Genomewide Association for Sugar Yield in Sweet Sorghum

Payne S. Burks, Chris M. Kaiser, Elizabeth M. Hawkins, Patrick J. Brown*

ABSTRACT
Sweet sorghum [Sorghum bicolor (L.) Moench] is characterized by juicy stems with high concentrations of fermentable sugars. The genetic basis of these phenotypes is not understood. A sweet sorghum diversity panel (n = 252), consisting of sweet sorghum cultivars (n = 80) and diverse landraces selected for matching plant height and maturity (n = 172), was genotyped and phenotyped in three environments over 2 yr for sugar-yield-related traits. Sugar yield is the product of juice volume and sugar concentration (Brix). Juice volume but not Brix was significantly higher in sweet sorghum cultivars than in diverse landraces. Most diverse landraces had white midribs, whereas most sweet cultivars had green midribs. The presence of green midribs was strongly correlated with increased sugar yield, juice volume, and moisture but was not significantly correlated with dry biomass. Genomewide association identified a major quantitative trait locus for midrib color, sugar yield, juice volume, and moisture at ~51.8 Mb on chromosome 6, a genomic region previously reported to contain the Dry midrib (D) locus. Midrib color itself was more highly predictive of sugar yield than any significant single nucleotide polymorphism (SNP) in this region, suggesting that either the causal mutation at the D locus is not in high linkage disequilibrium with any SNP in the dataset or that multiple mutations affect midrib color in sorghum. Significant negative correlations between Brix and grain harvest index indicate the existence of trade-offs between grain and sugar yields.

doi: 10.2135/cropsci2015.01.0057
Freely available online through the author-supported open-access option.
© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA
All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.
sugar sorghum is an annual diploid crop that has received relatively little breeding attention, suggesting that rapid genetic improvement may be possible.

Physiological mechanisms underlying sugar accumulation in sweet sorghum are not well understood. Like sugarcane, sweet sorghum accumulates sucrose in stem parenchyma cells, but phloem loading and the timing of sucrose accumulation both appear to differ between the two species. In sugarcane, sucrose transfer in mature internodes is predominantly symplasmic, with sugars loaded into the phloem via plasmodesmata (Tarpley and Vietor, 2007), whereas in sorghum, sucrose transfer is symplasmic in growing internodes but apoplastic in mature internodes (Tarpley and Vietor, 2007). Lingle (1987) reported that sweet sorghum sugar accumulation increases rapidly after panicle emergence, once internode elongation has ceased. Other studies suggest that the timing of sucrose accumulation is similar in sweet sorghum and sugarcane, beginning while stem elongation is still occurring (Hoffmann-Thoma et al., 1996; Gutjahr et al., 2013). Some of these contradictory findings might result from the use of photoperiod-sensitive (PS) versus photoperiod-insensitive (PI) sorghum genotypes. In PI genotypes or PS genotypes grown under short days, internode maturation is generally coincident with flowering, whereas in PS genotypes grown under long days, flowering is delayed relative to internode maturation, and sugar levels at anthesis are considerably higher (Gutjahr et al., 2013). Overall, the data suggest that sugar accumulation is dependent on internode maturation rather than flowering, but for practical purposes sugar accumulation before flowering is limited to the lower internodes of PS genotypes grown under long days.

Genetic mechanisms underlying sugar hyperaccumulation in sweet sorghum are unknown. However, sugar yield and its component phenotypes, Brix and juice volume, appear to be quantitative traits in sorghum with significant environmental and genotype-by-environment effects (Ferraris, 1981; Shiringani et al., 2010). Previous linkage mapping studies in sweet sorghum (Murray et al., 2008a,b; Ritter et al., 2008; Guan et al., 2011; Felderhoff et al., 2012) involved crosses between sweet lines and dwarf grain sorghums, and major QTL for plant height and flowering time were identified in each study. In addition, parents of each mapping population included the sweet sorghum variety ‘Rio’ (Murray et al., 2008a,b; Felderhoff et al., 2012) or a dwarf Rio derivative (Ritter et al., 2008). Two previous association mapping studies in sweet sorghum (Murray et al., 2009; Lv et al., 2013) both included sweet cultivars and dwarf grain sorghum lines, with few nonsweet varieties of comparable height and maturity to the sweet cultivars. In addition, both these studies used small numbers of markers (n = 369 and n = 51, respectively) and individuals (n = 125 and 119, respectively). Given that sugar yield and its components are quantitative traits, their genetic dissection will require both larger single nucleotide polymorphism (SNP) datasets and larger panels of individuals in which height and maturity are not strongly confounded with sugar yield.

The potential for rapid genetic gain in sorghum sugar yield depends on the genetic architecture of its component traits, Brix and juice volume. Juice volume is itself a function of total biomass and biomass juiciness per unit weight. Sugarcane-yield increases in recent decades have been driven almost entirely by increasing cane yields, with no increase in Brix (Jackson, 2005), suggesting that further Brix increases in sugarcane may not be physiologically feasible or may have pleiotropic negative effects on plant growth and biomass. Maximum Brix levels in sweet sorghum are similar to those in sugarcane at ~25% of fresh weight (Murray et al., 2008b), supporting the idea of a physiological plateau. Juice volume in sorghum is reportedly affected by the Dry midrib (D) locus, which conditions allelic differences between dry, white (D-), and juicy, green (dd) midrifs (Rooney, 2000), but a previous linkage mapping study found no effect of the segregating D locus on juiciness (Felderhoff et al., 2012). Sugar yield in sugarcane appears to be sink-limited, because transgenic sugarcane engineered to accumulate isomaltose did not produce any less sucrose, resulting in nearly doubled total sugar yield (Wu and Birch, 2007). Sweet sorghum is unlike sugarcane in that it produces grain, suggesting that competition between carbon sinks might affect sugar yield in sweet sorghum. However, current evidence is equivocal. Murray et al. (2008b) reported a quantitative trait locus (QTL) on chromosome 3 that increased Brix without affecting grain yield, while Felderhoff et al. (2012) reported a QTL in this region that increased Brix but reduced biomass. Murray et al. (2008b) also reported several QTL that exhibited trade-offs for grain and sugar yield, increasing yield of one carbon sink while decreasing yield of the other, but these QTL were environment-specific and mapped near major loci for plant height and flowering time. This study attempted to address these questions more conclusively by designing an association panel that included sweet and nonsweet sorghum accessions with similar height and flowering time, by blocking all experiments by maturity group to minimize neighbor effects from any remaining height and flowering time variation, and by collecting a much larger marker dataset that allowed genomewide association with sugar yield traits.

**MATERIALS AND METHODS**

**Plant Material**

An association panel (n = 252) was assembled to study the genetic basis of sugar accumulation in sorghum (Table S1). This panel included 80 sweet sorghum cultivars from a previous sweet sorghum association study by Murray et al. (2009), 60 landraces from a study by Wang et al. (2009), and 112 landraces...
selected from the exotic progenitors of the sorghum conversion program (EP lines; [Thurber et al., 2013]) on the basis of height and maturity. Height and maturity criteria for the landraces were established on the basis of the distributions of these phenotypes in the sweet cultivars: each landrace had to be at least 2-m tall and had to reach the grain-filling stage of maturity by early October in central Illinois. Sorghum germplasm was obtained from the National Plant Germplasm System, from the USDA–ARS in Lincoln, NE, and from the Institute for Genomic Diversity at Cornell University. Seed increase for earlier-maturing entries was performed during the 2012 summer at the Energy Farm in Urbana, IL, and seed increase for later-maturing entries was performed in Puerto Vallarta, Mexico, during the winter of 2012–2013.

**Experimental Design and Agronomic Practices**

Experimental data was collected from 252 inbreds at the University of Illinois Energy Farm in Urbana, IL, in 2013 (13EF) and 2014 (14EF) and in Dixon Springs, IL, in 2013 (13DS). Genotypes were arranged in a group balanced block design to adjust for maturity differences, which were assessed in a preliminary growout in 2012. Five maturity groups (n = 42–59 accessions each) were established, each containing a mix of sweet sorghum cultivars and landraces (Fig. S1). Maturity groups 1 and 2 contained roughly equal proportions of sweet cultivars and landraces, whereas maturity groups 3 to 5 contained more landraces than sweet cultivars. Each genotype was randomized within its maturity group (block), and each maturity group was replicated twice within each location with randomization of maturity groups within replicates. To eliminate border effects associated with neighboring plots of varying maturities, a two-row border separated each maturity group. Days to anthesis phenotypes measured in 2013 and 2014 generally supported the maturity group assignments established in 2012, with only a few exceptions (Fig. S2). Single 15-foot rows with 30-inch row spacing were used, and plant populations were adjusted to 120,000 individuals/ha. Nitrogen fertilizer was applied as UAN 28% at a rate of 100 lb N/acre preplant for both summers in Urbana, and 90 lb N/acre was side-dressed in the form of UAN 32% for the 2013 summer in Dixon Springs. In addition, 195 lb P$_2$O$_5$/acre and 225 lb K$_2$O/acre of MAP 11–52–0 (monoammonium phosphate) and KCL 0–0–60 (potassium chloride), respectively, were applied during 2013 in Urbana. In Dixon Springs 115 lb P$_2$O$_5$/acre and 180 lb K$_2$O/acre of DAP 18–46–0 (diammonium phosphate) and KCL 0–0–60, respectively, were applied during 2013. Phosphorus and potassium application rates are calculated for 2 yr of uptake at both Urbana and Dixon Springs locations. Plants were rain-fed, and weed control was accomplished by hand-hoeing.

**Phenotyping**

KDC100 barcode scanners (Koamtac) were incorporated into phenotyping protocols previously established by Burks et al. (2013) and Felderhoff et al. (2012). Days to anthesis were recorded for each plot when 50% of individuals were shedding pollen. Midrib color was scored qualitatively at anthesis for the presence of a green (0), intermediate (1), or white (2) midrib by examining the flag leaf. Harvests were scheduled for each maturity group ~30 d after anthesis. Plant height was measured in centimeters shortly before harvest as the distance between the ground and the top of the panicle. At harvest, a 1-m section of row was cut just above the soil surface and total plot weight was measured (kilograms). Panicles were removed to obtain vegetative weight (kilograms), and stalks were processed in a large, trailer-mounted, three-roller sugarcane mill (SCM-APL–3C, Edwards Engineering). Juice volume was measured in milliliters for the entire meter of row, and the concentration of soluble solids within the juice was measured in degrees Brix using a digital refractometer (Atago Pocket Refractometer PAL-1). A bagasse sample was collected and weighed (grams), then dried in a forced air convection oven at 60°C for 3 d and reweighed to calculate the dry matter content of the bagasse. All phenotypes are listed in Table 1.

**Statistical Analysis**

Raw phenotypes were box-cox transformed. Lambda ($\lambda$) values between −8 and 8 in increments of 0.1 were tested for each

---

Table 1. Phenotypes.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Units</th>
<th>Description</th>
<th>Measurement/Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midrib</td>
<td>%</td>
<td>Midrib color</td>
<td>0 = green; 1 = intermediate; 2 = white</td>
</tr>
<tr>
<td>Brix</td>
<td></td>
<td>Soluble solids</td>
<td>Refractometer</td>
</tr>
<tr>
<td>Wet weight</td>
<td>kg</td>
<td>Wet vegetative weight</td>
<td>Stalks from 1 m of row, panicles removed</td>
</tr>
<tr>
<td>Dry weight</td>
<td>kg</td>
<td>Dry vegetative weight</td>
<td>[Wet wt – (juice volume/10000)] × dry matter</td>
</tr>
<tr>
<td>Moisture</td>
<td>%</td>
<td>Total plot moisture</td>
<td>(Wet wt – dry wt)/Wet wt × 100</td>
</tr>
<tr>
<td>Sugar yield</td>
<td>g/m²</td>
<td>Sugar yield per unit area</td>
<td>(Brix × juice yield)/100</td>
</tr>
<tr>
<td>Sugar yield by weight</td>
<td>g/kg</td>
<td>Sugar yield per unit weight</td>
<td>Sugar yield/wet wt</td>
</tr>
<tr>
<td>Juice volume</td>
<td>mL/m²</td>
<td>Juice volume per unit area</td>
<td>Juice from 1 m of row, panicles removed</td>
</tr>
<tr>
<td>Juiciness</td>
<td>mL/kg</td>
<td>Juice volume per unit weight</td>
<td>Juice yield/wet wt</td>
</tr>
<tr>
<td>Height</td>
<td>cm</td>
<td>Plant height</td>
<td>Distance from ground to panicle</td>
</tr>
<tr>
<td>Grain HI</td>
<td>%</td>
<td>Harvest index for grain</td>
<td>(Panicle wt/total wt) × 100</td>
</tr>
<tr>
<td>Anthesis</td>
<td>days</td>
<td>Days to anthesis</td>
<td>50% of plants flowering</td>
</tr>
<tr>
<td>Dry matter†</td>
<td>%</td>
<td>Bagasse dry matter</td>
<td>Bagasse dry wt/bagasse wet wt</td>
</tr>
<tr>
<td>Panicle weight†</td>
<td>Kg</td>
<td>Panicle weight</td>
<td>Total wt – wet wt</td>
</tr>
<tr>
<td>Total weight†</td>
<td></td>
<td>Total weight</td>
<td>Total wt of 1 m of row</td>
</tr>
</tbody>
</table>

†Not used for further analysis.
phenotype using the boxcox() function in the “MASS” package in R (R Core Team, 2014), and the $\lambda$ value with maximum likelihood was used to transform raw phenotypes using the equation:

$$Y_t = (Y^\lambda - 1)/\lambda$$

where $Y_t$ and $Y_i$ are the vectors of raw and transformed phenotypes, respectively.

Best linear unbiased predictors (BLUPs) were predicted for raw and transformed values of each phenotype across all locations using ASREML 3.0 (Gilmour et al., 2009) to fit the following mixed linear model:

$$Y_{ijklm} = \mu + L_i + MG_j + L\times MG_{ji} + L(HD)_{ik} + L(R)_{il} + G_m + L\times G_{im} + HAI_j + e_{ijklm}$$

where $Y$ is a vector of phenotypes, $L$ is the effect of the $i$th location, $MG$ is the effect of the $j$th maturity group, $L\times MG_{ji}$ is the interaction effect between the $i$th location and the $j$th maturity group, $L(HD)_{ik}$ is the effect of the $k$th harvest date nested within the $i$th location, $L(R)_{il}$ is the effect of the $l$th replicate nested within the $i$th location, $G_m$ is the effect of the $m$th genotype, $L\times G_{im}$ is the interaction effect between the $i$th location and the $m$th genotype, HAI is the harvest–anthesis interval covariate, and $e$ is a vector of residuals. All variables were considered random. Broad-sense heritabilities ($H^2$) were calculated using variances for genotype ($G$), genotype by location ($L\times G$), and error using the equation:

$$H^2 = U_{G}/[V_{G}+(V_{L\times G}/n_l)+(V_e/n_e \times n_r)]$$

where $n_e$ and $n_r$ are the number of locations and replicates within each location, respectively. Phenotypic correlations were calculated from BLUPs calculated using both raw and transformed phenotypes. All raw phenotypic data are provided in Table S2.

**Single Nucleotide Polymorphism Library Construction and Genotyping**

Genomic DNA was extracted from etiolated seedlings 3 d after germination using a revised CTAB protocol (Mace et al., 2003) and quantified using Pico-Green (Invitrogen). As described in Thurber et al. (2013), libraries were prepared using a protocol modified from Poland et al. (2012). Briefly, genomic DNAs (~250–500 ng) were double digested with either PstI-HF and BfaI or PstI-HF and HinP1I, ligated to adapters containing one of 384 unique barcodes, pooled, amplified, and subjected to single-end sequencing on an Illumina HiSeq2000 instrument with 100 bp read lengths. Twelve inbreds were not genotyped owing to failure of germination, DNA extraction, or genomic library preparation (Table S1).

**Genotyping-by-Sequencing Bioinformatics and Imputation**

Single nucleotide polymorphism calling from raw Illumina fastq files was performed using the GBS pipeline implemented in TASSEL 4 (Glaubitz et al., 2014). Tags (64 bp sequences) present at least 10 times in the dataset were aligned to version 2.1 of the sorghum genome (www.phytozome.net, accessed 25 June 2015) using bowtie2 (Langmead and Salzberg, 2012) with alignment settings set to “sensitive.” Single nucleotide polymorphism filtering was performed on the basis of the proportion of heterozygous genotypes (~mF 0.9), the minimum taxon coverage (~mTCov 0.05), and the minimum site coverage (~mSCov 0.05). Missing data were imputed in Beagle4 (Browning and Browning, 2007) using a window size of 500 SNPs and an overlap of 100 SNPs. High-heterozygosity SNPs were pruned after imputation using a cutoff of 0.10. The complete genotypic dataset is provided in Table S3.

**Population Structure, Linkage Disequilibrium, and Association Mapping**

The SNP dataset was pruned to remove redundant SNPs, defined as SNPs with identical genotypes that lay within 64 bp of each other, and SNPs with minor allele frequencies <5%, resulting in a final dataset of 42,926 SNPs across the set of lines genotyped and phenotyped ($n = 240$). Principal component analysis was performed using the prcomp() function in R. Linkage disequilibrium ($r^2$) was calculated in R for all pairs of SNPs within 1 Mb of each other. The GAPIT package in R (Lipka et al., 2012) was used to conduct all marker-trait associations using the arguments group.from = nrow(myY) and group.to = nrow(myY) to run a mixed linear model (MLM) rather than the default compressed mixed linear model (CMLM), using arguments Model. selection = T and PCA.total = 5 to select the optimal number of principal components to use as covariates (0–5), and using the default F3D option. Marker-trait association results were pruned for nonindependent associations on the basis of linkage disequilibrium and physical distance. Two SNPs associated with a given trait were considered nonindependent if they were within 25 kb of each other and had $D' > 0.9$ and $r^2 > 0.6$.

**RESULTS AND DISCUSSION**

**Controlling for Neighbor Effects and Developmental Timing**

A preliminary experiment in 2012 used a randomized complete block design in which plots were at risk of significant neighbor effects because of differences in plant height and flowering time. The 2012 experiment also employed a small, single-plant sugarcane mill with which it was only feasible to juice two stalks per plot. Repeatability of wet vegetative weight (wet weight) between reps of the 2012 experiment was extremely low (Table 2), and no data from 2012 were used for further analysis. These problems were addressed in 2013–2014 by obtaining a much larger sugarcane mill that allowed 1 m of stalks from each plot to be processed and through the use of a group balanced block design. Group balanced block designs are well suited for variety trials in which the varieties with similar morphological characteristics are grouped together (Gomez and Gomez, 1984). In this study, genotypes were blocked by maturity group using flowering time data from
Strong population structure was observed in the association panel, with the first two principal components explaining more than 25% of the variance in our dataset of 42,926 SNPs (Fig. 2). However, the classification of accessions as sweet cultivars or landraces was not strongly associated with population structure: sweet cultivars were less diverse than landraces but still covered much of the genetic space represented in the panel. Decay of linkage disequilibrium as assessed by median $r^2$ fell to below 0.2 between 2 and 5 kb and to below 0.1 between 5 and 20 kb in this panel (Fig. S3), in close agreement with previous studies (Hamblin, 2005; Wang et al., 2013).

### Performance of Sweet Cultivars Versus Landraces

All phenotypes were compared between sweet cultivars, which had prior breeding for sugar yield, and landraces, which did not. For ease of interpretation, Fig. 3 shows BLUPs calculated from raw phenotypic values, but analysis of BLUPs from box-cox transformed phenotypes leads to similar conclusions. Some landraces had high sugar yields comparable to the best sweet sorghum cultivars. There were no significant

### Table 2. Wet weight repeatability and correlations with yield traits by location.

<table>
<thead>
<tr>
<th>Location</th>
<th>Blocked by maturity</th>
<th>Stalks processed</th>
<th>Wet weight repeatability</th>
<th>Correlations ($r$) with wet weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sugar yield</td>
</tr>
<tr>
<td>12EF†</td>
<td>No</td>
<td>2 stalks</td>
<td>15.8%</td>
<td>0.42</td>
</tr>
<tr>
<td>13EF</td>
<td>Yes</td>
<td>1 m of row</td>
<td>71.4%</td>
<td>0.79</td>
</tr>
<tr>
<td>13DS</td>
<td>Yes</td>
<td>1 m of row</td>
<td>69.2%</td>
<td>0.82</td>
</tr>
<tr>
<td>14EF</td>
<td>Yes</td>
<td>1 m of row</td>
<td>80.3%</td>
<td>0.88</td>
</tr>
</tbody>
</table>

† Preliminary growout; these phenotypes are not presented further.

Figure 1. Distribution of harvest-anthesis interval (HAI) values by harvest date within each location. The horizontal red dashed line shows the HAI target value of 30 d.
differences between groups for height or dry weight, indicating that differences between groups were not strongly confounded with plant height or biomass variation in this study. All other phenotypes were significantly (two-sided *t* test; *α* < 0.05) different between sweet cultivars and landraces with the exception of Brix. Sweet cultivars had numerically higher Brix values than landraces, but the significance of this *t* test (*p* = 0.008) did not survive a Bonferroni correction for multiple testing. The phenotype with the most dramatic difference between groups was Midrib: almost all sweet cultivars had a green midrib, whereas most diverse landraces had a white midrib. Sweet cultivars had higher moisture, wet weight, juice volume, and sugar yield than landrace genotypes. Sweet cultivars also had higher juiciness and sugar yield by weight, and lower grain harvest index (HI). Though we aimed to control for differences in flowering time as well as plant height between groups, landraces flowered slightly later than sweet cultivars (Fig. 3).

Lower values for grain HI (the proportion of total wet weight accounted for by the panicle) in sweet cultivars could reflect lower grain yields and/or higher wet biomass in sweet cultivars compared with landraces, which would decrease the numerator and increase the denominator of grain HI, respectively. Because sweet cultivars tend to have green midribs and green midribs are correlated with higher moisture, differences in midrib color might confound grain HI values. However, when only green midrib accessions are considered, the difference in grain HI values between sweet cultivars and landraces is significant (*F* = 9.9; *p* = 0.02) and becomes much more significant after controlling for differences in wet weight (*F* = 16.0; *p* = 0.0001), strongly suggesting that sweet cultivars have lower grain yields. A limitation of the current study is that a single time point was sampled, ~30 d after anthesis. Follow-up studies should examine the relative dynamics of grain starch and stem sugar accumulation in sweet and nonsweet sorghum varieties.

**Partitioning Genetic Variance**

Table 3 lists broad-sense heritability estimates (*H*²) for all quantitative traits. In all cases, *H*² estimates were above 80%, indicating that genetic effects were large in this experiment compared with G × E effects and error. Heritability estimates for similar traits in a previous study (Murray et al., 2008b) were lower, possibly because of greater genetic variation in our diverse association panel and/or reduced error variance in our group balanced block design. Moisture and juice volume had the highest
$H^2$ after anthesis and plant height, suggesting the segregation of loci with large effects on these traits. Brix had the lowest $H^2$, and 34% of the variance in Brix was explained by harvest date, underscoring the strong influence of weather fluctuations on plant water status and the Brix phenotype. Grain HI was affected most strongly by location. This might be due in part to the larger HAI values for most harvest dates in the 13DS location (Fig. 1). For each phenotype, inclusion of the HAI covariate improved the model fit even though it explained less than 1% of the phenotypic variance.

**Phenotypic Correlations**

Pearson correlations between BLUPs were calculated for two groups of lines: all genotypes in the panel (AG; $n = 252$; Table 4 below diagonal) and green midrib lines only (GM; $n = 128$; Table 4 above diagonal). A separate correlation analysis was conducted within green midrib lines because so many traits were highly correlated with midrib color in the AG group. Green midrib lines were defined as those that were phenotyped as green (“0”) in at least two of the three replicates phenotyped. In both groups, sugar yield was most highly correlated with juice volume ($r = 0.94$ and 0.92, respectively, in AG and GM groups) and wet weight ($r_{AG} = r_{GM} = 0.88$). Brix, another yield component, had lower correlations with sugar yield ($r_{AG} = 0.53$; $r_{GM} = 0.76$). Correlations of sugar yield with height ($r_{AG} = 0.27$; $r_{GM} = 0.71$) and anthesis ($r_{AG} = 0.13$; $r_{GM} = 0.47$) were both much higher in the GM group. Moisture was positively correlated with sugar yield across all genotypes ($r_{AG} = 0.50$) but negatively correlated with sugar yield in green midrib lines ($r_{GM} = -0.26$). Grain HI

---

Figure 3. Phenotypic differentiation between landraces ($n = 172$) and sweet cultivars ($n = 80$). Phenotypes that differ significantly between groups at $p < 0.05$ are highlighted in color. HI, harvest index.
consistently showed significant negative correlations with all yield–related traits, including sugar yield \( (r_{AG} = -0.63; \ r_{GM} = -0.47) \) and Brix \( (r_{AG} = -0.38; \ r_{GM} = -0.57) \).

Juice volume was a much better predictor of sugar yield than Brix in this study, consistent with previous results (Murray et al., 2008b). Brix has a higher correlation with yield in the GM group, which is more similar to a sweet sorghum breeding population. However, selection for Brix alone (i.e., without measuring juice volume on whole plants or plots) is not recommended because Brix is so highly affected by the overall water status of the plant. Height was highly and positively correlated with Brix in both AG and GM groups \( (r_{AG} = r_{GM} = 0.57) \). This result was especially surprising for the AG group, in which height has low correlations with other sugar-related traits.

Negative correlations were observed between grain HI and all sugar-related traits in both AG and GM groups.

Previous studies have reported that the developing grain is not a significant sink (Lingle, 1987) and that competition between grain starch and stem sugar sinks is low to moderate (Murray et al., 2008b). In contrast, our results suggest that, after controlling for plant height and flowering time, there is a strong direct trade-off between the amount of carbon devoted to starch in the developing grain and sugars stored in the culm. Segregation of major plant height and maturity loci in previous studies might have masked this effect.

**The Sorghum D Locus Is a Major Sugar Yield QTL**

Table 5 shows associations for all phenotypes with \( q \) values \(< 0.10\), filtered to exclude associations judged redundant on the basis of proximity and high linkage disequilibrium (see Methods; Table S4 provides a complete unfiltered
list). For all sugar yield traits the most significant associations were found at ~51.8 Mb on chromosome 6 (Fig. 4), a region previously reported to contain the *Dry Midrib* (*D*) locus (Hart et al., 2001; Srinivas et al., 2009; Xu et al., 2000; Hilson, 1916). The same SNP at 51,801,476 bp was the most significant association for midrib color, sugar yield, sugar yield by weight, moisture, juice volume, and juiciness, strongly suggesting a pleiotropic effect of the *D* locus on all these phenotypes.

Because the QTL detected at the *D* locus did not explain all the variation in midrib color, we also used the midrib color phenotype itself as a “pseudo-SNP” for phenotypic association. For this analysis, average midrib scores were rounded to the nearest whole number. The midrib color pseudo-SNP is much more significant for all traits than the most significant SNP detected at the *D* locus (Table S4), indicating that midrib color itself is a better predictor of sugar yield traits than any SNP in our dataset. There are several possible explanations for this result. First, higher SNP density might be needed in this region to accurately “tag” the causal mutation at the *D* locus. Second, midrib color might be affected by several different mutations in sorghum. Two additional SNPs with significant effects on midrib and moisture were detected, on chromosomes 1 and 5. These results suggest that simple phenotypic selection for green midribs is currently the most effective solution when working with exotic material in sweet sorghum breeding programs.

### Colocalization with QTL Reported in Other Studies

Previous QTL analysis in a (grain × sweet) sorghum RIL population (Felderhoff et al., 2012) concluded that the *D* locus controls midrib color but does not influence moisture or yield, in strong contrast to the present results. However, the authors acknowledge that the white midrib parent used in this study, BTx3197, did not have a “true” pithy phenotype. Therefore, this population might not be segregating for strongly contrasting alleles at the *D* locus. We also cannot exclude the possibility that the effects at the *D* locus are strongly influenced by environment, with greater expressivity in Illinois relative to Texas. Previous studies found Brix associations at ~5.5 Mb on chromosome 1 and at ~57.6 Mb on chromosome 9 (Murray et al., 2009) and at ~55–60 Mb on chromosome 3 (Natoli et al., 2002; Murray et al., 2008b; Felderhoff et al., 2012), but Brix was the lowest-heritability phenotype in this study, and no Brix QTL were detected.

The same SNP at 55.13 Mb on chromosome 7 was significant for both height and dry weight. This is several Mb away from *Dw3*, a major dwarfing locus in sorghum that lies at ~58.6 Mb (Multani et al., 2003; Thurber et al., 2013), and it is not clear whether this association represents *Dw3* or a novel QTL for plant height.

### Table 5. Marker-trait associations (q < 0.1). HI, harvest index.

<table>
<thead>
<tr>
<th>chr</th>
<th>bp</th>
<th>AAF†</th>
<th>Phenotype(s)</th>
<th>p‡</th>
<th>q</th>
<th>Effect§</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48,672,641</td>
<td>0.52</td>
<td>Height</td>
<td>E-06</td>
<td>0.026</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>48,672,661</td>
<td>0.53</td>
<td>Anthesis</td>
<td>E-09</td>
<td>0.0002</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>48,678,973</td>
<td>0.05</td>
<td>Anthesis</td>
<td>E-05</td>
<td>0.083</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>50,321,739</td>
<td>0.06</td>
<td>Anthesis</td>
<td>E-05</td>
<td>0.083</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>63,473,533</td>
<td>0.10</td>
<td>Midrib</td>
<td>E-06</td>
<td>0.018</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>66,043,935</td>
<td>0.05</td>
<td>Midrib</td>
<td>E-05</td>
<td>0.084</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>75,416,421</td>
<td>0.06</td>
<td>Wet weight</td>
<td>E-06</td>
<td>0.072</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>60,717,722</td>
<td>0.07</td>
<td>Anthesis</td>
<td>E-06</td>
<td>0.083</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>8,308,381</td>
<td>0.07</td>
<td>Midrib</td>
<td>E-06</td>
<td>0.018</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>8,494,767</td>
<td>0.40</td>
<td>Midrib</td>
<td>E-06</td>
<td>0.066</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>51,766,559</td>
<td>0.81</td>
<td>Juice volume</td>
<td>E-08</td>
<td>0.001</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>51,801,476</td>
<td>0.76</td>
<td>Juice volume</td>
<td>E-09</td>
<td>0.0002</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>52,348,631</td>
<td>0.25</td>
<td>Midrib</td>
<td>E-06</td>
<td>0.084</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>53,058,104</td>
<td>0.40</td>
<td>Midrib</td>
<td>E-05</td>
<td>0.084</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>53,070,941</td>
<td>0.23</td>
<td>Midrib</td>
<td>E-07</td>
<td>0.004</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>55,130,655</td>
<td>0.16</td>
<td>Height</td>
<td>E-08</td>
<td>0.002</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>55,130,655</td>
<td>0.16</td>
<td>Dry weight</td>
<td>E-06</td>
<td>0.046</td>
<td>–</td>
</tr>
</tbody>
</table>

† Alternate allele frequency (non-Tx623).
‡ Only the exponent of the *p* value in scientific notation is listed.
§ Effect of the alternate allele relative to the reference (Tx623) allele.

A QTL that affects both height and anthesis was detected at 48.67 Mb on chromosome 1. This is over 10 Mb away from the maturity locus *Ma3* (Mace and Jordan, 2010) and likely represents a novel QTL. Multiple significant SNPs in this region are in low linkage disequilibrium with each other and have opposite effects on anthesis, suggesting the existence of an allelic series or closely linked QTL. A significant association with grain HI was detected at 49.94 Mb on chromosome 6, in approximately the same location as a thousand-seed weight QTL reported by Murray et al. (2008b). In both cases, the reference (grain sorghum) allele increases grain weight.

### CONCLUSIONS

The *D* locus is a major sugar yield QTL in sorghum. Most phenotypic differences in sugar yield traits between sweet sorghum cultivars and landraces can be explained in part by differences in midrib color. However, our current genetic model does not explain all the variance in midrib color in our dataset: 49% of the accessions in our study have green midribs, but only 24% of the accessions in our study carry the reference (green) allele at the most significant *D*-linked SNP. Phenotypic selection for green

---

A QTL that affects both height and anthesis was detected at 48.67 Mb on chromosome 1. This is over 10 Mb away from the maturity locus *Ma3* (Mace and Jordan, 2010) and likely represents a novel QTL. Multiple significant SNPs in this region are in low linkage disequilibrium with each other and have opposite effects on anthesis, suggesting the existence of an allelic series or closely linked QTL. A significant association with grain HI was detected at 49.94 Mb on chromosome 6, in approximately the same location as a thousand-seed weight QTL reported by Murray et al. (2008b). In both cases, the reference (grain sorghum) allele increases grain weight.

### CONCLUSIONS

The *D* locus is a major sugar yield QTL in sorghum. Most phenotypic differences in sugar yield traits between sweet sorghum cultivars and landraces can be explained in part by differences in midrib color. However, our current genetic model does not explain all the variance in midrib color in our dataset: 49% of the accessions in our study have green midribs, but only 24% of the accessions in our study carry the reference (green) allele at the most significant *D*-linked SNP. Phenotypic selection for green
midribs is expected to confer higher juice and sugar yields when breeding with diverse material. Selection for Brix alone is not recommended since it is highly influenced by water status. A large number of stalks (>>2) should be juiced from each plot to obtain adequate repeatabilities. The results reported here would be greatly bolstered by replication in additional environments, preferably in the southern and southeastern United States where sweet sorghum could fit into existing, sugarcane-based sugar production schemes. Future studies should assess the potential of genomic selection to accelerate genetic gain for sugar yield in elite sweet sorghum breeding programs.

Acknowledgments
This research was funded by the USDA National Institute of Food Agriculture–Institute of Bioenergy, Climate and Environment, competitive grant award no. 2010–67009–21665, and by a Monsanto Foundation fellowship to PSB.

References
Hilson, G.R. 1916. On the inheritance of certain stem characters in
sorghum. Agric. J. India 11:150.