Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum

Jing Zhao, Maria B. Mantilla Perez, Jieyun Hu, Maria G. Salas Fernandez*

Abstract

Sorghum [Sorghum bicolor (L) Moench], an important grain and forage crop, is receiving significant attention as a lignocellulosic feedstock because of its wateruse efficiency and high biomass yield potential. Because of the advancement of genotyping and sequencing technologies, genome-wide association study (GWAS) has become a routinely used method to investigate the genetic mechanisms underlying natural phenotypic variation. In this study, we performed a GWAS for nine grain and biomass-related plant architecture traits to determine their overall genetic architecture and the specific association of allelic variants in gibberellin (GA) biosynthesis and signaling genes with these phenotypes. A total of 101 single-nucleotide polymorphism (SNP) representative regions were associated with at least one of the nine traits, and two of the significant markers correspond to GA candidate genes, GA2ox5 (Sb09 g028360) and KS3 (Sb06 g028210), affecting plant height and seed number, respectively. The resolution of a previously reported quantitative trait loci (QTL) for leaf angle on chromosome 7 was increased to a 1.67 Mb region containing seven candidate genes with good prospects for further investigation. This study provides new knowledge of the association of GA genes with plant architecture traits and the genomic regions controlling variation in leaf angle, stem circumference, internode number, tiller number, seed number, panicle exsertion, and panicle length. The GA gene affecting seed number variation (KS3, Sb06 g028210) and the genomic region on chromosome 7 associated with variation in leaf angle are also important outcomes of this study and represent the foundation of future validation studies needed to apply this knowledge in breeding programs.

Core Ideas:

- The 101 SNPs were associated with at least one of nine plant architecture traits
- KS3 gene was associated with variation in seed number
- GA2ox5 gene included in a significant region on chromosome 9 controlling plant height
- Novel genomic regions were associated with stem circumference and internode number
- Novel genomic regions were associated with tiller number, panicle exsertion, and length

The increasing interest in biomass production for biofuel use is resulting in a paradigm shift in breeding for plant architecture parameters. The genetic manipulation of these traits can positively affect biomass production (Yuan et al., 2008) as suggested by the high correlations between biomass yield and plant height (Lubberstedt et al., 1997; Salas Fernandez et al., 2009) or leaf angle (Morinaka et al., 2006). Sorghum, the fifth most widely grown cereal crop in the world, is receiving significant attention as one of the most productive annual species for bioenergy production (Rooney et al., 2007) in addition to its well-known value as a grain and forage crop. Therefore, understanding the genetic control...
of plant architecture traits and applying that knowledge in sorghum breeding programs might be instrumental to develop improved germplasm for the incipient lignocellulosic feedstock market as well as contribute to increase yield in grain and forage sorghum breeding programs.

Several linkage mapping studies have been conducted in sorghum to dissect the genetic mechanisms controlling plant architecture. Traits such as plant height, flowering time, and panicle length have been characterized in different segregating populations (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Zou et al., 2012; Nagaraja Reddy et al., 2013). Other traits such as panicle exertion (Klein et al., 2001; Brown et al., 2006; Feltus et al., 2006), tiller number (Paterson et al., 1995; Hart et al., 2001; Feltus et al., 2006; Murray et al., 2008; Shiringani et al., 2010; Alam et al., 2014), and internode number (Srinivas et al., 2009; Zou et al., 2012; Nagaraja Reddy et al., 2013) have also been investigated by several groups. However, limited information is available for leaf angle (Hart et al., 2001), stem circumference (Zou et al., 2012), and seed number (Brown et al., 2006; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). Most QTL identified in these studies were specific to a single population, a finding consistent with the nature of biparental populations, but in some cases, the comparative analysis of multiple independent studies allowed for the identification of a QTL consistent across populations. Panicle length was an example in which a QTL was identified by four groups in the region 58,285,987 to 61,171,968 bp on chromosome 7 (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013).
objectives in this study were to (i) determine the genomic regions controlling plant height, flowering time, tiller number, internode number, panicle exertion, panicle length, seed number, stem circumference, and leaf angle as biomass yield components; (ii) compare our results with previously identified QTL for those traits, if available; and (iii) investigate the association between allelic variation in GA genes and our traits of interest.

Materials and Methods

Germplasm

The panel of 315 sorghum accessions used in this study has been previously described and characterized (Casa et al., 2008; Morris et al., 2013a; Mantilla Perez et al., 2014). It includes 214 conversion lines and 101 historical and elite lines from public breeding programs.

Phenotypic Data

Sorghum lines were planted in a randomized complete block design in three locations in Iowa, with two replications per location, during summer 2010 and 2012. Each plot was a single 3-m-long row with 76-cm row spacing. In 2010, three representative plants per genotype per replication were evaluated in Ames, Crawfordsville, and Lewis, IA, for eight agronomic traits: plant height, flowering time, panicle length, panicle exsertion, stem circumference, internode number, tiller number, and seed number. In 2012, two representative plants per genotype per replication were characterized for leaf angle in Ames, Crawfordsville, and Greenfield, IA. Protocols implemented to measure plant height, flowering time, panicle exertion, stem circumference, and leaf angle have been previously described (Mantilla Perez et al., 2014). Internode number was determined after stripping leaves from the stem. The three panicles per genotype per replication were threshed and manually cleaned to reduce the number of small seeds that could be discarded by air-blowing procedures. Counting was performed using a mechanical seed counter, and the number of seeds was expressed per panicle. The number of tillers was destructively determined by manual separation from the main stalk.

Phenotypic data was analyzed by ANOVA in SAS version 9.2 (SAS Institute, 2008) in which location, genotype, genotype × location interaction, and replication nested within location were treated as random effects. Heritability (H2) for each trait was calculated across environments as follows:

\[ H^2 = \frac{\sigma^2_G}{\sigma^2 + (\sigma^2_{GE}/n) + (\sigma^2_e/(nr))}, \]

where \( \sigma^2_G \) is the genotypic variance, \( \sigma^2_{GE} \) is the genotype × environment interaction variance, \( \sigma^2_e \) is the error variance, \( n \) is the number of environments, and \( r \) is the number of replications. Best linear unbiased prediction (BLUP) was calculated by fitting the following linear model in the R package lme4 for the estimation of breeding values:

\[ Y = (1|\text{Genotype}) + (1|\text{Loc}) + (1|\text{Loc/Rep}) + (1|\text{Genotype:Loc}) \]

where \( Y \) is trait data, 1| indicates random effects, and a colon (:) denotes interaction. Genotype refers to the 315 sorghum accessions, Loc refers to the three environments, and Loc/Rep is replication nested within location. Correlation coefficients were calculated using BLUPs and Pearson’s statistics cor procedure in R software (R Core Development Team, 2013).

Genotypic Data

The association panel was genotyped using GBS methodology (Elshire et al., 2011). The imputed genotypic data has been previously reported (Morris et al., 2013a) and is publicly available at http://www.morrislab.org/data. A total of 136,285 SNPs with minor allele frequency (MAF) >5% and missing data <40% were used in this study. Physical positions of SNPs was determined using Phytozome v1.4. A total of 263 SNPs corresponding to BR genes have been previously investigated for their potential association with plant architecture traits using the same phenotypic data set (Mantilla Perez et al., 2014) and were, thus, excluded from this study.

The Sequenom (SQNM) MassARRAY iPLEX Platform (Gabriel et al., 2009) at the Genomic Technologies Facility (Iowa State University) was used to genotype newly developed markers within GA candidate genes if no GBS data or limited number of markers were available. The identification of sorghum homologous GA genes was performed in silico following a procedure similar to the one described by Mantilla Perez et al. (2014) for BR genes. In summary, previously reported GA protein sequences from model species (Yamaguchi, 2008; Chebotar and Chebotar, 2010; Hedden and Thomas, 2012; Daviere and Achard, 2013) were obtained from the National Center for Biotechnology Information databases and BLASTed against the sorghum genome sequence from Phytozome v1.4 (Paterson et al., 2009) using TBLASTN. Their common domains were predicted using Pfam (Punta et al., 2012). A total of 27 GA candidate genes were identified: 19 from the biosynthesis pathway and eight from the signaling pathway (Supplemental Table S1). Twelve candidate GA genes had no marker or only one SNP from the GBS data set. Therefore, 54 new markers covering these twelve genes were developed by sequencing on ABI 3730 DNA analyzer (Applied Biosystems) and then scored using SQNM (Supplemental Table S2). After markers were filtered for a MAF > 0.05, two genes, Sb03 g035000 (Gibberellin 2-oxidase) and Sb09 g020080 (Gibberellin receptor GIDI), did not have SNPs representing them and were thus excluded from the analysis. In summary, a total of 225 GA-related markers (from both GBS and SQNM data sets) within 25 candidate genes, or 5 kb upstream or downstream from them, were particularly targeted and evaluated as part of the GWAS.

To better capture the variation between 56,624,926 and 61,171,968 bp on chromosome 7, spanning the region of a previously identified QTL for leaf angle (Hart et al., 2014).
et al., 2001; Paterson et al., 1995; Feltus et al., 2006; Mur- et al., 2006); (iii) tiller number (Hart Srinivas et al., 2009; Zou et al., 2012; Nagaraja Reddy et al., 2013). Markers used
panicle length (Hart et al., 2001; Brown et al., 2006; presented in Fig. 1: (i) plant height, flowering time, andlowering studies and the comparison with our results

The same SNP data set was used to calculate the kinship

calculated using the Loiselle algorithm (Loiselle et al.,

Population structure (Q) for this panel has been previ-
ously estimated as five subpopulations using 702 SNPs
with a minimum physical distance of 350 kb (MAF > 0.05 and missing data <15%; Mantilla Perez et al., 2014). The same SNP data set was used to calculate the kinship matrix (K), an estimate of the level of relatedness among individuals, using the Loiselle algorithm (Loiselle et al., 1995) as implemented in SPAGeDi 1.4 (Hardy and Veke-

Both general linear model (GLM, including Q) and mixed linear model (MLM, including Q + K) were used to test phenotype–genotype associations as implemented in TASSEL (Bradbury et al., 2007). False discovery rate, a procedure designed to control false positives as a result of multiple comparisons (Storey and Tibshirani, 2003), was calculated using the q-value package in R software (R Core Development Team, 2013).

The physical positions of previously identified QTL for our traits of interest were extracted from the fol-
lowing studies and the comparison with our results

Results

Significant Phenotypic Variation

The 315 accessions used in this study exhibited a significant variation for all plant architecture traits. As previously reported (Mantilla Perez et al., 2014), genotype, location, and genotype × location interaction were significant sources of variation for plant height, panicle exsertion, pan-
icle length, stem circumference, flowering time, and leaf angle. The analysis of variance also indicated that there was a significant effect of genotype, location, and genotype × location interaction (P < 0.05) for seed number, tiller num-
ber, and internode number. Detail results of the ANOVA
for all traits are presented in Supplemental Table S3. The
BLUPs ranged from 0.05 to 3.3 for tiller number, 6.59 to 13.88 for internode number, and 387 to 3099 for seed num-
ber. All heritability values were high (0.75–0.99) with stem
circumference, tiller number, and seed number being the
only traits with heritability lower than 0.90 (Table 1). The
mean, standard deviation and range of variation for all
traits, calculated using BLUPs, are summarized in Table 1.

The correlation coefficients between all phenotypes
are presented in Table 2. The highest correlation (r = 0.77)
was observed between flowering time and internode num-
ber. Both traits were significantly and positively correlated (P < 0.001) with stem circumference (r = 0.46 and 0.57, respectively) and seed number (r = 0.28 and 0.41, respec-

The correlation coefficients between all phenotypes
are presented in Table 2. The highest correlation (r = 0.77)
was observed between flowering time and internode num-
ber. Both traits were significantly and positively correlated (P < 0.001) with stem circumference (r = 0.46 and 0.57, respectively) and seed number (r = 0.28 and 0.41, respec-

Summary of Genome-Wide Association

Study Results

Only MLM association results are presented in detail and
further compared with previous knowledge of the nine
traits investigated here, since this model greatly reduced
the number of false positive associations when compared
with GLM results, as shown in quantile–quantile plots
(Supplemental Fig. S1). The q-value threshold was set
specifically for each trait (Table 3): (i) q = 0.0000488 to
0.00995 (corresponding P values 2.67 × 10⁻² to 1.69 × 10⁻⁴) for leaf angle, panicle length, tiller number, and plant
height; (ii) q = 0.02539 to 0.06287 (corresponding P values
1.01 × 10⁻⁶ to 7.08 × 10⁻⁴) for panicle exsertion, internode
number, seed number, and flowering time; and (iii) q = 0.1126 (corresponding P value = 4.86 × 10⁻³) for stem
circumference. Based on these thresholds, the expected
number of false positive associations was only one for leaf
angle, tiller number, plant height, flowering time, panicle
length, panicle exsertion, internode number, and seed

Association Analysis

The 315 accessions used in this study exhibited a significant variation for all plant architecture traits. As previously
reported (Mantilla Perez et al., 2014), genotype, location, and genotype × location interaction were significant

4 OF 14
THE PLANT GENOME • JULY 2016 • VOL. 9, NO. 2

2001), we collected additional marker data. DNA from 160 accessions with imputed or missing data in the original
GBS data set for the significant markers S7_58576095
and S7_59049004 were PCR amplified and sequenced
using ABI 3730 to either confirm the imputed data or complete the GBS data set. Missing data was reduced to <2%. The tandem duplication reported as the causal polymorphism of Dw3 (Multani et al., 2003) was geno-
typed in all accessions because there were no GBS SNPs available within this important gene. This was accom-
plished to test the hypothesis that this hormonal-related
gene localized within the target interval is associated
with variation in leaf angle. In total, 683 high quality
markers (MAF > 0.05 and missing data <14%) were geno-
typed. This data set was used to do regional single SNP
association analysis in our attempt to refine the physical
interval previously reported for this leaf angle QTL on
chromosome 7.

Association Analysis

Population structure (Q) for this panel has been previ-
ously estimated as five subpopulations using 702 SNPs
with a minimum physical distance of 350 kb (MAF > 0.05 and missing data <15%; Mantilla Perez et al., 2014). The same SNP data set was used to calculate the kinship
matrix (K), an estimate of the level of relatedness among

Both general linear model (GLM, including Q) and mixed linear model (MLM, including Q + K) were used to
test phenotype–genotype associations as implemented in TASSEL (Bradbury et al., 2007). False discovery rate, a
procedure designed to control false positives as a result of
multiple comparisons (Storey and Tibshirani, 2003), was
calculated using the q-value package in R software (R
Core Development Team, 2013).

The physical positions of previously identified QTL
for our traits of interest were extracted from the fol-
lowing studies and the comparison with our results

presented in Fig. 1: (i) plant height, flowering time, and
panicle length (Hart et al., 2001; Brown et al., 2006;
Srinivas et al., 2009; Zou et al., 2012; Nagaraja Reddy et
al., 2013); (ii) panicle exsertion (Klein et al., 2001; Feltus
et al., 2006; Brown et al., 2006); (iii) tiller number (Hart
et al., 2001; Paterson et al., 1995; Feltus et al., 2006; Mur-
ray et al., 2008; Mace et al., 2009; Shirangi et al., 2010;
Alam et al., 2014); (iv) internode number (Srinivas et al.,
2009; Zou et al., 2012; Nagaraja Reddy et al., 2012; Naga-
raga Reddy et al., 2013; (v) leaf angle (Hart et al., 2001;
(vi) stem circumference Zou et al., 2012); and (vii) for
seed number (Nagaraja Reddy et al., 2013). Markers used
in the aforementioned QTL studies were not specifically
scored in this diversity panel, but their corresponding
physical position was determined based on the sorghum
genome sequence (Phytozome v1.4; Paterson et al., 2009)
and graphically indicated in Fig. 1.
number and four for stem circumference. A total of 101 unique genomic regions were associated with our plant architecture traits of interest, out of 136,320 SNPs tested (Fig. 1; Table 3). Single-nucleotide polymorphisms in close physical proximity and in LD were considered part of the same significantly associated genomic region. Complete information of all significant associations identified using MLM is presented in Supplemental Table S4.
**Table 1. Phenotypic variation of all traits based on best linear unbiased predictions (calculated as genotype performance across environments).**

<table>
<thead>
<tr>
<th>Traits†</th>
<th>Units</th>
<th>Mean ± SD</th>
<th>Range</th>
<th>H²§</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH‡</td>
<td>cm</td>
<td>153.55 ± 58.36</td>
<td>68.56–365.91</td>
<td>0.99</td>
</tr>
<tr>
<td>PL‡</td>
<td>cm</td>
<td>25.70 ± 6.15</td>
<td>9.95–55.84</td>
<td>0.98</td>
</tr>
<tr>
<td>PE‡</td>
<td>cm</td>
<td>10.35 ± 7.63</td>
<td>0.02–39.03</td>
<td>0.95</td>
</tr>
<tr>
<td>SC‡</td>
<td>cm</td>
<td>5.78 ± 0.84</td>
<td>3.42–8.26</td>
<td>0.88</td>
</tr>
<tr>
<td>TN</td>
<td>number</td>
<td>0.65 ± 0.56</td>
<td>0.05–3.73</td>
<td>0.75</td>
</tr>
<tr>
<td>IN</td>
<td>number</td>
<td>10.65 ± 1.34</td>
<td>6.59–13.88</td>
<td>0.92</td>
</tr>
<tr>
<td>FT‡</td>
<td>day</td>
<td>67.46 ± 3.94</td>
<td>54.45–77.16</td>
<td>0.94</td>
</tr>
<tr>
<td>LA‡</td>
<td>degree</td>
<td>50.52 ± 13.31</td>
<td>12.92–88.64</td>
<td>0.95</td>
</tr>
<tr>
<td>SN</td>
<td>number</td>
<td>1567.63 ± 502.9</td>
<td>387–3099</td>
<td>0.88</td>
</tr>
</tbody>
</table>

† PH, plant height; PL, panicle length; PE, panicle exsertion; SC, stem circumference; TN, tiller number; IN, internode number; FT, flowering time; LA, leaf angle; SN, seed number.

§ H², heritability; R² = σ²G/(σ²G + σ²E + σ²G/|n|), where σ²G is the genotypic variance, σ²E is the genotype × environment interaction variance, σ²G/|n| is the error variance, |n| is the number of environments, and |r| is the number of replications.

Considering that the use of a MLM could generate false negative results if causal variants are structured with kinship or between subpopulations, we identified the most significant associations obtained with GLM and compared them with MLM results. The associations uniquely identified by GLM are summarized in Supplemental Table S5.

### Genome-Wide Association Study by Trait

Few markers or genomic regions with major effects were associated with plant height, leaf angle, and flowering time, as shown in Table 3. For plant height, one region on chromosome 9, represented by SNP S9_57236778, explained 29% of the phenotype variation and another region on chromosome 6 (SNP S6_39106643) contributed 20% of the variation. Similar genetic architecture was observed for flowering time, since one significant SNP (S5_51577750) with a large effect (R² = 0.157) was identified on chromosome 5.

On chromosome 7, several SNPs spanning a 1.67 Mb region (between S7_58178513 and S7_59850040), were significantly associated with variation in leaf angle, and one of them (S7_59818811) accounted for more than 15% of the variation (Fig. 2). The level of LD was variable within this region (Fig. 2c), but some markers were in high LD, and thus, further studies are needed to dissect this important chromosomal segment and fully understand the genes or polymorphisms controlling this phenotype.

For some traits, like stem circumference and panicle exsertion, many markers or regions with small effects were identified across several chromosomes (Table 3). Significant SNPs for variation in stem circumference were detected on all chromosomes except 6 and 9 with R² values that ranged from 0.055 to 0.122. Variation in panicle exsertion was explained by markers localized on chromosomes 1, 2, 3, 6, 7, 9, and 10 with small effects (0.076 < R² < 0.118).

Tiller number and internode number were plant architecture traits controlled by several SNPs with relatively large effects located on multiple chromosomes. Markers on chromosomes 1, 3, 4, 8, and 9 explained 8.9 to 14.4% of the variation in tiller number. Genomic regions on chromosomes 1, 2, 5, 6, 8, and 9 controlled between 9.1 and 14.3% of the variation in internode number (Table 3). A small region on chromosome 6 was the most significantly associated with variation in seed number and five markers in that region (S6_57048727, S6_57049108, S6_57049169, S6_57049184, and S6_57049320) correspond to polymorphisms on KS3, a GA biosynthetic gene similar to 15-ketourease synthase (KS) (Supplemental Table S4).

### Same Single-Nucleotide Polymorphisms Associated with Different Traits

In several cases, one or more SNPs were significantly associated with more than one trait, a phenomenon that could be due to pleiotropy or different causal genes in LD (Supplemental Table S4), for example: (i) SNP S9_52325578 associated with variation in both flowering time and internode number; (ii) SNPs S9_57836978 and S9_58005176 explained variation in plant height and panicle exsertion; (iii) SNPs S6_42703814, S6_42726564, and S6_42764790 were significant for both plant height and internode number.

### Table 2. Phenotypic correlations between traits based on best linear unbiased predictions.

<table>
<thead>
<tr>
<th>Traits†</th>
<th>PH</th>
<th>PL</th>
<th>PE</th>
<th>SC</th>
<th>TN</th>
<th>IN</th>
<th>FT</th>
<th>LA</th>
<th>SN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH‡</td>
<td>–</td>
<td>0.15</td>
<td>0.47***</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PL‡</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PE‡</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.11</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SC‡</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.02</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>TN</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IN</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FT‡</td>
<td>0.19</td>
<td>0.01</td>
<td>0.02</td>
<td>0.07</td>
<td>0.57***</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>LA‡</td>
<td>0.15</td>
<td>0.16**</td>
<td>–</td>
<td>–</td>
<td>0.46***</td>
<td>–</td>
<td>0.31***</td>
<td>0.77***</td>
<td>–</td>
</tr>
<tr>
<td>SN</td>
<td>0.30***</td>
<td>–</td>
<td>0.03</td>
<td>0.20**</td>
<td>–</td>
<td>0.06</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.

† PH, plant height; PL, panicle length; PE, panicle exsertion; SC, stem circumference; TN, tiller number; IN, internode number; FT, flowering time; LA, leaf angle; SN, seed number.

‡ Corresponding traits previously reported (Mantilla Perez et al., 2014).
Fig. 2. Increased resolution of a previously identified quantitative trait loci (QTL) for leaf angle. (a) Leaf angle locus (QTL) QLea txs-E was previously mapped between simple-sequence repeat markers Xtxp92 and Xtxp295 on chromosome 7 (Hart et al., 2001). (b) Narrower region of 1.67 Mb on chromosome 7 significantly associated with leaf angle; blue triangle represents the position and association significance of Dw3 (tandem duplication identified as functional polymorphism for plant height was scored as a marker). (c) Linkage disequilibrium plot of markers within 1.67 Mb region. (d) Candidate genes within the 1.67 Mb region are indicated with colored arrows and ordered based on physical map. 1, Sb07 g023360 (ZF-HD homeobox); 2, Sb07 g023380 (Kinase); 3, Sb07 g023575 (AP2 domain); 4, Sb07 g023730 (Dw3) [blue]; 5, Sb07 g023803 (AP2 domain); 6, Sb07 g024110 [similar to SPINDLY]; 7, Sb07 g024740 [similar to SAUR36]; 8, Sb07 g024750 [similar to SAUR36].

Table 3. Summary of significant single-nucleotide polymorphisms (SNPs) for the nine plant architecture traits.

<table>
<thead>
<tr>
<th>Trait†</th>
<th>False discovery rate threshold</th>
<th>Corresponding P-value</th>
<th>Chromosome</th>
<th>R² range</th>
<th>No. of significant SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>q ≤ 0.0000488</td>
<td>P ≤ 2.67 × 10⁻²</td>
<td>6,9</td>
<td>0.1–0.290</td>
<td>6</td>
</tr>
<tr>
<td>LA</td>
<td>q ≤ 0.003982</td>
<td>P ≤ 6.42 × 10⁻²</td>
<td>1,3,6,7,9</td>
<td>0.092–0.159</td>
<td>7</td>
</tr>
<tr>
<td>PL</td>
<td>q ≤ 0.00995</td>
<td>P ≤ 1.69 × 10⁻³</td>
<td>3,5,7,10</td>
<td>0.091–0.139</td>
<td>6</td>
</tr>
<tr>
<td>TN</td>
<td>q ≤ 0.00773</td>
<td>P ≤ 1.77 × 10⁻³</td>
<td>1,3,4,8,9,10</td>
<td>0.089–0.144</td>
<td>17</td>
</tr>
<tr>
<td>FT</td>
<td>q ≤ 0.02539</td>
<td>P ≤ 1.01 × 10⁻³</td>
<td>5,8,9,10</td>
<td>0.123–0.157</td>
<td>5</td>
</tr>
<tr>
<td>IN</td>
<td>q ≤ 0.06287</td>
<td>P ≤ 7.08 × 10⁻³</td>
<td>1,2,5,6,8,9</td>
<td>0.091–0.134</td>
<td>8</td>
</tr>
<tr>
<td>SN</td>
<td>q ≤ 0.04597</td>
<td>P ≤ 8.82 × 10⁻³</td>
<td>6,10</td>
<td>0.08–0.091</td>
<td>2</td>
</tr>
<tr>
<td>PE</td>
<td>q ≤ 0.03566</td>
<td>P ≤ 7.27 × 10⁻³</td>
<td>1,2,3,6,7,9,10</td>
<td>0.076–0.118</td>
<td>14</td>
</tr>
<tr>
<td>SC</td>
<td>q ≤ 0.1126</td>
<td>P ≤ 4.86 × 10⁻³</td>
<td>1,2,3,4,5,7,8,10</td>
<td>0.055–0.122</td>
<td>36</td>
</tr>
</tbody>
</table>

† PH, plant height; LA, leaf angle; PL, panicle length; TN, tiller number; FT, flowering time; IN, internode number; SN, seed number; PE, panicle exsertion; SC, stem circumference.
‡ A single region is defined by SNPs in physical proximity and in LD.
number; and (iv) SNP S7_59261924 was identified associated with panicle length and stem circumference.

**Association of Gibberellin Candidate Genes with Plant Architecture**

Two sorghum GA candidate genes, KS3 and Gibberellin 2-oxidase 5 (GA2ox5), were significantly associated with plant architecture characteristics and explained 9.1% of the variation in seed number and 14.6% of the variation in plant height, respectively. Ten markers were discovered within KS3 and genotyped by SQNM. Seven of them, with high quality data (MAF > 0.05 and missing data <10%), were associated with seed number variation and five were the most significant genome-wide markers for the trait (Fig. 3). GA2ox5 is a strong candidate for further investigation because SNPs within this gene were not only significant in this study but they colocalized with previously identified QTL for plant height.

Association results for all GA candidate genes represented by one SNP with the lowest P-value for each trait, irrespective of the significance threshold, is included in Supplemental Table S6. The genomic position of all 25 GA genes is indicated in Fig. 1 by one representative SNP per gene highlighted in red to simplify the graphic representation.

**Discussion**

**Comparison Between our Genome-Wide Association Study Results and Previous Studies**

Information for 119 previously identified QTL controlling our plant architecture traits of interest was collected and compiled into 73 narrow regions, indicated as shaded areas in Fig. 1. When compared with our GWAS results (using MLM), we observed that (i) there were 10 overlapping regions between our significant SNPs and previously identified QTL; (ii) nine significant SNPs did not fall into any previously reported QTL regions but were relatively close to them (86 kb to 2.5 Mb); (iii) three SNPs out of the nine described in (ii) were in LD with
were not the most significant marker in this chromosomal interval (Fig. 1). In spite of the quantitative genetic evidence proposing that allelic variation in GA2ox5 controlling plant height (Brown et al., 2008; Wang et al., 2012; Morris et al., 2013a; Thurber et al., 2013), Ordonez et al. (2014) concluded that this GA catabolic enzyme could not be Dw1 based on two arguments. First, they indicated that if indeed GA2ox5 was Dw1/dw1, the short phenotype should be accompanied by a bending stem, the observed response to the GA inhibitor uniconazole. However, the authors did not address the fact that GA2ox is encoded by a gene family in sorghum, and thus, because of the functional redundancy, an extreme bending phenotype would be unlikely. Second, expression differences for GA2ox5 were not statistically significant between Dw1 and dw1 lines, but RNA sampling was only performed from elongating internodes in seedlings. It would be pertinent to test expression patterns from multiple developmental stages and tissues considering that, in rice, members of the GA2ox family have differential expression in various tissues (Sakamoto et al., 2004). In general, we can conclude that current knowledge of the significantly associated region on chromosome 9 suggests that GA2ox5 is still an important candidate gene worth studying and validating.

Panicle length and flowering time have also been widely investigated by linkage mapping, and those studies, used as independent validations, provide robustness to our data. A panicle length QTL was consistently identified in the same physical interval between SNPs S7_58285987 and S7_611796858 by four groups using three different biparental populations (Hart et al., 2001; Brown et al., 2006; Srinivas et al., 2009; Nagaraja Reddy et al., 2013). In our study, the region on chromosome 9 associated with variation in flowering time (Multani et al., 2003). Strong association signals that correspond to Dw1 and Dw2 on chromosomes 9 and 6, respectively, were detected in LD mapping studies conducted on a diversity panel (Morris et al., 2013a; Zhang et al., 2015a) and on a minicore collection across multiple environments (Upadhya et al., 2012a). In our study, markers on chromosome 6 and 9 that correspond to Dw1 and Dw2 were also significantly associated with variation in plant height, which validates our results. Additionally, Higgins et al. (2014) identified a SNP associated with variation in plant height that was very close to a GA2ox gene on chromosome 9 (at ~57 Mb). They, and other researchers (Brown et al., 2008; Morris et al., 2013a), suggested that this GA2ox gene could underlie the Dw1 locus. Our results directly confirmed the significant association of GA2ox5 (Sb09 g028360 at 57,265,477 bp on chromosome 9) with variation in plant height and provides additional evidence as the possible underlying gene in Dw1 locus. However, only GA2ox5 SNP, located on the 5’UTR (S9_57266896), was not the most significant marker in this chromosomal interval (Fig. 1). In spite of the quantitative genetic evidence proposing that allelic variation in GA2ox5 is

### Table 4. Linkage disequilibrium (LD) analysis between significant single-nucleotide polymorphism (SNPs) and previously identified quantitative trait loci (QTL). This analysis was performed to determine the novelty of regions identified in this study.

<table>
<thead>
<tr>
<th>Trait†</th>
<th>Chromosome</th>
<th>SNP</th>
<th>Distance‡</th>
<th>LD‡</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT‡¶</td>
<td>8 54138800</td>
<td>S8_49721204</td>
<td>1,517,882</td>
<td>0–0.2</td>
<td>Novel region</td>
</tr>
<tr>
<td>FT‡¶</td>
<td>10 54070775</td>
<td>S10_54425412</td>
<td>81,763</td>
<td>0–0.7</td>
<td>Overlap</td>
</tr>
<tr>
<td>IN#</td>
<td>6 40250000</td>
<td>S6_42703814</td>
<td>2,053,814</td>
<td>0–0.1</td>
<td>Novel region</td>
</tr>
<tr>
<td>IN#</td>
<td>8 52050000</td>
<td>S8_49721204</td>
<td>2,328,796</td>
<td>0–0.1</td>
<td>Novel region</td>
</tr>
<tr>
<td>PE#</td>
<td>6 51050130</td>
<td>S6_52934397</td>
<td>1,095,206</td>
<td>0–0.1</td>
<td>Novel region</td>
</tr>
<tr>
<td>PL††</td>
<td>3 1479755</td>
<td>S3_788281</td>
<td>691,474</td>
<td>0–0.15</td>
<td>Novel region</td>
</tr>
<tr>
<td>SC#</td>
<td>4 55850000</td>
<td>S4_55482999</td>
<td>401,701</td>
<td>0–0.59</td>
<td>Overlap</td>
</tr>
<tr>
<td>SC#</td>
<td>7 59540000</td>
<td>S7_59461924</td>
<td>188,076</td>
<td>0–0.74</td>
<td>Overlap</td>
</tr>
<tr>
<td>SN††</td>
<td>6 50761007</td>
<td>S6_5049320</td>
<td>2,538,644</td>
<td>0–0.17</td>
<td>Novel region</td>
</tr>
</tbody>
</table>

† FT, flowering time; IN, internode number; PE, panicle exertion; PL, panicle length; SC, stem circumference; SN, seed number.

‡ Between newly associated SNPs and previously identified QTL. The superscript following each trait indicates the QTL study.

§ Srinivas et al. (2009).

¶ Hart et al. (2001).

# Zou et al. (2012).

Knowledge about the genetic architecture controlling stem circumference and seed number is limited in sorghum. Two QTL for stem circumference were localized on bin 1535 of chromosome 4 and bin 2461 of chromosome 7 (Zou et al., 2012) under two contrasting conditions: short and long days. No markers were identified in our study within those intervals, but two significant SNPs, S4_55448299 and S7_59261924, were in LD ($r^2 = 0.59$ and $r^2 = 0.74$) with the respective QTL (Table 4). Seed number, an important yield component for grain and forage sorghums, was investigated by two different groups (Brown et al., 2006; Nagaraja Reddy et al., 2013), but only one QTL was discovered on chromosome 6 between markers gspb069 and Xcup12 that explained 5% of the phenotypic variance (Nagaraja Reddy et al., 2013). We have identified five significant markers (S6_57049320, S6_57048727, S6_57049108, S6_57049169, and S6_57049184) in the candidate GA biosynthetic gene KS3 (Sb06 g028210) that explained 9.1% of the phenotypic variation and that were 2.5 Mb from the previously identified QTL but not in LD ($r^2 < 0.17$) (Fig. 3b,c); therefore, these markers belong to a novel genomic region controlling seed number per panicle. KS3 (Sb06 g028210) is an interesting candidate gene for validation and further studies because it is in tandem with another KS-similar gene KSI (Sb06 g028220), whose markers were not significantly associated with variation in seed number (Fig. 3b). Single-nucleotide polymorphisms representing Sb06 g028210 were not originally present in the GBS data set and in spite of the intermediate-to-high level of LD previously reported in sorghum (Hamblin et al., 2005; Morris et al., 2013a), this important genomic region on chromosome 6 would have been undetected if we had not collected additional marker data based on previous knowledge of GA genes and their effects on plant architecture from model species. Additional novel associations were identified for both stem circumference and seed number, as indicated in Table 3 and Supplemental Table S4.

In summary, several significant markers colocalized with previously identified QTL regions for all our target traits except seed number. We recognize that some of the novel regions identified in our study could have been previously discovered, but the comparison with a few QTL studies was not possible because of limitations inherent to the marker technology used at the time, for example, amplified fragment length polymorphism and diversity array technology markers. Both the validated and novel regions reported in this study represent valuable knowledge that could be further investigated and exploited in breeding programs and significantly enrich our understanding of the genetic control of traits with limited previous information in sorghum.

Increasing Resolution of a Previously Identified Quantitative Trait Loci for Leaf Angle

Considering the importance of leaf angle for the genetic improvement of both biomass and grain yield on a per-area basis, we further investigated the significantly associated chromosomal segment that corresponds to the QTL QLea.
txs E identified by Hart et al. (2001), and we were able to reduce the physical region to a 1.67 Mb interval. Leaf angle, defined as the inclination between leaf blade and the vertical culm (Zhao et al., 2010), is mainly determined by the joint connecting the blade with the sheath. Most mutants for leaf angle in model species had been described as having an abnormal division and expansion of adaxial cells in the collar (Nakamura et al., 2009; Zhao et al., 2010) and allelic changes in BR biosynthesis and/or signaling genes (Wada et al., 1981; Yamamuro et al., 2000; Wang et al., 2008; Tanaka et al., 2009). In sorghum, BR genes have also been associated with natural variation in leaf angle (Manilla Perez et al., 2014), but increasing evidence suggests that other phytohormones, such as auxin, ethylene, abscisic acid, and gibberellins, are involved in leaf angle determination as well (Cao and Chen, 1995; Shimada et al., 2006; Xu et al., 2014). Detail mechanistic information for this individual group of hormones on leaf angle is not available since many of them work synergistically with BR (Cohen and Meudt, 1983; Shimada et al., 2006; Hardtke et al., 2007; Song et al., 2009).

After scanning the refined genomic region on chromosome 7 for candidate genes, we detected seven that have been reported in other species as directly or indirectly affecting leaf angle (Fig. 2d; Supplemental Table S7). These seven candidate genes can be divided into two categories: hormone-related and non-hormone-related genes. In the hormone-related category, one gene, Sb07 g023360, was associated with abscisic acid (Xu et al., 2014); two genes, Sb07 g023575 and Sb07 g023803, were related to the ethylene pathway because of their predicted AP2 domains (Jiang et al., 2012); two genes, Sb07 g024740 and Sb07 g024750, were involved in auxin regulation (Kant et al., 2009); and one gene, Sb07 g024110, was related to the GA signaling pathway (Shimada et al., 2006). The only non-hormone-related gene (Sb07 g023380) was a type of CAMK that includes calcium- and calmodulin-dependent protein kinases (Yang and Komatsu, 2000). Current knowledge of the seven novel candidates in this region indicates that Sb07 g023360 is the sorghum orthologue of OsZHD1 gene, a zinc finger homodomain class homeobox transcription factor that plays an important role in rice morphogenesis especially in the formation and distribution of bulbiform cells. Overexpression of OsZHD1 in rice induced abaxially curled and drooping leaves (Xu et al., 2014). Sb07 g023380 is predicted to be the ortholog of a Calcium-dependent protein Kinase (CDPK) involved in the Ca2+–dependent protein phosphorylation leading to brassinolide, thus affecting lamina inclination in rice (Yang and Komatsu, 2000). The rice SPINDLY (SPY) gene (orthologous to Sb07 g024110) encodes an O-linked N-acetylglucosamine transferase considered to be a negative regulator of GA signaling. Transgenic rice plants transformed with an OsSPY RNAi construct showed a larger bending angle at the lamina joint (Shimada et al., 2006). Sb07 g023575 and Sb07 g023803 were selected as candidate genes in this region for future studies because they contain an AP2 domain. An AP2 transcription-factor-like gene affected internode length, leaf shape, and leaf angle in maize because of a rearrangement of leaf epidermal cells and internode parenchyma cells (Jiang et al., 2012). Both Sb07 g024740 and Sb07 g024750 belong to a SAUR family, and they were predicted to be orthologous to SAUR36 genes involved in auxin regulation. Although there is no evidence to demonstrate that SAUR36 functions in altering leaf angle, their family member, SAUR39, has been verified as a key player in changing leaf angle in rice (Kant et al., 2009). Transgenic plants with single copy insertions of SAUR39 developed more horizontal young and old leaves in 10 wk, while wild-type rice plants maintained small leaf angles.

Dw3, a well-known auxin transporter gene with a major effect on sorghum plant height, is also physically located in this important region, and it has recently been reported having pleiotropic effects on leaf angle (Truong et al., 2015). Considering that there were no GBS markers representing this gene in our genotypic data set, we specifically genotyped the association panel for the tandem repeat reported as the causal polymorphism for plant height (Multani et al., 2003). Even though we identified an interval on chromosome 7 controlling leaf angle that coincides with previously reported QTL (Supplemental Fig. S2), our results do not support the hypothesis that Dw3 underlies variation in leaf angle (Fig. 2b). Several experimental differences between our study and Truong et al. (2015) could potentially explain these apparent contradicting results. The angle investigated in our study corresponds to the leaf immediately under the flag leaf, and it was determined at flowering time when vegetative growth ceased. Truong et al. (2015) investigated angles of the third, fourth, and fifth leaf under the leaf whorl at several intervals before and during flowering (based on reported dates). Considering the known function of Dw3 as an auxin transporter (Multani et al., 2003; Brown et al., 2008), the hormonal concentration would decrease from top to bottom in a plant carrying the dw3 allele but would be homogenous throughout the stem and canopy in a Dw3 plant. If Dw3/dw3 is indeed controlling leaf angle, it is logical to conclude that phenotypic differences between Dw3 and dw3 plants would be maximized in lower leaves, in agreement with Truong et al. (2015), but not on upper leaves as suggested by our results. Therefore, we propose that additional genes in this region of chromosome 7 control leaf angle. This hypothesis is also supported by Truong et al. (2015), in which another QTL controlling leaf angle was detected close to Dw3 in a RIL population in which both parents carried the Dw3 allele (R07018 × R07020) (Supplemental Fig. S2). Finally, it should be acknowledged that the apparent contradicting results about the role of Dw3 in leaf angle control could be the consequence of synthetic associations in the region, a phenomenon that has been previously described as the cause of inaccurate association signals (Dickson et al., 2010; Morris et al., 2013b; Higgins et al., 2014).

Conclusions
Our study has generated new knowledge of the genetic mechanisms underlying plant architecture parameters in sorghum, an important grain, forage, and bioenergy crop species. Nine traits, some of them highly correlated, were...
simultaneously investigated and our results compared with genomic regions previously identified by GWAS and linkage mapping studies. In summary, we have (i) discovered new genomic regions for all traits that could be further validated and narrowed down for their application in breeding programs; (ii) confirmed previously identified QTL for some of the traits that provides independent validation to some of our results; (iii) identified a few genomic regions that simultaneously control more than one trait, a phenomenon that could be due to pleiotropy or LD between different causal polymorphisms; and (iv) investigated, in detail, a previously identified QTL for leaf angle that was reduced to a 1.67 Mb region with seven good candidate genes for future validation and cloning experiments. Once our results are confirmed, further studies on pleiotropic effects and modeling data could make these markers useful in breeding programs to design the best sorghum ideotype for each environment and production system, and this knowledge could also be transferred to other important grass species such as rice, maize, and wheat through comparative genomics.

Supplemental Information Available

Supplemental Figure S1. Comparison of quantile–quantile plots for general linear model (GLM) and mixed linear model (MLM) results for each trait. 

Supplemental Figure S2. Comparative analysis of studies reporting QTL on chromosome 7 controlling variation in leaf angle. Detail information about population, leaf number, and experimental conditions is indicated between brackets. Physical position of Dw3 gene relative to QTL intervals is included.

Acknowledgments

The authors appreciate Xiaochen Sun’s contribution for the development of the R code for the estimation of BLUPs. This work was supported by the USDA, National Institute of Food and Agriculture [Project #IOW05298] and by the R.F. Baker Endowment, R.F. Baker Center for Plant Breeding, Iowa State University, Ames, IA 50011, USA.

References


