12	13	Sobic.001G304700	Methylenetetrahydrofolate reductase family protein	51682368	51685451
BSLMM	TGW	2014	S1_51671938	T/G	0.16	–1.98	0.011	5.84×10^{-5}	12	13	Sobic.001G304500	GRAS family transcription factor (scarecrow)	51665095	51668192
BSLMM	TGW	2014	S1_66818439	A/G	0.08	2.89	0.011	2.70×10^{-5}	11	7	Sobic.001G466100	Glycosyl hydrolase family 10 protein	66796146	66798756
BSLMM	TGW	2014	S1_66818439	A/G	0.08	2.89	0.011	2.70×10^{-5}	11	7	Sobic.001G466900	Glycosyl hydrolase family protein 43	66829024	66833502
BSLMM	TGW	2014	S1_66818439	A/G	0.08	2.89	0.011	2.70×10^{-5}	11	7	Sobic.001G466400	Glycosyl hydrolase family 10 protein	66806878	66809917
BSLMM, MLM	TGW	2014	S2_69688557	T/A	0.06	–3.64	0.071	1.75×10^{-6}	3	2	Sobic.002G327600	Signal transduction histidine kinase, ethylene sensor	69690587	69694133
BSLMM, MLM	TGW	2014	S3_58483216	A/C	0.05	4.55	0.169	2.22×10^{-7}	1	1	Sobic.003G245600	Werner syndrome-like exonuclease	58498071	58498679
BSLMM, MLM	TGW	2014	S3_58483216	A/C	0.05	4.55	0.169	2.22×10^{-7}	1	1	Sobic.003G245700	Werner syndrome-like exonuclease	58500934	58501987
BSLMM	YPP	2013	S2_66605247	A/G	0.49	–2.62	0.01	1.27×10^{-4}	13	45	Sobic.002G286600	O-fucosyltransferase family protein	66609175	66615018
BSLMM	YPP	2013	S3_65357373	C/G	0.17	4.44	0.048	1.19×10^{-5}	1	4	Sobic.003G327900	Nodulin MtN21 /EamA-like transporter family protein	65372822	65375378
BSLMM	YPP	2013	S5_7346373	T/G	0.12	–4.99	0.021	1.13×10^{-5}	2	3	Sobic.005G064900	Cytochrome P450, family 71, subfamily B, polypeptide 2	7350144	7352139
BSLMM	YPP	2014	S2_61521584	C/A	0.31	3.8	0.015	2.01×10^{-4}	5	66	Sobic.002G224200	Heat shock transcription factor B4	61529017	61531174
BSLMM	YPP	2014	S3_52986810	G/A	0.36	–5.28	0.031	1.24×10^{-5}	1	7	Sobic.003G201000	DROUGHT SENSITIVE 1	52995227	52999078
BSLMM	YPP	2014	S6_52573680	G/A	0.46	–3.4	0.011	1.46×10^{-4}	20	48	Sobic.006G157700	Glycosyl hydrolase family protein	52552973	52556867
BSLMM	YPP	2014	S6_52573680	G/A	0.46	–3.4	0.011	1.46×10^{-4}	20	48	Sobic.006G157800	Galacturonosyltransferase-like 3	52559107	52561154
BSLMM, MLM	YPP	2014	S8_47626872	A/G	0.08	7.37	0.014	7.45×10^{-6}	11	5	Sobic.008G126400	Heavy metal transport/detoxification superfamily protein	47625510	47627678

† BSLMM, Bayesian sparse linear mixed model; MLM, mixed linear model.

‡ The minor allele is listed on the right.

§ MAF, minor allele frequency.

¶ The allelic effect is in respect to the minor allele. The effects were estimated from the MLM implemented using the Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al., 2012).

PIP, posterior inclusion probability. Higher values represent greater association significance. All SNPs with a PIP > 0.01 were considered significant.

†† The significance threshold of $p = 10^{-5}$ for the MLM was empirically determined and supported in an outside study (Zhang et al., 2015).

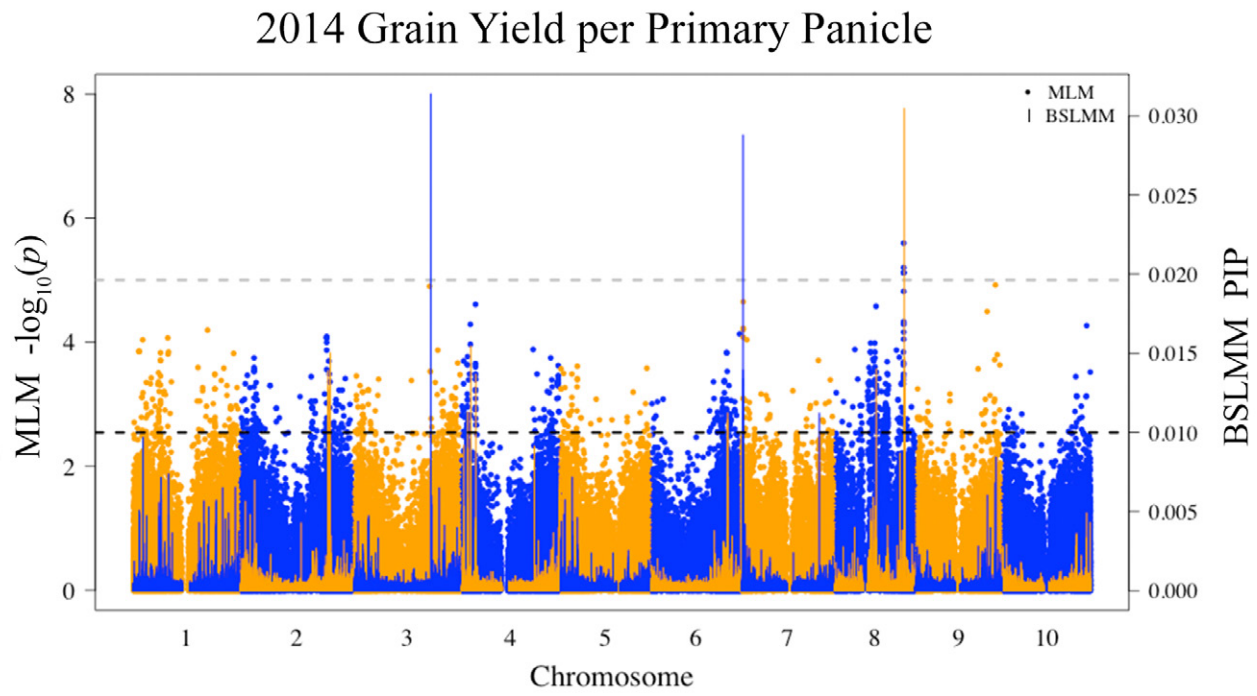

Fig. 2. Superimposition of Manhattan plots displaying genome-wide association study results for grain yield per primary panicle in 2014 generated from the mixed linear model (MLM) in GAPIT and Bayesian sparse linear mixed (BSLMM) model in GEMMA. Physical single-nucleotide polymorphism (SNP) position on the genome is provided on the *x*-axis. Alternating colors distinguish chromosomes. For both linear models, SNPs with higher *y*-axis values have greater associations with grain yield. The light gray horizontal dashed line (*y* = 5) denotes the empirically derived significance threshold for the MLM, while SNPs above the black horizontal dashed line (*y* = 0.01) were considered significant in the BSLMM. PIP, posterior inclusion probability.

with YPP were found to be statistically significant across both years. Several significant associations for YPP did however lie within previously mapped QTL regions for grain yield and related traits. Two association peaks on chromosome 3, one at ~53 Mb and another at ~65.3 Mb, were located within separate grain yield QTL identified by Ritter et al. (2008). A region containing five significant SNPs near 43.4 Mb on chromosome 6 was encompassed within QTL for grain yield and panicle weight (Srinivas et al., 2009). Finally, an association located at 52.4 Mb on chromosome 10 was within a previously identified grain yield QTL (Ritter et al., 2008). Outside of known QTL for grain-yield-related traits, significant SNPs for YPP were commonly found within previously mapped QTL for both stay-green (Subudhi et al., 2000; Kebede et al., 2001; Haussmann et al., 2002; Harris et al., 2007) and sucrose content (Ritter et al., 2008; Shiringani et al., 2010).

Of the three yield-related traits, genome-wide association analysis for GNP produced the most significant marker–trait associations (Supplemental Fig. S3). If relatively small-effect QTL are responsible for phenotypic variation in yield-related traits, this result would be expected, since GNP displayed the greatest amount of trait variation across years (Supplemental Table S3). When combining GNP associations from 2013 and 2014, there were significant loci found on all chromosomes except chromosome 2 and chromosome 9. The strongest association peak was found at ~58.4 Mb on chromosome 3. Within this region, there were 13 SNPs that were above the BSLMM significance threshold (Supplemental Table S5). This locus was within several identified QTL for multiple traits including panicle weight (Shiringani et al., 2010), primary branch length (Brown et al., 2006), plant height (Ritter et al., 2008), and DTM (Srinivas et al., 2009). In addition, the locus at 43.4 Mb on chromosome 6 identified in the GWAS for YPP was also associated with GNP.

Grain weight association scans were most consistent across years among the yield-related traits, yet there were fewer identified loci (n = 17) than GNP and YPP. Grain weight SNPs located 65.3 to 69.7 Mb on chromosome 2 and 36.9 Mb on chromosome 6 found in this study colocalized with previous seed size QTL identified using different biparental mapping populations (Paterson et al., 1995; Feltus et al., 2006; Srinivas et al., 2009). The region on chromosome 2 was also within a recently mapped 100-grain weight QTL (q*GW2*) that was identified by Han et al. (2015) in multiple years. There were multiple TGW association peaks distributed across chromosome 4 that were located within grain weight QTL including one SNP at 33.9 Mb (Paterson et al., 1995; Zhang et al., 2015) and another association near 62.8 Mb that was previously identified by Brown et al. (2006). The strongest associated SNP (S3_58483216) with TGW was within a grain yield QTL previously mapped by Ritter et al. (2008); however, this SNP also colocalized with QTL for plant height (Lin et al., 1995; Ritter et al., 2008). This region could contain a pleiotropic gene or multiple genes in strong LD regulating plant height and grain weight. There is no reason

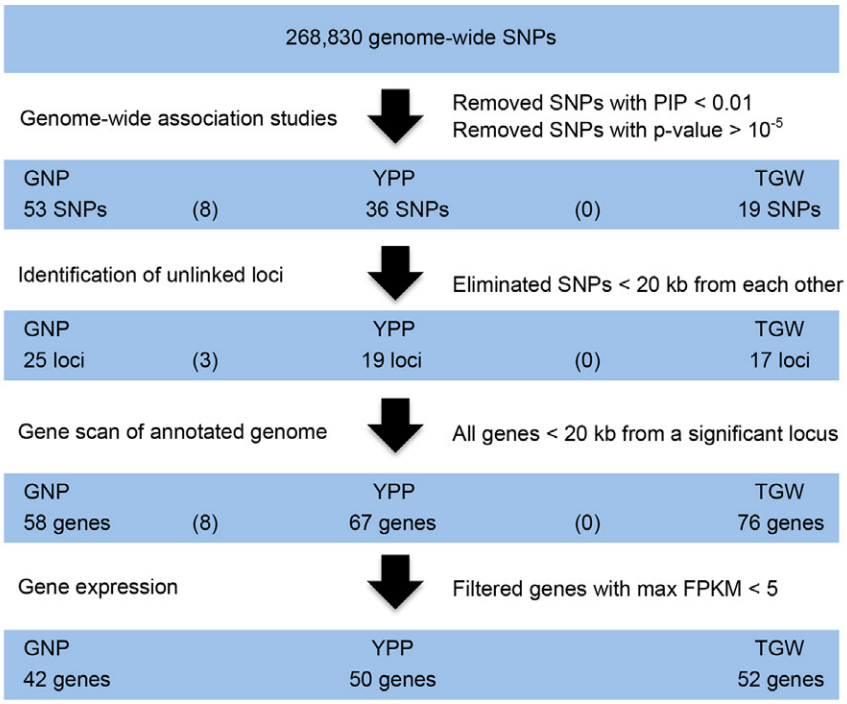

Fig. 3. Schematic of the progression from association mapping to candidate gene selection used to determine candidate genes of highest likelihood to be causing variation in the phenotype. To the left of the arrow is each selection process, while the selection criterion used for each process is located right of the arrow. Numbers in parentheses between traits signify shared single-nucleotide polymorphisms (SNPs) or genes between those two traits. GNP, grain number per primary panicle; YPP, grain yield per primary panicle; TGW, 1000-grain weight.

to believe this colocalization is a result of confounding phenotypes because there was no significant correlation between these two traits in either year (Table 1).

Candidate Gene Identification and Expression

Gene transcripts linked to SNPs (within 20 kb) significantly associated with each yield-related trait were selected by scanning the most recently annotated version of the sorghum genome (Supplemental Table S5) (www.phytozome.net). The distance of 20 kb had previously been found to be the approximate average LD in sorghum based on a wide array of elite and exotic germplasm (Bouchet et al., 2012; Mace et al., 2013). After scanning genes in close proximity to significant SNPs generated from the association scans (BSLMM and MLM) across both years, there were 67 YPP, 58 GNP, and 75 TGW genes identified (Fig. 3). Of these positional candidates, 14% (8/58) of genes for GNP were shared with YPP, while there were no genes shared between TGW and YPP as well as no shared genes between GNP and TGW. Associations for TGW were typically found in more gene-dense regions (4.4 genes <20 kb of each locus) than the other yield-related traits. There were ~3.5 genes <20 kb of each locus associated with YPP and only 2.3 genes surrounding each GNP locus.

The complete set of associated SNPs and nearby functional candidate genes (<20 kb) for YPP as well as GNP and TGW are listed in Table 3. Putative functions of YPP candidates spread across several categories, similar to GNP and TGW. The only putative gene superfamilies that were common to all three yield-related traits were the glycosyl hydrolase family and protein kinase family, although no specific transcript was associated with each trait. Specific to YPP associations, the strongest SNP association (posterior inclusion probability = 0.031) in 2014 was <10 kb from a drought sensitive 1 homolog (*Sobic.003G201000*). In addition, there were a surprisingly high number of candidate genes encoding enzymes involved in cell wall biosynthesis and metabolism (Table 3). These putative transcripts included cellulose synthase-like C5 (*Sobic.002G208200*), O-fucosyltransferase (*Sobic.002G286600*), glycosol hydrolase (*Sobic.006G157700*), and galacturonosyltransferase-like 3 (*Sobic.006G157800*). With YPP in 2014, there were four SNPs on chromosome 8 significantly associated in both linear models within a heavy-metal transporter gene (*Sobic.008G126400*). While two of these nucleotide variants caused synonymous changes, the other two (S8_47626872 and S8_47626978) resulted in missense mutations.

Among the genes located within 20 kb of the associated SNPs for GNP, two homologs, hexokinase 1 (*Sobic.003G291800*) and cytochrome P450 (*Sobic.001G195200*), stood out as potential targets for resequencing based on putative function. Hexokinases are not solely important for their role in glycolysis but have various functions within glucose sensing and signaling pathways, which are critical in regulating plant growth and development (Xiao et al., 2000; Moore et al., 2003; Cho et al., 2009). The putative cytochrome P450 on chromosome 1 that encompassed a significant SNP (S1_17488332) located within its transcript

was also flanked by two additional cytochrome P450 homologs (*Sobic.001G195100* and *Sobic.001G195300*). There was another grain number association located on chromosome 1 that was identified in both linear models near a putative jasmonate ZIM-domain protein (*Sobic.001G259700*).

The strongest SNP associated with TGW across years (S3_58483216) was linked to a gene homologous to Werner syndrome-like exonuclease, a gene that is involved in posttranslational gene silencing in *Arabidopsis* (Glazov et al., 2003). A cluster of SNPs at ~8.9 Mb on chromosome 4 were within LD of several genes, with one transcript encoding a putative early-nodulin related protein (*Sobic.004G099900*). A candidate gene (*Sobic.001G304700*) located at ~51.7 Mb on chromosome 1 was functionally annotated as a member of the methylenetetrahydrofolate reductase family. Another association on chromosome 1 was <4 kb from a GRAS family transcription factor. This candidate gene annotated as a homolog to scarecrow, a key gene in maize involved in development of critical components for C_4 photosynthesis (Slewinski et al., 2012). Perhaps the most intriguing candidate (*Sobic.002G327600*) was an ethylene receptor homolog that was located 2 kb from the third SNP (S2_69688557) of highest significance in 2014 (Fig. 4b; Table 3).

Publically available RNA-seq data from sorghum genotype BTx623 (reference genome) were compiled from various studies (Dugas et al., 2011; Davidson et al., 2012; Makita et al., 2015) to understand where and when candidate genes were expressed. Expression profiles for the complete list of positional candidate genes for yield-related traits are located in Supplemental Table S5. Genes with a maximum FPKM expression of less than five across all tissues included in the expression profile were removed, eliminating 14 to 24% of genes depending on the yield-related trait (Fig. 3). The remaining candidates were individually examined to find gene transcripts with differential expression across tissues. There were 42 candidate genes for GNP remaining after elimination of lowly expressed transcripts. Of these 42 genes, only three had highest expression within early developing inflorescence tissue, while eight genes had maximum expression within the anther, eight within the leaves 20 d after emergence, and 15 genes with highest expression in the developing seed (5–25 DAP). Homologs for the three gene transcripts with maximum expression within the developing inflorescence encoded xyloglucan endotransglucosylase/hydrolase 9 (*Sobic.006G228100*), cation efflux family protein (*Sobic.007G130500*), and an uncharacterized protein (*Sobic.001G259800*). The GNP candidate hexokinase 1 was prominently expressed throughout developing inflorescence tissues of BTx623, with its greatest expression being in the anther. We expected genes contributing to variation in TGW to be most strongly expressed within the developing seed, embryo, and endosperm. In fact, 13 of the 52 TGW positional candidate genes with maximum FPKM > 5 had highest expression within these tissues, including the ethylene receptor and methylenetetrahydrofolate reductase (Fig. 4d) homologs previously mentioned. As an additional step to screen TGW gene candidates, we also observed comparative differential gene expression in developing maize kernels of genotypes with segregating seed sizes (Sekhon et al., 2014). Several genes in LD with a significant SNP for TGW and most prominently expressed in developing sorghum seed also had differential expression between large- and small-seeded maize 'Krug' recombinant inbred lines (Fig. 4f, 5, respectively).

DISCUSSION

Genome-Wide Association Studies

Genome-wide association studies have been useful in detecting novel marker–trait associations for quantitative traits in various plant species (Korte and Farlow, 2013) including sorghum (Brown et al., 2008; Sukumaran et al., 2012; Morris et al., 2013a; Rhodes et al., 2014; Zhang et al., 2015). In this study, there was sufficient statistical power in the GWAS using a multilocus BSLMM to detect significant associations for all yield-related traits. While there were significant SNP associations found with the BSLMM for all yield-related traits, few SNPs were significant across the 2 yr of collected field data, suggesting there were significant environmental influences. This result was not surprising given the contrast in weather between 2013 and 2014 growing seasons in Florence, SC (Supplemental Table S6). To summarize, the growing season (May–Sept.) of 2013 had over 220 mm more rainfall and over 300 fewer growing degree days from planting to mean harvest date than 2014 (www.wunderground.com/history [accessed 1 May 2015]). Genome-wide association studies using the single SNP testing MLM and the multilocus BSLMM did however yield many overlapping loci (Fig. 2; Table 3; Supplemental Table S4) to strengthen the confidence of associations not being false positives, which can be a problem with association mapping in complex traits (Shen and Carlborg, 2013). These results from GWAS conducted in sorghum revealed functional candidate genes for grain-yield components that can be further evaluated to identify causal variants and potentially lead to the discovery of new genes associated with yield-related traits using available gene coexpression network tools (Makita et al., 2015).

Grain Number

Overall, GNP had the highest number of significant SNPs and independent loci generated from GWAS (Fig. 3). Because grain number had the highest amount of variation observed among yield-related traits and is highly quantitative, identification of more loci contributing small effects to the phenotype would be expected. Among the top 100 SNP associations for YPP, 59 and 18 were shared with GNP in 2013 and 2014, respectively. This large overlap in SNP associations between YPP and GNP, which was much greater than the overlap between YPP and TGW, is consistent with the strong positive relationship between

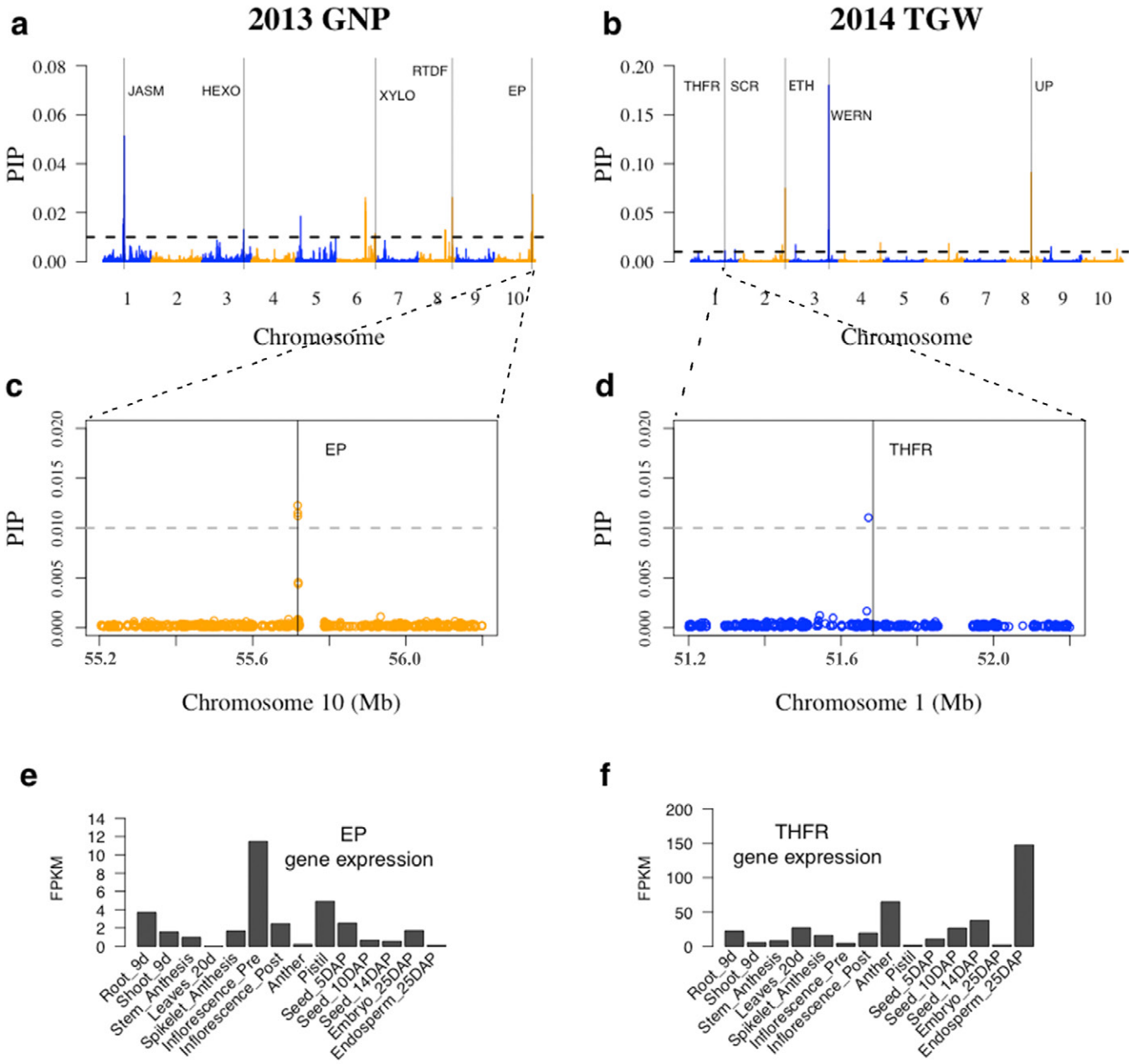

Fig. 4. Grain number and weight candidate genes in linkage disequilibrium with significant single-nucleotide polymorphisms (SNPs) identified from association mapping reveal strong expression within the developing inflorescence and endosperm at 25 d after pollination, respectively. (a, b) Manhattan plots displaying genome-wide association studies for grain number per primary panicle (GNP) in 2013 and 1000-grain weight (TGW) in 2014. The vertical black lines highlight the locations of several a posteriori candidate genes. (c, d) A 1-Mb region of chromosome 10 and chromosome 1 are shown to emphasize the close proximity of the candidate genes to significant SNPs identified from the Bayesian sparse linear mixed (BSLMM). (e, f) Expression profiles of two candidate genes, encoding a putative expressed protein (*Sobic.010G216600*) and a methylenetetrahydrofolate reductase (*Sobic.001G304700*), show elevated expression within the developing inflorescence and grain, respectively. JASM, jasmonate ZIM-domain containing protein; HEXO, hexokinase 1, XYLO, xyloglucan endotransglucosylase/hydrolase 9; RTDF, rotundifolia-like 8; EP, expressed protein; THFR, methylenetetrahydrofolate reductase; SCR, scarecrow (GRAS transcription factor); ETH, ethylene receptor; WERN, Werner syndrome-like exonuclease; UP, uncharacterized protein.

grain yield and number. These results suggest that while determining the complete genetic basis of grain number may be more difficult to accomplish because of the complexity of the trait and genotype × environment interactions, the impact of GNP on maximizing grain yield of the primary panicle appears to be much larger than grain weight. A homolog encoding hexokinase 1 within the glycolytic pathway was tightly linked with a SNP that was associated with GNP in 2013. This candidate supports findings from a transcriptome study in maize that identified grain yield to be associated with a large number of genes involved in glycolysis (Fu et al., 2010). Other

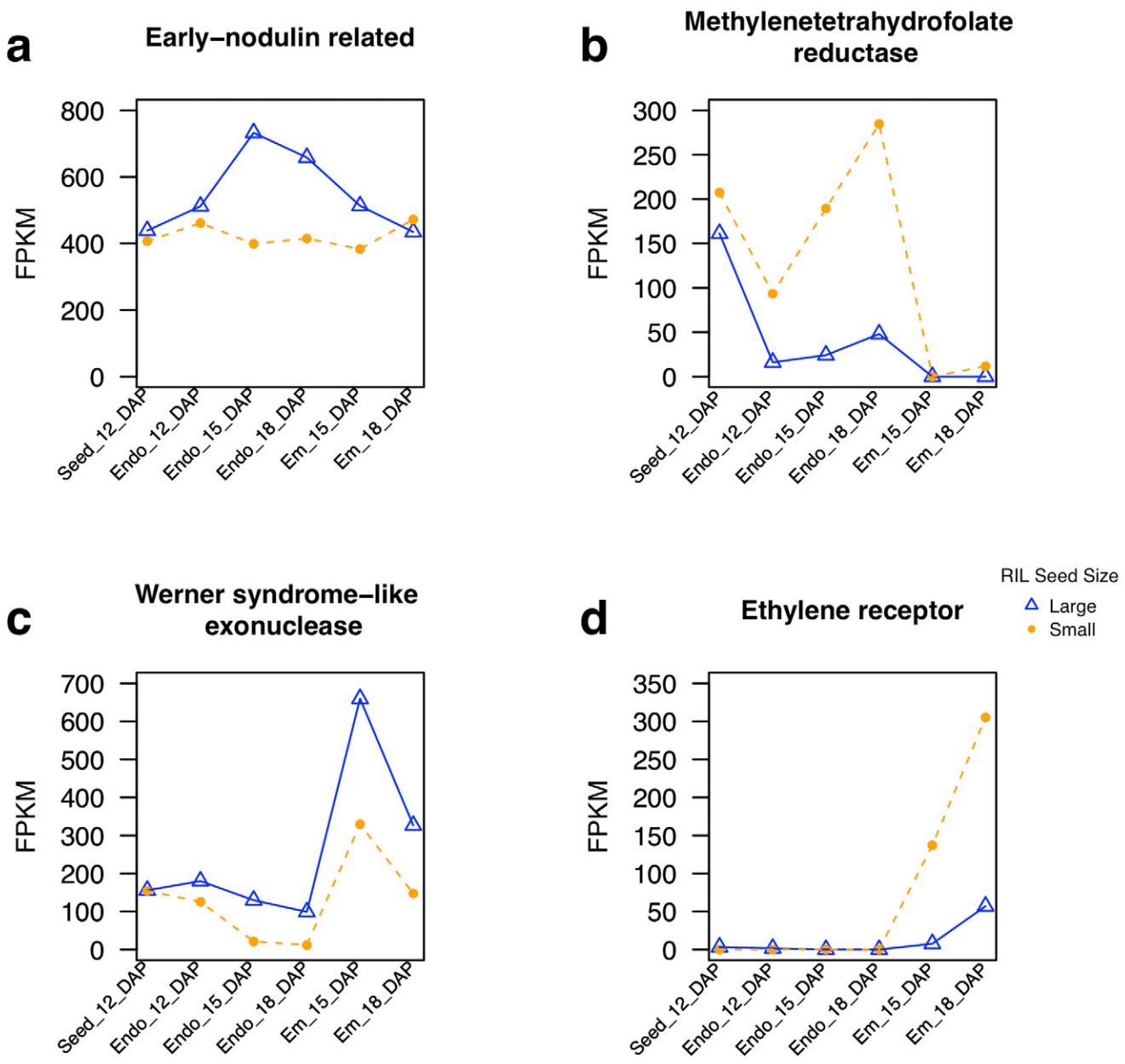

Fig. 5. Maize homologs to sorghum candidate genes for grain weight show differential expression across genotypes segregating for kernel size. Expression profiles were taken from maize 'Krug' large and small kernel recombinant inbred lines (Sekhon et al., 2014). The large and small kernel values presented represent an average from three selected recombinant inbred lines (RILs) within the population. Gene expression was measured in whole seed, endosperm, and embryo tissue at several different time periods postanthesis. (a, c) The majority of homologs to grain weight candidate genes were greater expressed in the large kernel lines, while (b, d) the expression of homologs of methylenetetrahydrofolate reductase and ethylene receptor were actually lower in the developing tissues of large kernel genotypes. FPKM, fragments per kilobase of exon per million reads mapped; DAP, days after pollination; Endo, endosperm; Em, embryo.

positional candidate genes that have functional relevance to grain number include homologs encoding glycosyl hydrolase and a jasmonate ZIM-domain containing protein. The latter protein was found to be involved in the regulation of a number of plant development processes, including senescence (Oh et al., 2013).

Grain Weight

Although grain number appears to have a stronger influence on grain yield, increasing grain weight in sorghum without sacrificing grain number would also increase yield and has other beneficial implications such as improved grain quality and processing (Lee et al., 2002). Because of the high heritability of TGW, grain weight SNP associations were more similar between years (r = 0.19). However, there were fewer associations detected above the significance threshold than the other yield-related traits, which suggests that variation in TGW is controlled by fewer larger-effect loci or true associations failed to reach the significance level. The genomic regions

significantly associated with TGW were much more gene dense, containing nearly double the number of genes within 20 kb than GNP. This greater number of genes surrounding a potential causal gene underlying TGW variation may result in greater linkage drag and thus have undesirable phenotypic effects. With these genes in such strong LD, backcrossing to eliminate this linkage drag may be difficult. Based on sequence homology with *Arabidopsis*, several candidate genes for TGW encode putative scarecrow (GRAS transcription factor), ethylene receptor, methylenetetrahydrofolate reductase, early-nodulin related protein, and multiple glycosyl hydrolases. The signal transduction histidine kinase ethylene receptor (*Sobic.002G327600*) is responsible for regulating the downstream processes controlled by the hormone ethylene, which include fruit ripening and cell death of the endosperm (Chen and Gallie, 2010). This transcript was almost solely expressed within the embryonic tissue of BTx623 at 25 d after pollination (Supplemental Table S5). Several of the additional candidates, including homologs of methylenetetrahydrofolate reductase and early-nodulin related protein, were highly expressed in the developing endosperm (Supplemental Table S5), the storage compartment responsible for the majority of total grain weight. Methylenetetrahydrofolate reductase contributes to lignin production in maize (Tang et al., 2014), but the gene may have multiple roles in cereals, since it is primarily expressed in sorghum endosperm during the grain filling stage (Fig. 4e). While several TGW candidates had greater expression in large kernel genotypes within the maize 'Krug' recombinant inbred population segregating for kernel size (Sekhon et al., 2014), expression of the methylenetetrahydrofolate reductase homolog within the developing endosperm was actually much higher (~sixfold in endosperm_18DAP) in small-kernel lines (Fig. 5). Also, the ethylene receptor homolog (*Sobic.002G327600*) that was predominantly expressed in the developing embryo of sorghum BTx623 was much more expressed (~18-fold in embryo_15DAP) in the small-kernel maize lines. These expression profiles indicate that if these genes are involved in regulating grain weight in cereals, they likely act to limit grain size, which is highly correlated with grain weight (Gnan et al., 2014).

Grain Yield

Strong positive relationships and colocalized association peaks of GNP and, to a lesser extent, TGW, with YPP support the common theory that these two traits are critical in determining final grain yield produced by the primary inflorescence. Association analyses of grain yield revealed associations across the genome including strong peaks located on chromosomes 2, 3, and 8 (Fig. 2). The large number of significant associations for YPP for each year was expected because of the extreme quantitative nature of the trait. The number of independent loci, with SNPs <20 kb apart considered one locus, identified from the GWAS for YPP was $n = 19$, which was in-between grain number ($n = 25$) and weight ($n = 17$) (Fig. 3). Included within LD of the genomic regions associated with YPP were a surprising number of transcripts with putative functions in cell wall biosynthesis and metabolism. These transcripts included a cellulose synthase-like 5 (*Sobic.002G208200*), O-fucosyltransferase (*Sobic.002G286600*), glycosyl hydrolase (*Sobic.006G157700*), and galacturonosyltransferase-like 3 (*Sobic.006G157800*). All of these transcripts were expressed across developing root, shoot, and seed tissues in sorghum BTx623. Another candidate gene, a heavy-metal transporter (*Sobic.008G126400*), contained multiple intragenic SNPs associated with YPP that created missense mutations. Based on the reference genome, one G → C variant changes amino acid 152 from glycine to alanine and the second variant (A → G) changes amino acid 177 from serine to glycine. There was a drought sensitive 1 homolog (*Sobic.003G201000*) in LD with the strongest association found in 2014, the year with the least rainfall during the growing season out of the 2 yr. Interestingly, the public sorghum BTx623 RNA-seq data show *Sobic.003G201000* is most strongly expressed within the immature inflorescence at the time when crop growth rate is critical in determining grain yield of the primary panicle (Ritchie et al., 1998).

Conclusions

This study characterized a large number of diverse sorghum accessions for grain yield components to understand the breadth of natural diversity that exists for these traits and identify potential genes that contribute to this phenotypic variation. Broad-sense heritability calculations revealed that the majority of this variation within the yield-related traits could be genetically manipulated for crop improvement. Based on the negative relationships found between GNP and TGW after categorizing accessions based on overall YPP, the physiological trade-off between primary grain yield components GNP and TGW appears to be strong in sorghum especially within higher yielding accessions. The stronger trade-off observed between GNP and TGW in the top 100 accessions for YPP suggests that this trade-off may be due to limited assimilate availability and allocation required to fill more grains. However, the lack of colocalization observed between significant GNP and TGW loci supports the recent findings from Gnan et al. (2014) in *Arabidopsis* that suggest these major grain yield components are likely under independent genetic control. Targeting these independent GNP and TGW loci for favorable alleles to incorporate into elite germplasm could potentially increase one yield component without decreasing the other, thus increasing total grain yield. Not only elucidating the genetic basis of grain number and weight but also understanding the genetic interactions between these two grain yield components are critical steps to ultimately provide a means for increasing grain yield in sorghum and other important cereal crops.

Acknowledgments

The authors want to thank David Jordan for his helpful comments and suggestions. We wish to thank Kelsey Zielinski and Matthew Lennon for dedicated efforts in data collection and entry. We want to also thank the members of the Clemson Agronomy Club who significantly contributed to postharvest phenotyping. We thank William Bridges for insightful suggestions for implementing the statistical polygenic modeling approach. Computationally intensive analyses were performed on Clemson University's high performance computing cluster, Palmetto. This work was partially funded by the North Carolina Biotechnology Center and United Sorghum Checkoff Program.

References

Adeyanju, A., C. Little, J. Yu, and T. Tesso. 2015. Genome-wide association study on resistance to stalk rot diseases in grain sorghum. G3: Genes, Genomes Genet. 5:1165–1175 doi:10.1534/g3.114.016394

Amelong, A., B.L. Gambín, A.D. Severini, and L. Borrás. 2015. Predicting maize kernel number using QTL information. Field Crops Res. 172:119–131. doi:10.1016/j.fcr.2014.11.014

Austin, D.F., and M. Lee. 1998. Detection of quantitative trait loci for grain yield and yield components in maize across generations in stress and nonstress environments. Crop Sci. 38:1296–1308. doi:10.2135/cropsci1998.0011183X003800050029x

Bates D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48. doi:10.18637/jss.v067.i01

Begum, H., J.E. Spindel, A. Lalusin, T. Borromeo, G. Gregorio, J. Hernandez, P. Virk, B. Collard, and S.R. McCouch. 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). PLoS ONE 10:e0119873. doi:10.1371/journal.pone.0119873

Borrell, A.K., F.R. Bidinger, and K. Sunitha. 1999. Stay-green trait associated with yield in recombinant inbred sorghum lines varying in rate of leaf senescence. Int. Sorg. Mill. Newsl. 40:31–34.

Bouchet, S., D. Pot, M. Deu, J.F. Rami, C. Billot, X. Perrier, R. Rivallan, L. Gardes, L. Xia, P. Wenzl, A. Kilian, and J.C. Glaszmann. 2012. Genetic structure, linkage disequilibrium and signature of selection in sorghum: Lessons from physically anchored DArT markers. PLoS ONE 7:e33470. doi:10.1371/journal.pone.0033470

Brown, P.J., P.E. Klein, E. Bortiri, C.B. Acharya, W.L. Rooney, and S. Kresovich. 2006. Inheritance of inflorescence architecture in sorghum. Theor. Appl. Genet. 113:931–942. doi:10.1007/s00122-006-0352-9

Brown, P.J., W.L. Rooney, C. Franks, and S. Kresovich. 2008. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics 180:629–637. doi:10.1534/genetics.108.092239

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community resources and strategies for association mapping in sorghum. Crop Sci. 48:30–40. doi:10.2135/cropsci2007.02.0080

Chen, J.F., and D.R. Gallie. 2010. Analysis of the functional conservation of ethylene receptors between maize and *Arabidopsis*. Plant Mol. Biol. 74:405–421. doi:10.1007/s11103-010-9686-4

Cho, J., N. Ryoo, J.S. Eom, D.W. Lee, H.B. Kim, S.W. Jeong, Y.H. Lee, Y.K. Kwon, M.H. Cho, S.H. Bhoo, T.R. Hahn, Y.I. Park, I. Hwang, J. Sheen, and J.S. Jeon. 2009. Role of the rice hexokinases *OsHXK5* and *OsHXK6* as glucose sensors. Plant Physiol. 149:745–759. doi:10.1104/pp.108.131227

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, D.M. Ruden, and X. Lu. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso-2; iso-3*. Fly (Austin) 6:1–13. doi:10.4161/fly.19695

Cisse, N., and G. Ejeta. 2003. Genetic variation and relationships among seedling vigor traits in sorghum. Crop Sci. 43:824–828. doi:10.2135/cropsci2003.8240

Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luo, M.J. Kearsey, P. Werner, D. Harrap, C. Tapsell, H. Liu, P.E. Hedley, N. Stein, D. Schulte, B. Steuernagel, D.F. Marshall, W.T.B. Thomas, L. Ramsay, I. Mackay, D.J. Balding, R. Waugh, D.M. O'Sullivan, C. Booer, S. Pike, G. Hamilton, G. Jellis, N. Davies, A. Ross, P. Bury, R. Habgood, S. Klose, D. Vequaud, T. Christerson, J. Brosnan, A. Newton, J. Russell, P. Shaw, R. Bayles, and M. Wang. 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc. Natl. Acad. Sci. USA 107:21611–21616. doi:10.1073/pnas.1010179107

Comeault, A.A., V. Soria-Carrasco, Z. Gompert, T.E. Farkas, C.A. Buerkle, T.L. Parchman, and P. Nosil. 2014. Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. Am. Nat. 183:711–727. doi:10.1086/675497

Dalton, L.G. 1967. A positive regression of yield on maturity in sorghum. Crop Sci. 7:271. doi:10.2135/cropsci1967.0011183X000700030035x

Dahlberg, J., J. Berenji, V. Sikora, and D. Latkovic. 2011. Assessing sorghum [*Sorghum bicolor* (L.) Moench] germplasm for new traits: Food, fuels and unique uses. Maydica 56:85–92.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Davidson, R.M., M. Gowda, G. Moghe, H. Lin, B. Vaillancourt, S.H. Shiu, N. Jiang, and C.R. Buell. 2012. Comparative transcriptomics of three *Poaceae* species reveals patterns of gene expression evolution. Plant J. 71:492–502 doi:10.1111/j.1365-313X.2012.05005.x

Dugas, D.V., M.K. Monaco, A. Olson, R.R. Klein, S. Kumari, D. Ware, and P.E. Klein. 2011. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics 12:514. doi:10.1186/1471-2164-12-514

El Naim, A.M., K.E. Muhammed, P.J. Brown, E.A. Ibrahim, and E.E. Suleiman. 2012. Impact of salinity on seed germination and early seedling growth of three sorghum [*Sorghum bicolor* (L.) Moench.] cultivars. Sci. Technol. 2:16–20. doi:10.5923/j.scit.20120202.03

FAO. 2015. FAOSTAT Statistics Database. Food and Agricultural Organization of the United Nations, Rome, Italy. http://faostat3.fao.org/download/Q/QC/E (accessed 15 Mar. 2015).

Feltus, F.A., G.E. Hart, K.F. Schertz, A.M. Casa, S. Kresovich, S. Abraham, P.E. Klein, P.J. Brown, and A.H. Paterson. 2006. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor. Appl. Genet. 112:1295–1305. doi:10.1007/s00122-006-0232-3

Fu, J., A. Thiemann, T.A. Schrag, A.E. Melchinger, S. Scholten, and M. Frisch. 2010. Dissecting grain yield pathways and their interactions with grain dry matter content by a two-step correlation approach with maize seedling transcriptome. BMC Plant Biol. 10:63. doi:10.1186/1471-2229-10-63

Gambín, B.L., and L. Borrás. 2012. Genotypic diversity in sorghum inbred lines for grain-filling patterns and other related agronomic traits. Crop Pasture Sci. 62:1026–1036. doi:10.1071/CP11051

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346. doi:10.1371/journal.pone.0090346

Glazov, E., K. Phillips, G.J. Budziszewski, F. Meins, and J.Z. Levin. 2003. A gene encoding an RNase D exonuclease-like protein is required for post-transcriptional silencing in *Arabidopsis*. Plant J. 35:342–349. doi:10.1046/j.1365-313X.2003.01810.x

Gnan, S., A. Priest, and P.X. Kover. 2014. The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics 198:1751–1758. doi:10.1534/genetics.114.170746

Griffiths, S., L. Wingen, J. Pietragalla, G. Garcia, A. Hasan, D. Miralles, D.F. Calderini, J.B. Ankleshwaria, M.L. Waite, J. Simmonds, J. Snape, and M. Reynolds. 2015. Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. PLoS ONE. doi:10.1371/journal.pone.0118847

Hamblin, M.T., S.E. Mitchell, G.M. White, J. Gallego, R. Kukatla, R.A. Wing, A.H. Paterson, and S. Kresovich. 2004. Comparative population genetics of the panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483. doi:10.1534/genetics.167.1.471

Han, L., J. Chen, E.S. Mace, Y. Liu, M. Zhu, N. Yuyama, D.R. Jordan, and H. Cai. 2015. Fine mapping of q*GW1*, a major QTL for grain weight in sorghum. Theor. Appl. Genet. 128:1813–1825. doi:10.1007/s00122-015-2549-2

Harlan, J.R., and J.M. de Wet. 1972. A simplified classification of cultivated sorghum. Crop Sci. 12:172–176. doi:10.2135/cropsci1972.0011183X001200020005x

Harris, K., P.K. Subudhi, A. Borrell, D. Jordan, D. Rosenow, H. Nguyen, P. Klein, R. Klein, and J. Mullet. 2007. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J. Exp. Bot. 58:327–338. doi:10.1093/jxb/erl225

Hassan, S.A., and M.I. Mohammed. 2015. Breeding for dual purpose attributes in sorghum: Identification of materials and associations among fodder and grain yield and related traits. J. Plant Breed. Crop Sci. 7:94–100. doi:10.5897/JPBCS2015.0497

Haussmann, B., V. Mahalakshmi, B. Reddy, N. Seetharama, C. Hash, and H. Geiger. 2002. QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor. Appl. Genet. 106:133–142. doi:10.1007/s00122-002-1012-3

House, L.R. 1985. A guide to sorghum breeding. 2nd ed. Int. Crop Res. Inst. for the Semi-Arid Trop. Hyderabad, India.

Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q. Zhang, J. Li, and B. Han. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42:961–967. doi:10.1038/ng.695

Kebede, H., P.K. Subudhi, D.T. Rosenow, and H.T. Nguyen. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor. Appl. Genet. 103:266–276. doi:10.1007/s001220100541

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 9:29. doi:10.1186/1746-4811-9-29

Lee, W.J., J.F. Pedersen, and D.R. Shelton. 2002. Relationship of sorghum kernel size to physiochemical, milling, pasting, and cooking properties. Food Res. Int. 35:643–649. doi:10.1016/S0963-9969(01)00167-3

Lin, Y.R., K.F. Schertz, and A.H. Paterson. 1995. Comparative analysis of QTLs affecting plant height and maturity across the *Poaceae*, in reference to an interspecific sorghum population. Genetics 141:391–411.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:10.1093/bioinformatics/bts444

Lothrop, J.E., R.E. Atkins, and O.S. Smith. 1985. Variability for yield and yield components in IAP1R grain sorghum random-mating population. II. Correlations, estimated gains from selection, and correlated responses to selection. Crop Sci. 25:240–244. doi:10.2135/cropsci1985.0011183X002500020010x

Mace, E.S., S. Tai, E.K. Gilding, Y. Li, P.J. Prentis, L. Bian, B.C. Campbell, W. Hu, D.J. Innes, X. Han, A. Cruickshank, C. Dai, C. Frère, H. Zhang, C.H. Hunt, X. Wang, T. Shatte, M. Wang, Z. Su, J. Li, X. Lin, I.D. Godwin, D.R. Jordan, and J. Wang. 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat. Commun. 4:2320. doi:10.1038/ncomms3320

Makita, Y., S. Shimada, M. Kawashima, T. Kondou-Kuriyama, T. Toyoda, and M. Matsui. 2015. MOROKOSHI: Transcriptome database in *Sorghum bicolor*. Plant Cell Physiol. 56:e6. doi:10.1093/pcp/pcu187

Mason, S.C., D. Kathol, K.M. Eskridge, and T.D. Galusha. 2008. Yield increase has been more rapid for maize than for grain sorghum. Crop Sci. 48:1560–1568. doi:10.2135/cropsci2007.09.0529

Moore, B., L. Zhou, F. Rolland, Q. Hall, W.H. Cheng, Y.X. Liu, I. Hwang, T. Jones, and J. Sheen. 2003. Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. Science 300:332–336. doi:10.1126/science.1080585

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, and S. Kresovich. 2013a. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc. Natl. Acad. Sci. USA 110:453–458. doi:10.1073/pnas.1215985110

Morris, G.P., D.H. Rhodes, Z. Brenton, P. Ramu, V.M. Thayil, S. Deshpande, C.T. Hash, C. Acharya, S.E. Mitchell, E.S. Buckler, J. Yu, and S. Kresovich. 2013b. Dissecting genome-wide association signals for loss-of-function phenotypes in sorghum flavonoid pigmentation traits. G3: Genes, Genomes, Genet. 3:2085–2094. doi:10.1534/g3.113.008417

Moser, G., S.H. Lee, B.J. Hayes, M.E. Goddard, N.R. Wray, and P.M. Visscher. 2015. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. PLoS Genet. 11:e1004969. doi:10.1371/journal.pgen.1004969

Murray, S.C. 2012. Differentiation of grain, sweet, and biomass-producing genotypes in *Saccharinae* species. In: A.H. Paterson, editor, Genomics of the *Saccharinae*. Springer Science and Business Media, New York. p. 479–499.

Murray, S.C., A. Sharma, W.L. Rooney, P.E. Klein, J.E. Mullet, S.E. Mitchell, and S. Kresovich. 2008. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48:2165–2179. doi:10.2135/cropsci2008.01.0016

Neumann, K., B. Kobiljski, S. Denčić, R.K. Varshney, and A. Börner. 2011. Genome-wide association mapping: A case study in bread wheat (*Triticum aestivum* L.). Mol. Breed. 27:37–58. doi:10.1007/s11032-010-9411-7

Oh, Y., I.T. Baldwin, and I. Galis. 2013. A jasmonate ZIM-domain protein NaJAZd regulates floral jasmonic acid levels and counteracts flower abscission in *Nicotiana attenuata* plants. PLoS ONE 8:e57868. doi:10.1371/journal.pone.0057868

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D.G. Peterson, Mehboob-ur-Rahman, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, and D.S. Rokhsar. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. doi:10.1038/nature07723

Paterson, A.H., Y. Lin, Z. Li, K.F. Schertz, J.F. Doebley, S.R.M. Pinson, S. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718. doi:10.1126/science.269.5231.1714

Peterson B.G., P. Carl, K. Boudt, R. Bennett, J. Ulrich, E. Zivot, M. Lestel, K. Balkissoon, and D. Wuertz. 2014. PerformanceAnalytics: Econometric tools for performance and risk analysis. R package version 1.4.3541

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Development Team. 2013. R: A language and environment for statistical computing. http://www.R-project.org (accessed 12 June 2015). R Foundation Stat. Comput., Vienna, Austria.

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98:11479–11484. doi:10.1073/pnas.201394398

Reynolds, M., G. Molero, J. Mollins, and H. Braun. 2015. Proc. of the Int. TRIGO (Wheat) Yield Potential Workshop. 24–26 Mar. 2015. CENEB, CIMMYT, Cd. Obregón, Sonora, Mexico.

Rhodes, D.H., L. Hoffmann, W.L. Rooney, P. Ramu, G.P. Morris, and S. Kresovich. 2014. Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. J. Agric. Food Chem. 62:10916–10927. doi:10.1021/jf503651t

Ritchie, J.T., U. Singh, D.C. Godwin, and W.T. Bowen. 1998. Cereal growth, development and yield. In: G.Y. Tsuji, et al., editors, Understanding options for agricultural production. Systems approaches for sustainable agricultural development. Springer, Netherlands. p. 79–98.

Ritter, K.B., D.R. Jordan, S.C. Chapman, I.D. Godwin, E.S. Mace, and C.L. McIntyre. 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. doi:10.1007/s11032-008-9182-6

Sadras, V.O. 2007. Evolutionary aspects of the trade-off between seed size and number in crops. Field Crops Res. 100:125–138. doi:10.1016/j.fcr.2006.07.004

Sadras, V.O., and R.A. Richards. 2014. Improvement of crop yield in dry environments: Benchmarks, levels of organisation and the role of nitrogen. J. Exp. Bot. 65:1981–1995. doi:10.1093/jxb/eru061

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644. doi:10.1086/502802

Sekhon, R.S., C.N. Hirsch, K.L. Childs, M.W. Breitzman, P. Kell, S. Duvick, E.P. Spalding, C.R. Buell, N. de Leon, and S.M. Kaeppler. 2014. Phenotypic and transcriptional analysis of divergently selected maize populations reveals the role of developmental timing in seed size determination. Plant Physiol. 165:658–669. doi:10.1104/pp.114.235424

Severini, A.D., L. Borrás, M.E. Westgate, and A.G. Cirilo. 2011. Kernel number and kernel weight determination in dent and popcorn maize. Field Crops Res. 120:360–369. doi:10.1016/j.fcr.2010.11.013

Shehzad, T., H. Iwata, and K. Okuno. 2009. Genome-wide association mapping of quantitative traits in sorghum (*Sorghum bicolor* (L.) Moench) by using multiple models. Breed. Sci. 59:217–227. doi:10.1270/jsbbs.59.217

Shen, X., and Ö. Carlborg. 2013. Beware of risk for increased false positive rates in genome-wide association studies for phenotypic variability. Front. Genet. 4:93. doi:10.3389/fgene.2013.00093

Shiringani, A.L., M. Frisch, and W. Friedt. 2010. Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. Theor. Appl. Genet. 121:323–336. doi:10.1007/s00122-010-1312-y

Slewinski, T.L., A.A. Anderson, C. Zhang, and R. Turgeon. 2012. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. Plant Cell Physiol. 53:2030–2037. doi:10.1093/pcp/pcs147

Srinivas, G., K. Satish, R. Madhusudhana, R.N. Reddy, S.M. Mohan, and N. Seetharama. 2009. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. Theor. Appl. Genet. 118:1439–1454. doi:10.1007/s00122-009-0993-6

Subudhi, P.K., D.T. Rosenow, and H.T. Nguyen. 2000. Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): Consistency across genetic backgrounds and environments. Theor. Appl. Genet. 101:733–741. doi:10.1007/s001220051538

Sukumaran, S., S. Dreisigacker, M. Lopes, P. Chavez, and M.P. Reynolds. 2015a. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor. Appl. Genet. 128:353–363. doi:10.1007/s00122-014-2435-3

Sukumaran, S., M.P. Reynolds, M.S. Lopes, and J. Crossa. 2015b. Genome-wide association study for adaptation to agronomic plant density: A component of high yield potential in spring wheat. Crop Sci. 55:2609–2619. doi:10.2135/cropsci2015.03.0139

Sukumaran, S., W. Xiang, S.R. Bean, J.F. Pedersen, S. Kresovich, M.R. Tuinstra, T.T. Tesso, M.T. Hamblin, and J. Yu. 2012. Association mapping for grain quality in a diverse sorghum collection. Plant Gen. 5:126–135. doi:10.3835/plantgenome2012.07.0016

Tang, H.M., S. Liu, S. Hill-Skinner, W. Wu, D. Reed, C.T. Yeh, D. Nettleton, and P.S. Schnable. 2014. The maize brown midrib2 (*bm2*) gene encodes a methylenetetrahydrofolate reductase that contributes to lignin accumulation. Plant J. 77:380–392. doi:10.1111/tpj.12394

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler. 2001. *Dwarf*8 polymorphisms associate with variation in flowering time. Nat. Genet. 28:286–289. doi:10.1038/90135

Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, and L. Pachter. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511–515. doi:10.1038/nbt.1621

Ugarte, C., D.F. Calderini, and G.A. Slafer. 2007. Grain weight and grain number responsiveness to preanthesis temperature in wheat, barley and triticale. Field Crops Res. 100:240–248. doi:10.1016/j.fcr.2006.07.010

USDA–ARS–National Genetic Resources Program. 2014. Germplasm Resources Information Network (GRIN). http://www.ars-grin.gov (accessed 3 Apr. 2015). Nat. Germplasm Resource Lab., Beltsville, MD.

van Oosterom, E.J., and G.L. Hammer. 2008. Determination of grain number in sorghum. Field Crops Res. 108:259–268. doi:10.1016/j.fcr.2008.06.001

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Xiao, W., J. Sheen, and J. Jang. 2000. The role of hexokinase in plant sugar signal transduction and growth and development. Plant Mol. Biol. 44:451–461. doi:10.1023/A:1026501430422

Xu, F., F. Tang, Y. Shao, Y. Chen, C. Tong, and J. Bao. 2014. Genotype × environment interactions for agronomic traits of rice revealed by association mapping. Rice Sci. 21:133–141. doi:10.1016/S1672-6308(13)60179-1

Yang, Z., E.J. van Oosterom, D.R. Jordan, and G.L. Hammer. 2009. Preanthesis ovary development determines genotypic differences in potential kernel weight in sorghum. J. Exp. Bot. 60:139–408. doi:10.1093/jxb/erp019

Zhang, D., H. Guo, C. Kim, T. Lee, J. Li, J. Robertson, X. Wang, Z. Wang, and A.H. Paterson. 2013. CSGRqtl, a comparative quantitative trait locus database for *Saccharinae* grasses. Plant Physiol. 161:594–599. doi:10.1104/pp.112.206870

Zhang, D., J. Li, R.O. Compton, J. Robertson, V.H. Goff, E. Epps, W. Kong, C. Kim, and A.H. Paterson. 2015. Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (*Panicoidae*) and rice (*Oryzoidae*). G3: Genes, Genomes, Genet. 5:1117–1128. doi:10.1534/genetics.115.177170

Zhou, X., P. Carbonetto, and M. Stephens. 2013. Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genet. 9:e1003264. doi:10.1371/journal.pgen.1003264