THE ROLE OF STALK TRAITS IN INFLUENCING SORGHUM JUICE YIELD

A Dissertation

by

GERALDO AFONSO DE CARVALHO JUNIOR

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	William L. Rooney
Committee Members,	John Mullet
	Patricia Klein
	Seth C. Murray
Head of Department,	David D. Baltensperger

December 2016

Major Subject: Plant Breeding

ABSTRACT

Sweet sorghum is an outstanding feedstock choice for bioethanol production, but the gap between theoretical and commercial ethanol yields must be reduced to improve economic viability. Extractable juice yield is a primary limiting factor for higher ethanol yield. Juice yield is genetically complex and laborious to measure with current phenotyping techniques. Therefore, dissecting the genetic basis underlying juice yield and alternative approaches to measure the trait during selection are needed. Because stem properties directly influence sorghum juice yield, the major objectives of this study were to assess the ability to predict juice yield using stem traits and to map quantitative trait loci controlling such characters by conducting two independent and complimentary studies. Coefficient path analysis showed that stalk weight and stalk volume had the greatest total effect on juice yield, followed by stalk diameter, stalk length and plant height. The direct effects for stalk diameter on stalk volume and juice yield were positive and approximately two-fold greater than the direct effect for plant height on the same traits. Equation modeling demonstrated that juice yield was satisfactorily predicted by jointly assessing stalk weight and stalk moisture or stalk diameter, plant height, and stalk juiciness. More importantly, these traits were moderately to highly repeatable within and across environments. One constitutive QTL for stem pithiness was mapped on chromosome 6 that co-localized with a stem moisture QTL recently mapped in sorghum. Moreover, two other adaptive QTLs were found on chromosomes 7 and 9 for the same

trait, the presence of which appeared to be influenced by water availability. A consistent pleiotropic QTL affecting stem moisture, bagasse moisture, and stem diameter was detected on chromosome 1. Results from this research suggest that the genetic control of juice yield component stem-traits is mostly governed by adaptive QTLs, whose presence and magnitude vary greatly between experiments. Overall, the results indicated that phenotypic prediction models and marker-assisted selection can be used to accelerate the selection process of juice yield.

DEDICATION

This dissertation is dedicated to my mother and father, Roselia e Geraldo, who taught me the values of education, humility, and honesty. I also dedicate this to my wonderful wife, Thavilla, who always encouraged me to pursue my dreams, even by having to postpone hers sometimes. Remember: we never give our dreams up. I love you!

ACKNOWLEDGEMENTS

I would like to thank my committee members for their guidance and support throughout the course of this research. Individually I thank my committee chair, Dr. Rooney, first for letting me be a part of his lab, for teaching me sorghum breeding and beyond, and for the guidance during my entire Ph.D. To my committee members: Dr. Klein, Dr. Murray, and Dr. Mullet, thanks for helping me analyze and/or providing genotypic data to my projects, for the quick response to all my questions, and for the helpful inputs throughout the entire course of my research. Thank all of you for serving on my committee.

I would like to express special thanks to Delroy Collins, Steve Labar, Leo Hoffmann Jr., Vickie Horn, Natalie Patterson, Jaemin Cho, Ace Pugh, Bethany Andrews, Brian Pfeiffer, David Horne, Dustin Herb, Francisco Gomes, John Gill, Luke Vacek, Lloyd Mbulwe, Nicholas Boerman, and all student workers involved on my projects. Honestly, without your help my projects would probably never have been completed.

I also would like to thank my father in law, Helio Silverio, my uncle, Crisnio Rosa, and our family friend, Marcelo Ramos, for providing the initial financial support I needed to approve my visa paperwork. Without your help my Ph.D. would be still a dream. Finally, I thank my entire family for understanding my absence in difficult times and for your unconditional support, to my wife for the love and patience, and to my friends for making this journey unique and enjoyable.

CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of Professor William L. Rooney and Seth C. Murray of the Department of Soil and Crop Sciences, Professor Patricia Klein of the Department of Horticultural Sciences, and Professor John Mullet of the Department of Biochemistry and Biophysics. The genotyping procedure described in Chapter III as well as the marker data analyzed for this chapter was conducted and provided by Professor Patricia Klein. All other work conducted for the dissertation was completed by the student independently.

Graduate study was supported by a fellowship from the Brazilian National Council for Scientific and Technological Development (CNPq) and a dissertation research fellowship from the Tom Slick Foundation at Texas A & M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Recent volatility in oil prices and concerns about environmental issues in past decades have sparked interest in alternative fuel sources worldwide. Henceforth, plant-based ethanol production systems are now firmly established in both Brazil and the United States. Although the systems are evolving, continued constraints such as food security (food vs. fuel), climate change, increasing global energy demand, and energy supply vulnerability (Asif and Muneer, 2007, Siirola, 2014) emphasize the need to develop more efficient processes. One of the most logical approaches is the integration of alternative crops into established plant-based ethanol factories. This approach temporally extends the effective use of industrial infrastructure and allows new technology to develop without the initial risk of significant industrial capital. Accordingly, several crops are being tested for bioethanol production; however, finding and designing a suitable crop is complex.

Among potential alternative energy crops, sorghum is a unique choice for several reasons. First, the crop can be designed for different energy feedstock types. Currently, the starch in the grain is used as feedstock for ethanol production in the U.S. For sweet sorghum, the readily fermentable sugars in the succulent stem could be processed with the same infrastructure of sugar cane mills; likewise, its biomass residuals can serve as a raw material for lignocellulosic ethanol production (Kim and Day, 2011, Yuan, et al., 2008). Additionally, particular features such as yield potential, tolerance to soil and water

constraints, established production systems, and rapid growth make sorghum a reliable energy crop alternative (Reddy, et al., 2005, Rooney, et al., 2007).

Initial efforts to design sweet sorghum energy types started approximately 45 years ago (Monk, et al., 1984, Schaffert and Gourley, 1982). However, its development as such was substantially emphasized only in the past decade, when biofuels became a strategic economic issue worldwide. This lack of a breeding history associated with a relatively small number of scientific studies related to genetics and breeding of energy sorghum has reduced the ability of breeders to effectively design superior cultivars during cultivar improvement. Recently, more information about sweet sorghum genetic diversity, population structure, and quantitative genetics is becoming available (Felderhoff, et al., 2012, Murray, et al., 2009, Murray, et al., 2008, Shiringani and Friedt, 2011, Wang, et al., 2013, Wang, et al., 2009, Zheng, et al., 2011). Despite that fact, efficiently combining the desirable traits into a particular energy sorghum ideotype is still a challenge.

An important limiting factor of producing ethanol from both sugarcane and sorghum relates to the capability of a cultivar to yield high sugar content, particularly sucrose. Sugar yield is the major target trait in breeding programs of both crops, and in sorghum, the trait is routinely selected based on two major components: soluble solids content and juice yield (Corn, 2009). Soluble solids content (SSC), as measured by using digital refractometers, is an efficient method of estimating sucrose concentration in plants, and a primary selection target in energy sorghum breeding programs. However, improvement in sucrose content of commercial sugarcane by traditional breeding has stabilized (Grof and Campbell, 2001). Storage of sucrose in the culms of *Saccharum*

species has an apparent physiological limit of nearly 27% of the fresh weight (Bull and Glasziou, 1963). Total juice sugar concentration in current elite sweet sorghum lines can reach values of 23 % brix (Smith, et al., 1987), of which ~70% corresponds to sucrose (Wu, et al., 2010). Even though improvement in sucrose concentration appears viable, the same limiting phenomenon observed in sugarcane is likely to occur in sweet sorghum. Thus, increased production of juice is the most effective breeding strategy to increase sucrose yield.

Juice yield is a relatively complex trait. A major gene, *d*, has been described as having a large effect on pithy (dry) or juicy sorghum plant stems and was known to be associated with leaf midrib color (Hilson, 1916). Leaf midrib color has been frequently used as a preliminary selection trait in sorghum breeding programs; however, recent research has shown that it is not necessarily a good indicator of juice yield and percent moisture (Felderhoff, et al., 2012). Juice yield is routinely assessed by extracting and quantifying juice in the stem. The process is usually performed by using three roller crushers or hydraulic presses and measuring the amount of juice extracted. Although this assay is widely used and appears reliable; it is time consuming and limits the possibility of evaluating large populations and large scale projects. Currently, the most accepted strategy to enhance juice yield is to increase total plant above ground biomass. However, increasing total plant biomass frequently leads to selection of genotypes with undesirable architecture such as extremely tall plants, which are more inclined to lodge. Instead, larger, thicker stalks and increased population density would be better than increasing height but

are harder to select for. Therefore, indiscriminately increasing total plant biomass without accounting for plant architecture is not a desirable strategy.

Felderhoff, et al. (2012) demonstrated that pithiness of the stalks followed a quantitative distribution in the population they studied. The stalks did not follow a simple inherited segregation ratio for the trait, and often had a dry core surrounded by juicy stalk tissue. Moreover, juice accumulation was not evenly distributed across the whole plant stalk. Gravois, et al. (1990) reported that various secondary traits such as pithiness, tube, stem diameter, stem length, and stalk density, among other stem-related traits showed some level of genetic relationship and also affected major yield components in sugarcane. Performing traditional combining ability studies, Godoy and Tesso (2013) suggested that additive genes played a major role in controlling juice yield; however, no information at the genomic level was reported in this study, leaving an open question about the number, interaction mechanisms, and chromosome position of such genes. This information has suggested that juice yield is more complex than initially reported, indicating that juice accumulation in plant stems shows variation in both radial and vertical dimensions; and that different secondary stem-related traits play an important role in yield component characters.

To dissect such complexity, understanding the relationships between stem properties and juice yield is a critical starting point. Unfortunately, there is lack of information regarding the juice yield trait in both sorghum and sugarcane. Therefore, breeding research, as well as genetic studies, should be developed to allow the improvement of breeding processes for the trait. Accordingly, two different but

complementary experiments were conducted in order to improve the phenotyping methodologies currently used to assess juice yield and to elucidate its genetic nature: (1) a group of twenty diverse sorghum genotypes including grain, forage, and sweet types were used to employ different novel assays such as stem pithiness visual score, stem pithiness image analysis, and stem traits combined index to indirectly assess juice yield, and (2) a quantitative trait locus (QTL) mapping assessment was performed using a biparental linkage mapping population to investigate the genetic basis of juice yield and its relationships with stem related traits in sorghum using multiple phenotyping assays.

CHAPTER II

ASSESSMENT OF STALK PROPERTIES TO PREDICT JUICE YIELD IN SORGHUM

Introduction

As biofuel blending mandates are adopted, the demand for ethanol as a transportation fuel is projected to triple over the next 15 years (International Energy Agency, 2015). Increase in demand concomitantly increases the need for biomass sources to provide fermentable sugar. As a result, breeding efforts and agronomic studies have been conducted to develop sweet sorghum as one of these sources (Davila-Gomez, et al., 2011, Rutto, et al., 2013, Schaffert, 1992, Smith, et al., 1987, Smith and Buxton, 1993, Woods, 2001, Zhang, et al., 2010, Zhao, et al., 2009). However, commercial ethanol yields produced from pilot sweet sorghum production have been lower than many of the economic thresholds which is problematic for commercial application. The limitations are associated with both production and processing. Thus, continued improvement of cultivars for sugar yield and agro-industrial traits is needed to enhance the economics of production.

Sweet sorghum juice is high in soluble sugars which are easily fermented into ethanol (Yu, et al., 2012). As in sugarcane, sugar yield is a direct function of sugar concentration in the juice and the amount of juice produced per unit area (Corn, 2009). In *Saccharum* species, soluble sugar concentrations in the stalk have an apparent physiological limit of ~27% of the fresh weight (Bull and Glasziou, 1963), explaining why sucrose content of commercial sugarcane appears to have stabilized (Grof and Campbell,

2001). Not surprisingly, improvement in sugar yield in modern sugarcane cultivars has been largely attributed to increases in stalk yield (Jackson, 2005). Elite sweet sorghum cultivars are reported to accumulate up to 23% of soluble solids (Smith, et al., 1987), an approximately 15% of total fermentable sugars equivalent on a wet basis (Wu, et al., 2010), which indicates that further increases in sugar content could be possible. However, given the physiological limit to sugar concentrations, a plateau will eventually be reached. Therefore, increased juice production is perhaps a more essential breeding strategy to improve sugar yield in sorghum.

Juice yield is traditionally assessed in small plot samples by pressing sorghum fresh stalks in three roller crushers or hydraulic presses and measuring the amount of juice extracted. Although these methods are reliable and widely used, they are laborious and time-consuming. This limits the number of entries, treatments and replicates that can be evaluated and selected per season thus slowing the rate of genetic gain. While juice yield is a complex trait underlined by multiple morphological, anatomical, and physiological components, the genetic architecture of these component traits tend to be simple. This relationship implies that these component traits are easier to manipulate in breeding programs, and indirect selection schemes based on such traits could be useful for improving juice yield. Before this approach can be used, the traits and their relationships with juice yield must be assessed.

Prior research indicated stalk-related traits such as pithiness, volume, height, and diameter directly affect biomass yield in energy crops. Calviño and Messing (2012) and Salas Fernandez, et al. (2009) clearly demonstrated the importance of height in increasing

biomass yield in sweet sorghum and other grasses. In sweet sorghum, Pfeiffer, et al. (2010) attributed higher biomass in hybrids to increases in plant height and stalk diameter. Gravois, et al. (1990) demonstrated in sugarcane that stalk volume had the greatest influence on stalk weight, followed by negative effects of stalk pithiness and tube formation (stalk's hollow core). At the genetic level, different studies have suggested a pleiotropic role of the genetic determinants of multiple stalk-related traits on sorghum juice and sugar yields (Burks, et al., 2015, Felderhoff, et al., 2012, Guan, et al., 2011, Murray, et al., 2008). However, these and previous studies have not assessed the use of these and other stalk-related traits to predict juice yield.

Herein a comprehensive analysis of the effects of multiple agronomic and stalk-related traits on juice yield was performed. The ability to predict juice yield using these traits was investigated using a set of twenty diverse sorghum genotypes across three Texas environments. The specific objectives of this study were to: (1) determine the association of agronomic and stalk-related traits with juice yield in this panel; and (2) evaluate the ability to predict juice yield using specific combinations of stalk-related traits in this panel.

Material and Methods

Plant material

Twenty *Sorghum bicolor* genotypes contrasting for morpho-anatomical stalk features and energy yield traits were evaluated. Within this group were ten early maturity and ten late maturity genotypes that represented both hybrids and inbreds of grain, sweet, forage and biomass types (Table A-1).

Field design and phenotyping

Field experiments were conducted in a randomized complete block design with three replicates from February to July 2014 and February to June 2015 in Weslaco, Texas, and from May to August 2014 in College Station, Texas. Standard agronomic practices for fertilization, herbicide and insecticide application were followed. To minimize the variation of stalk morpho-anatomical properties due to competition between plants, plots were manually thinned to a target distance of 13 cm between plants in Weslaco and 17 cm in College Station. In addition, tillers were removed at ~15 and 45 days after emergence; tiller removal is a unique component to this study and most previous studies have not done this. This resulted in a population density of ~75,000 plants.ha^{-1} for both locations.

Before harvesting, flowering time and height were measured. Flowering time was estimated as the number of days from planting to when 50% of the plot reached mid-anthesis. Plant height, determined by measuring the distance from the ground to the top of the panicle or to the top of the whorl for genotypes that did not flower prior to harvesting, was measured one day prior to harvesting as an average for the whole plot. Most plots were harvested 25 to 30 days post-anthesis as the grain neared physiological maturity, except for the latest flowering materials which were evaluated on the last harvest date of each trial. At harvest, a center section of the plot consisting of ten sequential plants was hand-harvested by cutting the plants just above the soil surface. Once total biomass weight was recorded from the ten plant sample, the panicles and leaves were removed, and the stalks were reweighed to estimate stalk weight as described in Broadhead and Freeman (1980). Subsequently, the harvested culms were divided into three subsamples of three

plants each designated as samples 1, 2, and 3. The remaining plant was either discarded or used as a spare when needed.

From sample 1, juice was extracted by passing plant stalks twice through a three roller sugarcane crusher Maqtron Cana Shop 200 model (Vencedora Maqtron, Brazil). The extracted juice was weighed, and juice yield was then estimated in $g.plant^{-1}$.

From sample 2, stalk length, internode diameter, and pithiness (juiciness) ratings were assessed separately for each of three plants. Stalk length was measured as the distance between the first base internode above the soil surface and the last top internode. Stalk diameter was measured for the first, third and last plant internodes from base to top by using a digital caliper. Based on the former measurements, stalk volume was calculated using the following formula as described by Worley, et al. (1991): $V=\frac{\pi h\left(r_1^2+r_1 r_2+r_2^2\right)}{3}$, where V is the stalk volume, h is the height, r_1 is radius at the base of the stalk, and r_2 is the radius at the top of the stalk. Stalk density was estimated by dividing stalk fresh weight by stalk volume. Visual rating was performed to estimate percentage of juicy areas after cutting at every internode. A scale from 1 to 9 was adopted, with 1 representing an interval of 0% to 10% of juicy stalk cross section area, 9 representing 90% to 100% of juicy stalk cross section area, and each interval in between representing a 10% increase for juicy area percentage (juiciness) (Figure A-1). Averages and sum of visual juice ratings were calculated by adding and averaging individual internode ratings across the entire stalk. Fresh chopped stalk samples were weighed to estimate fresh weight and subsequently dried at 60º C until the samples reached stable weights to estimate stalk dry weight. Total

stalk moisture was estimated as stalk fresh weight minus stalk dry weight divided by stalk fresh weight times 100.

Sample 3 was used for image analysis of the stalk to estimate pithiness (juiciness). An internode cross section was made for the third, fifth, and seventh above-ground internodes from bottom to top of all three plants. Images of these cross sections were captured using a flatbed scanner (Model Epson Perfection V600) and posterior image analysis was performed using the ImageJ software (Schneider, et al., 2012) to calculate pithy stalk area by performing the following steps. Using standard menu options, images were first converted to RGB (red, green, and blue) format and colors were split into red, green, and blue channels. The blue channel was selected, and the image was cropped to individual internodes. Color thresholds were manually set in order to segment dead air-filled (pithy white tissue) from live (juicy green tissue) parenchyma cells (Figure A-1). Pithy white tissue area was selected and its area estimated in pixels. The same procedure was repeated for the green tissue area. Percentage of pithy stalk area was calculated as $P_{area\ (\%)} = \frac{\text{white tissue area}}{\text{white + green tissue area}} \text{ x } 100$. Percentage of juicy area was equal to 1 - $P_{area\ (\%)}$. Three-plant average estimates were used in statistical analysis for traits measured at the subsample level.

Statistical analysis

Preliminary analysis was performed to diagnose violations of classical assumptions for analysis of variance and mixed models (normality and additivity), and angular transformation was applied when needed. Next, a single-environment model (S1) was fitted as follows: response = block + genotype + error. In this case, block, genotype, and error were assumed to be random and normally distributed. Repeatability was estimated for each environment as $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/r}$, where σ_g^2 and σ_e^2 are the genotypic and error variances, respectively, and r corresponds to the number of blocks. A second model (S2): response = environment + genotype x environment + block within environment + genotype + error was fitted to add GxE. Again, every term was assumed random and normally distributed with constant variance across environments. Overall repeatability estimates were calculated as $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{l} + \frac{\sigma_e^2}{lr}}$, where σ_g^2, σ_{ge}^2, and σ_e^2 correspond to the genotypic, genotype by environment, and error variances in this order; r and l correspond to the number of blocks and environments, respectively. Finally, a third model (S3) was fitted, exactly as model (S2) except taking genotype, environment and the interaction terms as fixed in order to test their effects. In this model, the block and error terms were still assumed to be random and normally distributed with constant variance across environments.

Pair-wise Pearson´s correlation coefficients were calculated between all traits in each location, and phenotypic correlations were subjected to path analysis according to Bollen (1989). To compare and select the best predictive equations, multiple linear

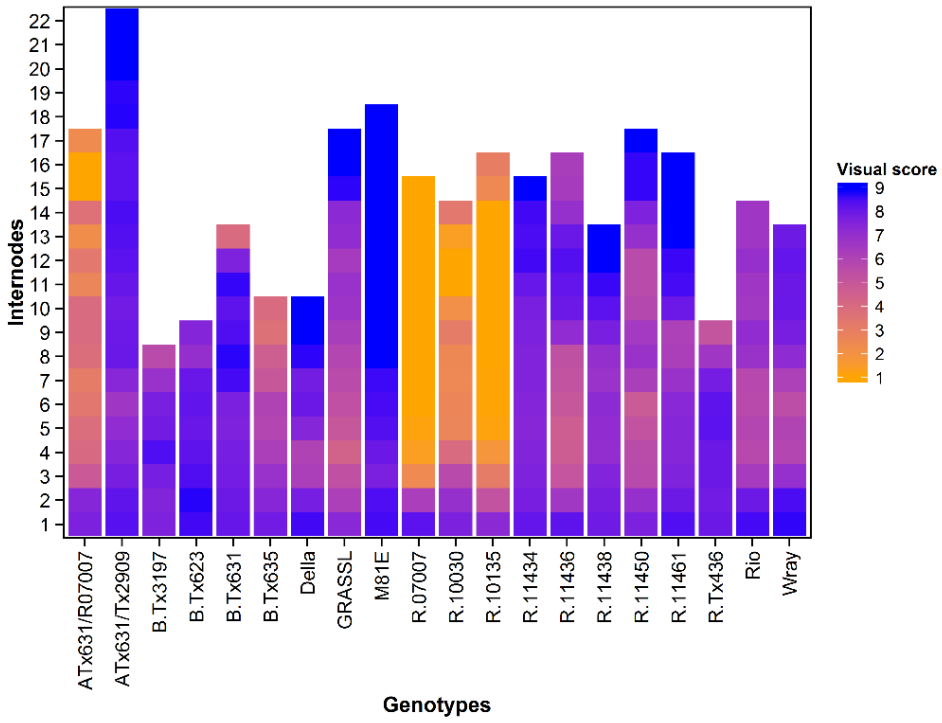
regression models were fitted according to the generalized model: juice yield = a + b x X_1 + c x X_2 + …+ n x X_n, where X_1 to X_n represents the full set of independent variables consisting of agronomic and stalk-related traits. All possible combinations of independent variables were regressed to juice yield by using a stepwise forward selection approach. Next, the two following linear models: (L1) juice yield = a + b x stalk weight + c x stalk moisture and (L2) juice yield = a + b x diameter of third internode + c x plant height + d x visual juice rating of seventh internode were selected after evaluating the physical plausibility, the potential application of each model, and statistical criterions such as Schwarz criterion (BIC), and coefficient of determination (R^2). Subsequently, random samples were assigned to a training (0.75 of total data set) and a testing set (0.25 of total data set) one thousand times. Models (L1) and (L2) were fitted to the training set and cross validated in the testing set every time a new sample was drawn. Then, R^2 and average prediction error (APE) values were computed upon each resampling for the training and testing sets, respectively. Prediction errors were estimated by subtracting the predicted juice yield from the observed juice yield. The mean prediction errors were calculated by averaging the absolute prediction error estimates. The model showing R^2 and APE estimates closest to the mean of their respective distributions was fitted to the entire data set. The two final models selected were then compared for their overall prediction performance. Statistical analyses were performed by using GenStat 18 (VSN International, 2014) and R Software (R Core Team, 2015).

Results

Stalk juiciness characterization

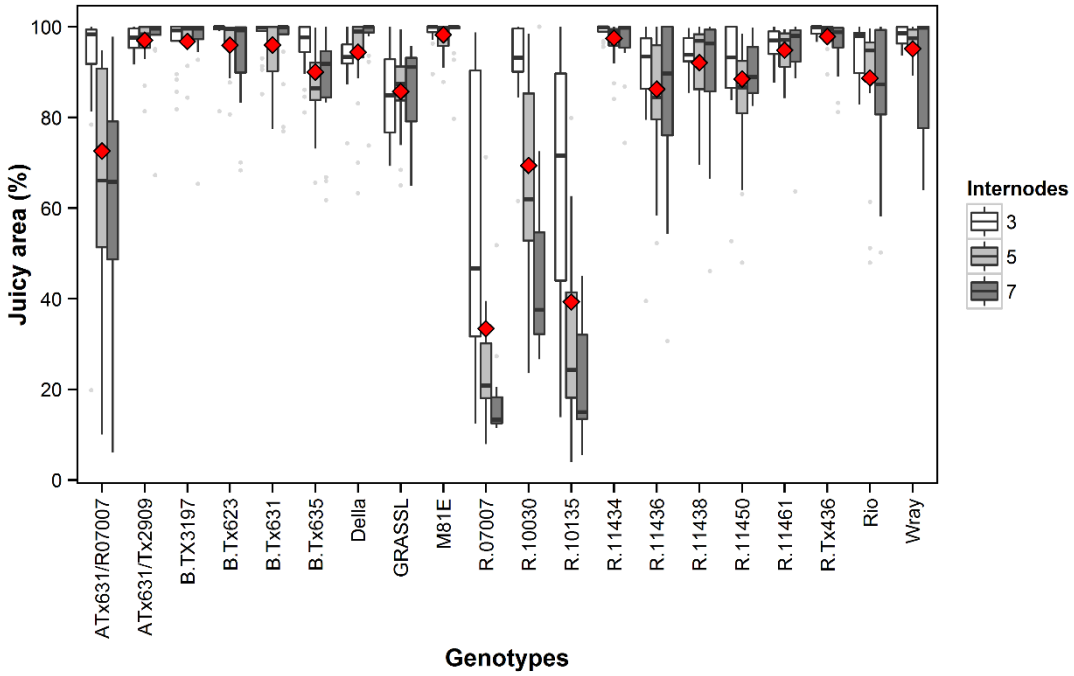
Overall percentage of juicy area estimates varied from ~38% to ~100%, suggesting there was genotypic variability for the trait. Visual screening of juicy area percentage revealed that the phenotypic expression of stalk pithiness follows distinct patterns along the stalk depending on the genotype (Figure 1).

Figure 1 Visual scores for stalk juiciness measured at every above-ground internode of twenty genotypes, and averaged across three Texas environments. A scale of 1 - 9 was adopted, with 1 representing 0% to 10% of stalk juicy area (stalk juiciness), 9 representing 90% to 100% of stalk juicy area, and each interval in between representing a 10% increase for stalk juiciness



Based on digital image analysis within and across internodes, significant variation was observed for percentage of juicy area for internodes three, five, and seven within and across genotypes (Figure 2, Table 1, Tukey´s 95%CI). Most genotypes showed median percentage of juicy area values > 80% across all three internodes, indicating the phenotypic expression of stalk pithiness was reduced or absent in these genotypes.

Figure 2 Percentage of juicy area (PJA) estimates for above-ground internodes three, five and seven from bottom to top of the stem across different genotypes. Summarized distributions of PJA are shown. Range of values, first and third quartile, and median are represented by the length of solid vertical black lines, bottom and top box edges, and horizontal black line, respectively. Red diamonds represent the overall PJA for each genotype as measured by averaging PJA values across internodes three, five, and seven. N = 1260



.

Table 1 Percentage of total variance component estimates and significance of Wald statistics of fixed effects obtained after fitting the mixed model: response = environment + genotype x environment + block within environment + genotype + error, underlined terms were considered random. Analysis was carried out for juice yield and stalk traits measured in twenty sorghum genotypes grown at three locations in Texas. Third, fifth, seventh and last internodes refer to the internode position across the stalk as measured above the ground from bottom to top of the plant

Traits	Environment (E)	Genotype (G)	GxE	Range	Unit
Juice yield	25 **	47 **	22 **	7.0 – 585.3	$g.plant^{-1}$
Diameter of first internode	43 **	27 **	14 **	0.07 – 0.28	dm
Diameter of third internode	35 **	37 **	11 **	0.09 – 0.28	dm
Diameter of last internode	2 **	62 **	18 **	0.05 – 0.20	dm
Stalk length	15 **	54 **	10 **	68.2 – 360.3	cm
Plant height	18 **	60 **	19 **	101.6 – 419.1	cm
Stalk volume	28 **	38 **	25 **	0.04 – 1.3	dm^3
Stalk weight	12 **	63 **	21 **	0.47 – 13.3	kg
Stalk density	13 **	46 **	12 **	0.61 – 3.43	$kg.dm^{3\ -1}$
Stalk moisture	11 **	53 **	20 **	55.7 – 89.9	%
Average of juice ratings	3 **	69 **	18 **	1.5 – 8.7	VR*
Sum of juice ratings	18 **	35 **	37 **	15.1 – 166.1	VR
Percentage of juicy area of the third internode	6 **	38 **	36 **	12.4 – 100	%
Percentage of juicy area of the fifth internode	1^{ns}	69 **	17 **	3.9 – 100	%
Percentage of juicy area of the seventh internode	3 **	67 **	20 **	5.5 – 100	%
Visual juice rating of the third internode	9 **	30 **	38 **	1 - 9	VR
Visual juice rating of the fifth internode	1 **	64 **	20 **	1 - 9	VR
Visual juice rating of the seventh internode	4 **	56 **	26 **	1 - 9	VR

*VR (visual ratings) = a scale from 1 to 9 was adopted, with 1 representing an interval of 0% to 10% of juicy stalk cross section area, 9 representing 90% to 100% of juicy stalk cross section area, and each interval in between representing a 10% increase for juicy area percentage (juiciness)
* significant at 5%; ** significant at 1%; *** significant at 0.1%; ns = non-significant at 5%
Average of juice ratings = average of juice ratings across internodes. Sum of juice ratings = sum of juice ratings across internodes

For some entries however, values as low as ~10% were recorded at internodes five and seven, resulting in a ~10 fold reduction in percentage of juicy area compared to some genotypes. Only two (R.07007 and R.10135) of twenty genotypes had a significant (LSD 95% C.I.) reduction of juicy area measurements at internode three (Figure 2).

Digital image analysis versus visual measurements

Estimates of the stalk juicy area were consistent between digital image analysis and visual screening ($r = 0.93$, $p < 0.01$). Visual measurements of juicy area had slightly lower repeatability compared to digital evaluations (Table 2). This indicated image analysis was more reliable than visual screening for estimating stalk juicy area. Visual screening appears to generally underestimate digital measurements (Figure A-2), and this trend was greatest for genotypes or internodes with the lowest digital juicy area estimates. For instance, visual juice ratings of 1, 2, and 3 had respective digital juicy area of ~20, 40, and 60% on a median basis; while it should correspond on average to ~10, 20, and 30% of juicy area. Digital juicy area estimations overlapped between visual juice ratings, especially from classes 4 to 6 (Figure A-2). However, every rating class was statistically distinct from the other based on digital mean estimates (Tukey's 99% C.I.). While visual phenotyping for stalk juiciness was sufficient to discriminate between every class of stalk juiciness, it might not be precise enough to capture smaller quantitative differences in juiciness as would be desired for studies such as QTL mapping.

Table 2 Single environment and overall repeatability estimates measured across environments for juice yield and stalk traits of twenty sorghum genotypes grown in three Texas environments. Third, fifth, seventh, and last internodes correspond to the internode position across the stalk as measured above the ground from bottom to top of the plant. Average of juice ratings = average of juice ratings across internodes. Sum of juice ratings = sum of juice ratings across internodes

Traits	Environments			Overall
	CS14	WE14	WE15	
Juice yield	0.98	0.98	0.94	0.85
Diameter of first internode	0.87	0.93	0.88	0.80
Diameter of third internode	0.88	0.93	0.89	0.87
Diameter of terminal internode	0.94	0.97	0.90	0.89
Stalk length	0.90	0.90	0.91	0.91
Plant height	0.99	0.99	0.98	0.90
Stalk volume	0.96	0.97	0.93	0.80
Stalk weight	0.99	0.98	0.98	0.90
Stalk density	0.80	0.85	0.89	0.86
Stalk moisture	0.96	0.96	0.88	0.84
Average of juice ratings	0.97	0.99	0.93	0.91
Sum of juice ratings	0.97	0.96	0.93	0.72
Percentage of juicy area of the third internode	0.84	0.90	0.95	0.73
Percentage of juicy area of the fifth internode	0.92	0.96	0.96	0.91
Percentage of juicy area of the seventh internode	0.96	0.98	0.95	0.90
Visual juice rating of the third internode	0.82	0.86	0.96	0.67
Visual juice rating of the fifth internode	0.90	0.97	0.94	0.88
Visual juice rating of the seventh internode	0.95	0.95	0.94	0.84

Environment (E), genotype (G), and GxE effects and repeatability (H^2) of traits

Genotypic effects were highly significant ($p < 0.01$) for all traits measured across all three environments (Table 1). Similarly, the GxE interaction was also highly significant for every trait. The effect of environment was highly significant for all traits except for percentage of juicy area of fifth internode. Repeatability estimates within environments were high for all traits ranging from 0.80 to 0.99 (Table 2). Plant height had the highest H^2 within environments at 0.99 and the second highest H^2 across environments at 0.90. Stalk density H^2 estimates within environments was consistently lower than most traits.

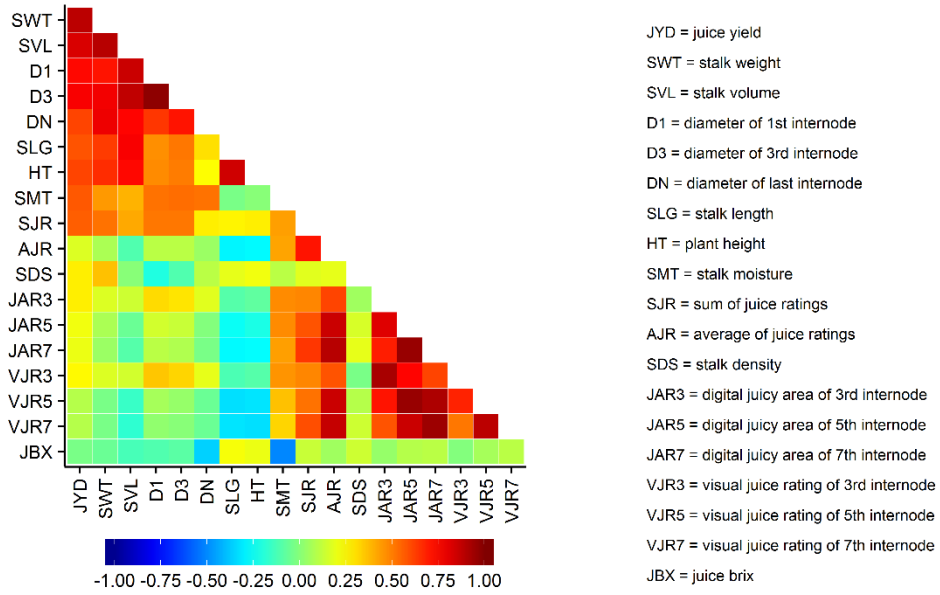
Similarly, H^2 estimates for percentage of juicy area and visual juice ratings was consistently lower for the third internode. Most traits were highly repeatable across the three environments with overall repeatability values greater than 0.80. The effects of GxE were more prominent on traits measured at internode three, and on traits derived from visual evaluations such as sum of juice ratings, with repeatability estimates ranging from 0.67 to 0.73.

Association between stalk traits and juice yield

Phenotypic correlation coefficients varied widely between various stalk traits and juice yield (Figure 3, Table A-2). Most stalk traits were significantly associated with juice yield, except for visual juice ratings of fifth and seventh internodes. Stalk weight showed the strongest association with juice yield (r = 0.88). Stalk volume, diameter of third internode, and diameter of first internode were also tightly associated with juice yield with r values of 0.82, 0.76, and 0.73, respectively. Lower correlations were observed for other traits underlying plant architecture. For example, plant height, stalk length, and diameter of top internode had r values of ~0.60 with juice yield. Similar correlation coefficients were also observed for stalk moisture (r = 0.57) and sum of juice ratings (r = 0.56). Measurements of stalk juicy area such as digital juicy area of third, fifth, and seventh internodes and visual juicy rating of third internode had a low correlation with juice yield (r = 0.20 – 0.27). The correlation between stalk density and juice yield was equally low (r = 0.27). The lowest significant association with juice yield was observed for average of juice ratings with an r value of 0.17. The association among multiple stalk traits indicated

that there was multicollinearity among the traits evaluated; therefore, Path-coefficient analysis was attempted to decouple these.

Figure 3 Pairwise Pearson's phenotypic correlations between eighteen traits measured in twenty sorghum genotypes across three environments in Texas. Internode orders were assigned from the first internode above the ground to the top of the plant. Sum and average of juice ratings were calculated by adding and averaging individual internode ratings across all internodes, respectively. Digital juicy areas were estimated in percentage by performing digital image analysis using ImageJ. Bottom color scale depicts correlation's magnitude. N = 139 (plots containing any missing data were excluded from analysis)



Path-coefficient analysis

The structure of relationships and phenotypic correlations used in the coefficient path analysis are shown in Figure 4. The direct effect of stalk volume on juice yield was positive and the most significant, followed by stalk weight (Table 3 and Table A-3). Similarly, diameter of third internode and plant height had positive direct effects on juice yield, but their values were slightly smaller than stalk volume and stalk weight. As

expected, diameter of third internode and plant height showed high direct effects on stalk volume as well. In this case, the effect of diameter of third internode was approximately two-fold greater than the effect of plant height.

Figure 4 Path diagram showing the cause and effect relationship of: (a) Juice yield (JYD) and its components stalk weight (SWT), and stalk moisture (SMT); and (b) JYD and its components visual juice rating of seventh internode (VJR7), diameter of third internode (D3), and plant height (HT). E indicates the residual effect. Path and correlation coefficients are shown

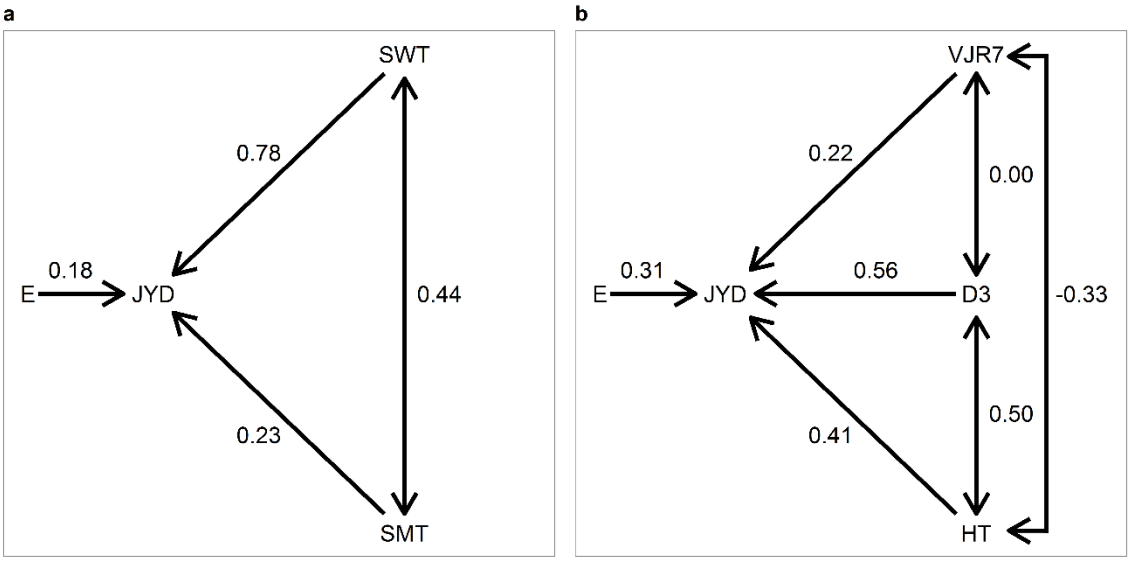


Table 3 Path analysis: unfolding of phenotypic correlations into components of direct and indirect effects, effect significance, and residual effects of six traits separated in two different models. Effects of stalk weight (SWT), stalk moisture (SMT), visual juice rating of seventh internode (VJR7), diameter of third internode (D3), and plant height (HT) on juice yield are shown. Traits were measured in twenty diverse sorghum genotypes across three environments in Texas

Models	Trait	Mode of action	Coefficient	Significance
Model L1	*SWT*	Direct effect on *JYD*	0.78	***
		Indirect effect via *SMT*	0.10	***
		Total effect (direct and indirect)	0.88	***
	SMT	Direct effect on *JYD*	0.23	***
		Indirect effect via *SWT*	0.34	***
		Total effect (direct and indirect)	0.57	***
		Residual	0.18	
Model L2	*VJR7*	Direct effect on *JYD*	0.22	***
		Indirect effect via *HT*	-0.13	***
		Indirect effect via *D3*	0.00	ns
		Total effect (direct and indirect)	0.09	***
	D3	Direct effect on *JYD*	0.56	***
		Indirect effect via *HT*	0.20	***
		Indirect effect via *VJR7*	0.00	ns
		Total effect (direct and indirect)	0.76	***
	HT	Direct effect on *JYD*	0.41	***
		Indirect effect via *VJR7*	-0.07	**
		Indirect effect via D3	0.28	***
		Total effect (direct and indirect)	0.61	***
		Residual	0.31	

* significant at 5%; ** significant at 1%; *** significant at 0.1%; ns = non-significant at 5%

The indirect effect of stalk volume on juice yield through visual juice rating of the seventh internode was significant and negative but almost negligible in magnitude. The same trend was observed for visual juice rating of the seventh internode whose effect on juice yield through stalk volume was also inverted. The indirect effects of plant height on

stalk volume via diameter of the third internode and vice-versa were highly significant and positive. Similarly, no trade-off was observed for indirect effects of stalk weight on juice yield via stalk moisture or vice-versa, however the indirect effect of stalk weight via stalk moisture was much smaller than its counterpart. Both stalk moisture and visual juice rating of the seventh internode had a much smaller direct effect on juice yield when compared to any other trait included in the same model. More importantly, indirect associations where visual juice rating of the seventh internode was involved had either a neutral or a slightly negative impact on juice yield. The indirect effects of plant height via diameter of the third internode on juice yield and vice-versa were also highly significant and positive, as observed for stalk volume.

Prediction of juice yield using stalk traits

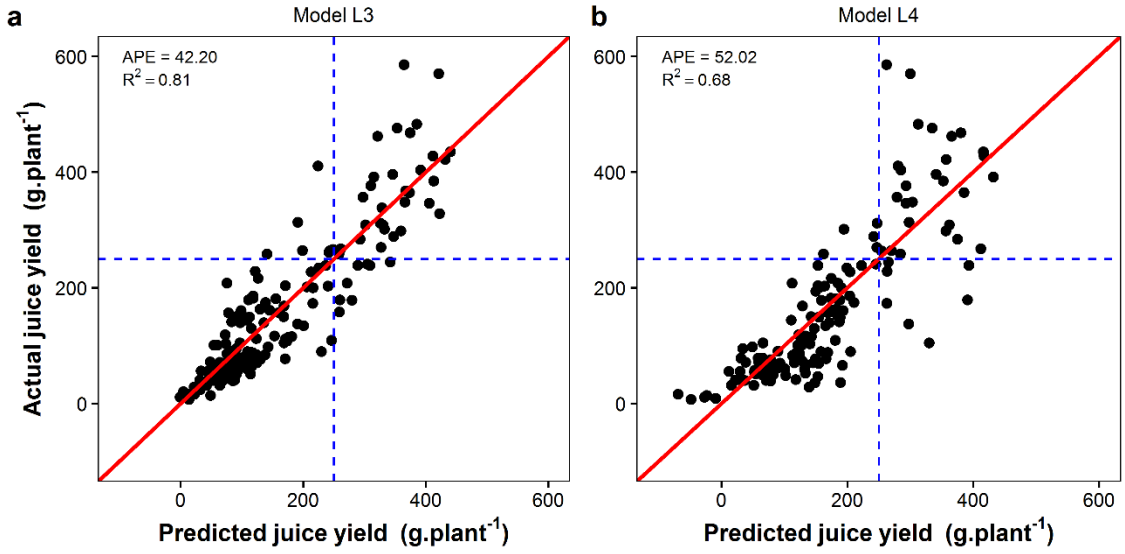
Two distinct sets of independent variables were selected after performing model evaluation analysis giving rise to models (L1) juice yield = a + b x stalk weight + c x stalk moisture and (L2) juice yield = a + b x diameter of third internode + c x plant height + d x visual juice rating of the seventh internode. After fitting and cross validating each model to one thousand random training and testing sets, respectively, model L1 yielded R^2 values that ranged from ~78 to ~86% (Figure A-3a). Model L2 produced a wider range of R^2 values but of smaller magnitude, ranging from ~62 to ~78%. The highest frequency of R^2 estimates were in distinct ranges for each model. Accordingly, values around 80% and 68% were observed at the highest frequency for models L1 and L2, respectively. A similar trend occurred for average prediction error (APE) estimates (Figure A-3b), but in this case

smaller APE values were obtained for model L1. Again, model L2 showed a wider range of APE estimates; nonetheless, greater overlapping of APE estimates from models L1 and L2 was observed in comparison to the R^2 results.

Interestingly, there was a positive linear relationship between APE and R^2 in both models (Figure A-4), in which higher R^2 values consistently yielded larger APE estimates in both situations. Models that showed R^2 and APE values closest to 0.80 and 42 for model L1, respectively, and 0.68 and 52 for model L2, respectively, were selected as final predictive models. These models (L3) juice yield = -364.19 + 27.70 x stalk weight + 5.21 x stalk moisture and (L4) juice yield = -365.20 + 1759.03 x diameter of the third internode + 0.78 x plant height + 9.97 x visual juice rating of the seventh internode were fitted to the entire data set to compare their prediction performance (Figure 5). By setting an arbitrary selection threshold for juice yield of 250 $g.plant^{-1}$ (which would be an approximately 50% selection cut), both models generated similar predictive ability. Although model L3 had larger R^2 and smaller APE, models L3 and L4 would miss similar numbers of genotypes if selection were applied. Accordingly, the reduction of ~10 units in APE from model L4 to model L3 was not sufficient to enhance the precision of such a model as a selection tool; even so its overall predictability was clearly improved. In general, model L3 tended to better predict values higher than 250 $g.plant^{-1}$ by shrinking such values towards the observed measurements, whereas the contrary was true for model L4. Finally, a similar behavior was observed for the jointly linear effect of stalk weight and stalk moisture, and diameter of the third internode, plant height and visual juice rating of the seventh internode on total soluble sugars (data not shown). But in this case, both

sets of independent variables explained a smaller portion of the variation for total soluble sugars than for juice yield.

Figure 5 Graphical display of actual versus predicted juice yield values of two different models. Model L3 (a) corresponds to juice yield = -364.19 + 27.70 x stalk weight + 5.21 x stalk moisture; Model L4 (b) corresponds to juice yield = -362.20 + 1759.03 x diameter of third internode + 0.78 x plant height + 9.97 x visual rating of seventh internode. Internode positions were assigned from bottom to top of the plant starting from the first internode above the ground. Average prediction error (APE) and adjusted R^2 are shown. Vertical and horizontal dashed blue lines depict an arbitrary 250 g.plant^{-1} selection threshold. First and fourth quadrants indicate the number of genotypes that would be misclassified if selection were applied. Second and third quadrants indicate correct classification upon selection



Discussion

Limited throughput of data directly impairs the genetic gain of a trait by limiting the number of progenies evaluated per breeding cycle. In this study, alternative methods to measure and select for juice yield were developed based on the genetic variation for juice yield and its underlying stalk traits. Although the sorghum panel evaluated herein was small, it consisted of elite sorghum breeding lines and hybrids developed for distinct

end uses that contrasted for juice yield and many morphological, anatomical and physiological stalk features. This diversity clearly explains the significant genotypic variation observed for all traits measured. In rather distinct populations and environments, significant genetic variability was also found for juice yield and stalk related traits (Burks, et al., 2015, Felderhoff, et al., 2012, Godoy and Tesso, 2013, Makanda, et al., 2009, Murray, et al., 2008, Pfeiffer, et al., 2010, Rutto, et al., 2013, Smith and Buxton, 1993, Vasilakoglou, et al., 2011, Zou, et al., 2011), demonstrating that improvement of these traits through genetic manipulation is feasible in sorghum.

Repeatability estimates were moderate to high for all traits, indicating the phenotyping techniques employed in this study were relatively consistent. Since tillering and plant population directly influences stalk traits in sorghum (Caravetta, et al., 1990) removing tillers was essential to maintain a consistent stalk number per unit area. This approach likely contributed to enhance the precision of our trials and to explain the relatively high repeatability estimates obtained. However, this procedure is not as easily applied to larger numbers of trials and the relative gain from this approach should be considered relative to the extra labor and cost. The number of locations and replications used in this study are standard in early phases of plant selection in the Texas A&M sorghum breeding program and other sorghum breeding programs, and based on the repeatability values obtained; no additional locations or replications are generally used or needed during cultivar testing.

As demonstrated by the variable selection analysis, stalk weight and stalk moisture were among the best predictive variables of juice yield and they comprise one of the cause

and effect relationship models determining juice yield in this study. Stalk weight is a major underlying component of stalk yield in sugarcane (Bell, et al., 2004, Kang, et al., 1989). Likewise, the trait strongly influences juice extractability in sorghum (Prasad, et al., 2013) as supported by the findings of this study. Another important component of fresh stalk yield in sugarcane is stalk moisture (Liu and Helyar, 2003) which is also strongly associated with juice yield in sorghum at both the phenotypic and genotypic levels (Burks, et al., 2015). In this study however, 70% of the total effect of stalk moisture on juice yield was due to its indirect effect through stalk weight. Interestingly, despite the small magnitude of the indirect effect of stalk weight on juice yield via stalk moisture, it was highly significant and positive. Altogether this indicates that cultivars with higher stalk moisture tended to have higher stalk weight and vice-versa. But since the major impact of stalk moisture was indirectly associated with stalk weight and the latter had a much higher direct impact on juice yield, selecting cultivars with higher stalk weights should be a primary goal when improving juice yield in sorghum.

Because the phenotyping methods used to measure stalk weight in sorghum are labor intensive and time-consuming, using indirect selection schemes based on these processes does not improve the throughput speed needed when screening for juice yield. Herein, coefficient path analysis demonstrated that the joint effects of diameter of the third internode (hereafter referred to as stalk diameter), plant height, and visual juice rating of the seventh internode (hereafter referred as stalk juiciness) on juice yield was substantial. These three traits can be easily and rapidly measured on site without the need of transporting samples from field to lab or using costly equipment. Therefore, using these

traits appears to be an effective way to predict juice yield especially under limited resource scenarios. However, better understanding of the underlying physiology can help to identify new sources of genetic diversity in component traits for improving juice yield that may otherwise be masked in an unsuitable background (e.g. grain sorghum).

Sugarcane stalk diameter and stalk length are the most important traits affecting stalk yield after stalk number (Miller and James, 1974, Miller, 1977). Because these traits underlie stalk volume, this explains the very strong relationship between stalk volume and stalk weight (Brown, et al., 1969), which extends to sorghum. More importantly, stalk volume had a very large effect on juice yield. In this study the direct effects for stalk diameter on stalk volume and juice yield were positive and approximately two-fold greater than the direct effect for plant height on the same traits. In addition, stalk juiciness had a highly significant but smaller direct effect on juice yield. All this implies that increasing juice yield can be achieved by prioritizing and selecting for stalk diameter, plant height and stalk juiciness in that order. The indirect effect of stalk diameter on juice yield through plant height was positive, indicating that genotypes with thicker diameters tend to be taller, which is a desirable combination for high biomass sorghum (Packer and Rooney, 2014). In contrast, the indirect effects of stalk diameter and plant height on juice yield via stalk juiciness was negligible, suggesting independence between these traits. As such, selection for stalk juiciness should be separate from selection for stalk diameter and plant height; since these are independent it may be possible to find novel variation for this trait in short sorghums that would otherwise be unsuitable for sugar production. Moreover, because

stalk juiciness appears to be highly heritable (Stephens and Quinby, 1939, Swanson and Parker, 1931) selection for this trait can be completed in early generations.

From the prediction modeling analyses, stalk weight and stalk moisture (model L3) and stalk diameter, plant height, and stalk juiciness (model L4) successfully predicted juice yield in sorghum. Model L3 was slightly better than model L4 in predicting juice yield. However, neither model accounted for all the variation. This additional variation may be due to genetic factors beyond the traits present in each model or error variance associated with the phenotyping methodologies and different models may be needed for different types/end-uses of sorghum. The subjectivity of visual phenotyping can likely increase the random error variation associated with visual measurements; therefore, plant height and stalk juiciness estimations as assessed in this study were probably less precise than equipment-based measurements of stalk weight and stalk moisture. Regardless, models 3 and 4 still had similar predictive power. At an arbitrary selection threshold of 250 g of juice per plant, both models effectively eliminated inferior materials.

Ideally, on-site crushing of entire plots would be the first choice to assessing and selecting for juice yield in sorghum. However, most crushers and pressers used to measure juice yield at the research level are stationary and operate with a small range of sample sizes, limiting the throughput of this method at multiple plot levels common in breeding programs. All these factors plus the ease of measuring the predictive traits in model L4 reinforces that until new equipment or techniques to efficiently measure juice yield in sorghum are developed, using this model is still more applicable when selecting for juice yield in sorghum breeding programs, especially in early generations. The smaller

explanatory power of stalk weight and stalk moisture as well as diameter of the third internode, plant height, and stalk juiciness on total soluble sugar in this study suggests sugar content is still an equally important component of sorghum sugar yield. The large variation observed for sugar content in this panel might explain its large effect on sugar yield. However, this phenomenon which was likely caused by the low sugar background effect of some grain, forage, and energy sorghums evaluated is rather weaker in dedicated sweet sorghum breeding programs, indicating prediction models based on stalk traits could be potentially applied to assess sugar yield.

Evaluation of each internode in the whole stalk allowed assessing the relationship between stalk-related traits and juice yield and predict juice yield using stalk-related traits in sorghum. In doing so, this research showed that visual estimations appeared more appropriate to distinguish contrasting stalk juiciness classes (i.e., pithy vs juicy). Moreover, the results herein provide compelling evidence that stalk traits could be used in indirect selection schemes when breeding sorghum for biomass and energy purposes. Given the development of high-throughput phenotyping technology, stalk traits (especially stem diameter and plant height) would be considered as primary traits for measurement when trying to predict both plant biomass and juice yield in sorghum. However, this approach may still have limited robustness of prediction. Therefore, these models must be tested in larger independent samples and multiple environments to assess their true applicability. Additional testing should also determine the relative efficiency of the prediction models.

CHAPTER III

QTL ANALYSIS IN A GRAIN X SWEET SORGHUM (*SORGHUM BICOLOR* L. MOENCH) POPULATION: GENETIC CONTROL OF JUICE YIELD AND SOME OF ITS COMPONENT TRAITS

Introduction

Stalk juice extractability is a primary component of sugar yield in sorghum (*Sorghum bicolor* L. [Moench]) (Corn, 2009). Not surprisingly, the trait is a key target of breeding programs aiming to develop sweet sorghum as an alternative source of fermentable sugars. Despite its relevance, studies dedicated to dissecting the genetics of this trait have been conducted predominantly in the past decade (Audilakshmi, et al., 2010, Felderhoff, et al., 2012, Godoy and Tesso, 2013, Murray, et al., 2009, Murray, et al., 2008, Ritter, et al., 2008, Rutto, et al., 2013, Zou, et al., 2011). These recent findings have been useful in developing breeding strategies and selection schemes to improve juice yield. At the same time, however, new questions about the genetic control and behavior of its component traits have arisen; and answering these questions would facilitate designing sorghum cultivars with superior juice and sugar yields.

Together with plant height, the genetic determinants of maturity have been deeply studied in sorghum (Brown, et al., 2008, Childs, et al., 1992, Childs, et al., 1997, Lin, et al., 1995, Mullet, et al., 2012, Murphy, et al., 2011, Murphy, et al., 2014, Pereira and Lee, 1995, Quinby and Karper, 1945, Quinby, 1966, Quinby, 1967, Rooney and Aydin, 1999, Upadhyaya, et al., 2013, Yang, et al., 2014). The knowledge resulting from these studies

has made the genetic manipulation of these traits more efficient. Across sorghum, flowering time is probably the most important trait as it pleiotropically affects numerous traits of economic importance (Childs, et al., 1997, Quinby, 1972). As a consequence, breeding schemes have been adopted to improve traits such as biomass yield via direct manipulation of flowering time (Mullet and Rooney, 2013, Mullet, et al., 2012, Murphy, et al., 2014). Since biomass and juice yields are strongly associated (Burks, et al., 2015); similar breeding approaches could, and are to some extent, be used to also improve juice yield. Undoubtedly, using similar schemes to manipulate traits influencing juice yield such as stem diameter, stem pithiness, stem length, and stem volume would facilitate improving juice productivity. However, further understanding of the genetics underlying these traits is needed before molecular-aided breeding approaches can be employed to improve stem related traits and juice yield.

In independent studies, quantitative trait loci (QTL) for flowering time in sorghum have been mapped in close proximity with plant height, juice yield index, juice yield, harvest yield, vegetative yield, stem fresh yield, dry biomass, stem juiciness, stem diameter, and stem moisture at distinct regions across the genome (Felderhoff, et al., 2012, Murray, et al., 2008, Ritter, et al., 2008, Shiringani, et al., 2010). These findings suggest either a broad pleiotropic effect of flowering time or a physical linkage between flowering time and the traits described. With different mechanisms but similar systemic effect, tillering directly influences grain yield and biomass production in sorghum (Ferraris and Charles-Edwards, 1986, Gerik and Neely, 1987, Hammer, 2006, Stickler and Pauli, 1961). In addition, there is an apparent interdependence between morpho-anatomical properties

of sorghum main stems and tillering (Caravetta, et al., 1990). Empirical observations at the Texas A&M Sorghum breeding program have also suggested that tillering affects distinct stem features in sorghum such as main culm diameter.

Association and linkage based QTL analysis have been commonly used for explaining the genetic basis of variation in sorghum juice yield and its component traits (Burks, et al., 2015, Felderhoff, et al., 2012, Guan, et al., 2011, Lv, et al., 2013, Murray, et al., 2008, Ritter, et al., 2008, Shiringani, et al., 2010). In all these studies, however, tillering was a component of the genetic background variation present in the mapping populations. Genetic segregation of this trait causes undesirable confounding effects that could overshadow the effects of juice yield component traits with smaller additive effects. Moreover, the influence of management and environmental signals on tillering could pose extra confounding variation in these QTL studies. Taken together, these likely reduce the power to identify potential genomic regions contributing to the genetic control of traits underlying juice yield. Thus, eliminating the variation for tillering as well as minimizing the influence of management and environment on this trait helps to reduce such confounding effects.

The hypothesis underlining this research was that systematically thinning and removing tillers at pre-determined stages would allow detection of what otherwise would be small effect QTL for stem-related traits influencing juice yield. Without thinning, these QTL would be masked by the confounding effects of tillering and population density. Accordingly, different stem traits underlying juice yield were measured in a subset of the recombinant inbred line (RIL) population studied by Felderhoff, et al. (2012) to meet the

following major objectives: (1) to genetically map QTL that control variation in stem diameter, stem moisture, and stem pithiness, (2) to investigate the behavior of these QTL in two distinct Texas environments, and (3) to validate QTL for other juice-related traits mapped in previous studies conducted by the Texas A&M Sorghum Breeding and Genetics program.

Material and Methods

Plant material

A recombinant inbred line (RIL) population consisting of a subset of 90 $F_{3:4}$ lines derived from the original population of a cross between 'BTx3197' and 'Rio' (Felderhoff, et al., 2012) was used. This population was developed by using a head to row scheme that consisted of selfing and advancing one panicle per family in each generation following the initial cross. BTx3197 is a grain type fertility maintainer sorghum line with intermediate dry stalks, and Rio is a standard sweet sorghum variety with juicy stalks (Broadhead, 1972). Progenies that were as similar as possible in height were used to minimize the confounding effect of plant height variation on yield traits.

Field experiments

The 90 RILs and both parents were grown in two field experiments in 2014, one in Weslaco, Texas (26.16° N, 97.99° W, 22.25 m) from February to July, and one in College Station, Texas (30.63° N, 96.33° W, 98.45 m) from May to August. Both trials were conducted under rainfed conditions in a randomized complete block design with two

replicates. Standard agronomic practices for fertilization, herbicide and insecticide application were followed. To minimize the variation of stalk morpho-anatomical properties due to competition between plants, plots were manually thinned to a target distance between plants of 13 cm in Weslaco and 17 cm in College Station. In addition, tillers were removed at ~15 and 45 days after emergence. This resulted in a population density of ~75,000 plants.ha^{-1} for both locations.

Phenotyping of the RIL population

Before harvesting, flowering time and height were measured. Plant height, determined by measuring the distance from the ground to the top of the panicle, was measured one day prior to harvesting as an average for the whole plot. Flowering time was estimated as the number of days from planting to when 50% of the plot reached mid-anthesis. Plots were harvested 25 to 30 days post-anthesis as the grain neared physiological maturity. At harvest, an ~ 1.5 m center section of the plot consisting of ten sequential plants was hand-harvested by cutting the plants just above the soil surface, similar to the procedure performed by Felderhoff, et al. (2012). Once total biomass weight was recorded from the ten plant sample, the panicles and leaves were removed, and the stalks were reweighed to estimate stalk weight. Subsequently, the harvested culms were divided into three subsamples of three plants each, designated as samples 1, 2, and 3. The remaining plant was either discarded or used as a spare when needed.

From sample 1, juice was extracted by passing plant stalks twice through a three roller sugarcane crusher Maqtron Cana Shop 200 model (Vencedora Maqtron, Brazil). Juice yield was then estimated in $g.plant^{-1}$.

From sample 2, stalk length, internode diameter, and juiciness ratings were assessed. Stalk length was measured as the distance between the first base internode and the last top internode. Stalk diameter was measured for the first, third and last plant internodes from base to top using a digital caliper. Based on the former measurements, stalk volume was calculated using the following formula as described by Worley, et al. (1991): $V=\frac{\pi h\left(r_1^2+r_1r_2+r_2^2\right)}{3}$, where V is the stalk volume, h is the height, r_1 is radius at the base of the stalk, and r_2 is the radius at the top of the stalk. Visual rating was performed to estimate percentage of juicy areas (stalk juiciness) at every internode. A scale from 1 to 9 was adopted, with 1 representing an interval of 0% to 10% of juicy stalk cross section area, 9 representing 90% to 100% of juicy stalk cross section area, and each interval in between representing a 10% increase for juicy area percentage. The visual juice ratings sum and mean were calculated by adding and averaging individual internode ratings across the entire stalk, respectively. Fresh chopped stalk samples were weighed to estimate stalk fresh weight and subsequently dried at 60º C until the samples reached stable weights to estimate stalk dry weight. Stalk moisture was estimated as stalk fresh weight minus stalk dry weight divided by stalk fresh weight times 100.

Sample 3 was used for image analysis of the stalk to estimate juiciness. An internode cross section was made for the 3rd, 5th, and 7th internodes from bottom to top of all three plants. Images of these cross sections were captured using a flatbed scanner

(Model Epson Perfection V600) and posterior image analysis was performed using the ImageJ software (Schneider, et al., 2012) to calculate juicy stalk area by performing the following steps. Using standard menu options, images were first converted to RGB (red, green, and blue) format and colors were split into green, blue and red channels. The blue channel was selected, and the image was cropped to individual internodes. Color thresholds were manually set in order to segment dead air-filled (pithy white tissue) from live (juicy green tissue) parenchyma cells (Figure A-1). Pithy white tissue area was selected and its area estimated in pixels. The same procedure was repeated for the green tissue area. Percentage of pithy stalk area was calculated as $P_{area\ (\%)} = \frac{\text{white tissue area}}{\text{white+ green tissue area}} \times 100$. Percentage of juicy area was equal to $1 - P_{area\ (\%)}$.

Statistical analysis

All statistical analyses described hereafter were performed using GenStat 18 (VSN International, 2014) and R software (R Core Team, 2015). Preliminary analysis was performed to diagnose violations of classical assumptions for mixed models analysis, and applying different transformation functions did not improve data normality and additivity. Next, a single-environment mixed model was fitted according to the model (3.1): response = genotype + block + row + range + number of millable stalks + harvest date + error. In this case, every term was assumed random with effects normally distributed. The appropriate random model was determined automatically by using the Automatic Spatial Analysis of Row-Column procedure. For this study, all feasible random models were tried including spatial models for rows and columns, and the final model was selected based on

Bayesian information coefficients. Broad sense heritability was estimated for each location as $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/r}$, where σ_g^2 and σ_e^2 are the genotypic and error variances, respectively, and r corresponds to the number of blocks. A second model (3.2), the same as model (2.1), was fitted following the same procedures previously described, and by considering genotype as fixed to test its effect. Model (3.3): response = environment + genotype x environment + genotype + block within environment + row within environment + range within environment + number of millable stalks within environment + harvest date within environment + error was fitted with every term assumed random and normally distributed. The appropriate random model was determined manually by following the same procedure described for model (3.1), but in this case spatial models for erratic spatial trends were not tried. Broad-sense heritability was estimated across environments as $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{l} + \frac{\sigma_e^2}{lr}}$, where σ_g^2, σ_{ge}^2, and σ_e^2 correspond to the genotypic, genotype by environment, and error variances, respectively. Both r and l correspond to the number of blocks and environments, respectively. Finally, model (3.4), the same as model (2.3), was fitted by taking genotype, environment, and genotype x environment as fixed in order to test their effects. In this model, all the remaining terms were assumed random, with effects normally distributed. For each RIL, best linear unbiased estimates (BLUEs) were calculated per and across environment(s) for further use in correlation and QTL analysis.

Genotyping

The restriction enzyme targeted genome resequencing technique, named Digital Genotyping, was used to genotype individual RILs and parental lines as described by Morishige, et al. (2013). In short, genomic DNA was isolated from the leaf tissue of ~10 seedlings of each line following the FastDNA Spin Kit protocol (MP Biochemicals). Leaf tissues were harvested from $F_{3:4}$ lines at ~14 days after seed emergence under controlled greenhouse conditions. Because the tissue was bulked from multiple plants the genotypes essentially represented the $F_{2:3}$ generation. Template DNA libraries were prepared using the restriction enzyme *Fse*I. Next, multiplex identifier barcodes were ligated to the fragments. And after multiple steps detailed in Morishige, et al. (2013) the template was sequenced on an Illumina GAIIx (Illumina). Base calls were generated using Illumina's Real Time analysis (RTA) software. Sequence text files were created using GERALD in Illumina's CASAVA v1.7 software package. These files were then processed to record genotype-specific read depth by using multiple customized perl and python scripts. Reads with a depth of 3 or greater from each parental genotype were aligned to the sorghum reference sequence by BLASTN analysis. The results from the BLASTN analysis were manually inspected to remove those reads that aligned to more than one position in the genome at the same e-value or percent identity and the files from the two parents were combined to identify potential polymorphisms using a custom python script (Morishige, et al., 2013). Finally, appropriate parental alleles (A, B, or H) were assigned to each progeny line using a third python script that searched for each parental sequence from a given *Fse*I site in each progeny line creating a suitable JoinMap (Van Ooijen, 2006) input

file. Missing and heterozygous allele calls were treated as missing data for linkage and QTL analyses.

Linkage map construction

The linkage map was constructed with 741 markers using JoinMap (Van Ooijen, 2006). Linkage groups were created by using the likelihood ratio χ^2 test at different significant levels of stringency. These levels ranged from the logarithm of the odds (LOD) values of 2.0 to 12.0 with a step level value of 1.0. Loci showing significant associations at current LOD threshold value with at least one member of a group were assigned to the same group. The recombination frequency threshold of 0.4 and LOD threshold of 3.0 were considered to assign markers to linkage groups with the Kosambi mapping function (Kosambi, 1943). The multipoint maximum likelihood mapping algorithm was used for map construction. Finally, the function "ripple" and goodness-of-fit test were performed to build map order.

QTL analysis

The BLUE values for juice yield and stalk traits as observed in each environment were analyzed separately in a single-trait and single-environment QTL analysis. For the QTL analysis a mixed model-based protocol as described for multi-trait and multi-environment QTL analyses (Alimi, et al., 2013, Boer, et al., 2007, Malosetti, et al., 2013, Malosetti, et al., 2008) was followed. The protocol consisted of an initial round of simple interval mapping (Lander and Botstein, 1989) that was followed by two consecutive

rounds of composite interval mapping (Jansen and Stam, 1994, Zeng, 1994). From the last round of composite interval mapping a candidate set of QTLs was obtained from which possibly redundant QTLs were removed by a backward elimination procedure, with a final set of QTLs being the end result.

In the simple interval mapping round, the following mixed model was fitted for the individual trait response at each environment; responses = QTL + G + errors, where QTL is the QTL effect along the genome, G represents the residual genetic main effects, i.e., genetic effects not explained by the QTL, and errors represent non-genetic variation. The QTL allele substitution effects were fixed. In the simple interval mapping scan, genomic positions were tested one by one for QTL presence by Wald tests. A multiple test correction was applied according to Li and Ji (2005). In composite interval mapping, potential QTLs elsewhere in the genome were included as cofactors; responses = cofactors + QTL + G + errors. The final multi-QTL model was of the form responses = QTLs + G + errors. The amount of genetic variance explained by a QTL was estimated from the drop in polygenic variance that occurred when the QTL was added to a model with exclusively polygenic effects; responses = G + errors.

To facilitate a more detailed study of pleiotropic action of QTLs, the BLUE values for QTLs observed within and across environments and that co-localized on similar chromosome regions were analyzed together in a multi-trait QTL analysis. For the QTL analysis a mixed model-based protocol as previously described was followed. In this case, the protocol consisted of an initial round of simple interval mapping (Lander and Botstein, 1989) that was followed by two consecutive rounds of composite interval mapping (Jansen

and Stam, 1994, Zeng, 1994). From the last round of composite interval mapping a candidate set of QTLs was obtained from which possibly redundant QTLs were removed by a backward elimination procedure, with a final set of QTLs being the end result.

In the simple interval mapping round, the following mixed model was fitted along the genome for the multiple traits responses jointly; responses = QTL + polygenic effects + errors, with an unstructured variance covariance model imposed on the polygenic effects, i.e., polygenic variances were trait specific and polygenic correlations were unique for each pair of traits. The QTL allele substitution effects were fixed and trait specific. In the simple interval mapping scan, genomic positions were tested one by one for QTL presence by Wald tests. A multiple test correction was applied according to Li and Ji (2005). In composite interval mapping, potential QTLs elsewhere in the genome were included as cofactors; responses = cofactors + QTL + polygenic effects + errors. The final multi-trait multi-QTL model was of the form responses = QTLs + polygenic effects + errors. The amount of genetic variance explained by a QTL was estimated from the drop in polygenic variance that occurred when the QTL was added to a model with exclusively polygenic effects; responses = polygenic effects + errors.

Table 4 Best linear unbiased estimator (BLUEs) for the parents BTx3197, Rio, and 90 derived $F_{3:4}$ sorghum lines evaluated in two Texas environments with respective combined analysis estimations

Trait	College Station, 2014						Weslaco, 2014					
	BTx3197	Rio	RIL mean	S.E.	Min	Max	BTx3197	Rio	RIL mean	S.E.	Min	Max
Average of juice rating	7.9 a	8.1 a	7.6	0.09	5.1	9.4	7.6 a	7.8 a	7.1	0.11	4.5	8.8
Bagasse moisture (%)	64.0 a	62.8 a	65.0	0.25	60.0	71.7	60.1 a	59.9 a	62.2	0.29	53.8	67.7
Diameter of 1st internode (dm)	0.15 a	0.17 b	0.16	0.00	0.1	0.2	0.12 a	0.12 a	0.12	0.00	0.10	0.17
Diameter of 3rd internode (dm)	0.15 a	0.18 b	0.16	0.00	0.1	0.2	0.11 a	0.13 b	0.13	0.00	0.10	0.17
Days to mid-anthesis (d)	68 a	87 b	75	0.73	66	89	73 a	79 b	76	0.27	71	83
Plant height (cm)	106 a	267 b	211	2.64	150	263	114 a	215 b	192	2.14	150	249
Juice rating of 3rd internode (VR)	8 a	9 b	7	0.13	3	9	8 a	8 a	7	0.11	4	9
Juice rating of 5th internode (VR)	9 a	7 b	7	0.17	2	9	8 a	7 b	7	0.14	4	9
Juice rating of 7th internode (VR)	8 a	6 b	8	0.16	3	9	5 a	8 b	6	0.23	2	9
Sum of juice ratings	65 a	90 b	71	1.15	30	95	47 a	56 b	50	0.95	31	75
Juicy area of 3rd internode (%)	98.5 a	100.0 b	98.3	0.22	89.2	100.0	99.8 a	96.3 b	96.8	0.36	84.5	100.0
Juicy area of 5th internode (%)	100.0 a	98.5 a	96.1	0.41	74.6	100.0	98.5 a	98.9 a	94.8	0.98	33.9	100.0
Juicy area of 7th internode (%)	99.8 a	97.8 a	97.5	0.35	81.0	100.0	93.2 a	97.2 a	92.5	1.04	68.2	100.0
Juice yield (k. plant^{-1})	0.05 a	0.24 b	0.20	0.00	0.04	0.25	0.03 a	0.09 b	0.08	0.00	0.02	0.15
Stalk length (cm)	79 a	238 b	181	3.01	113	257	89 a	187 b	166	2.56	103	230
Stalk moisture (%)	76.8 a	74.9 a	76.6	0.30	69.4	83.3	71.5 a	69.9 a	73.3	0.30	66.6	80.3
Stalk volume (dm^3)	0.1 a	0.3 b	0.2	0.01	0.1	0.4	0.1 a	0.2 b	0.1	0.00	0.1	0.3
Total biomass weight (kg)	2.5 a	8.1 b	5.1	0.10	3.1	8.1	1.7 a	3.5 b	3.0	0.06	1.7	4.9
Vegetative biomass weight (kg)	2.0 a	7.6 b	4.4	0.10	2.6	7.6	1.1 a	2.7 b	2.3	0.05	1.4	3.8

Table 4 Continued

Trait	Combined					
	BTx3197	Rio	RIL mean	S.E.	Min	Max
Average of juice rating	7.7 a	7.9 a	7.3	0.09	4.7	8.7
Bagasse moisture (%)	62.3 a	61.1 a	63.6	0.24	59.4	69.2
Diameter of 1st internode (dm)	0.14 a	0.14 a	0.14	0.00	0.11	0.17
Diameter of 3rd internode (dm)	0.13 a	0.15 b	0.14	0.00	0.13	0.18
Days to mid-anthesis (d)	70 a	83 b	75	0.45	68	84
Plant height (cm)	111 a	242 b	202	2.17	150	254
Juice rating of 3rd internode (VR)	8 a	8 a	7	0.11	3	9
Juice rating of 5th internode (VR)	8 a	7 a	7	0.13	3	9
Juice rating of 7th internode (VR)	7 a	7 a	7	0.15	2	9
Sum of juice ratings	55 a	72 b	61	0.91	30	87
Juicy area of 3rd internode (%)	99.1 a	98.1 a	97.5	0.24	86.8	100.0
Juicy area of 5th internode (%)	87.9 a	84.0 a	80.5	0.60	63.9	90.3
Juicy area of 7th internode (%)	83.0 a	82.9 a	81.3	0.67	62.6	90.0
Juice yield (k. plant^{-1})	36.9 a	160.6 b	115.1	3.38	35.9	192.6
Stalk length (cm)	84 a	212 b	174	2.49	109	223
Stalk moisture (%)	74.1 a	72.3 a	75.0	0.27	68.9	81.3
Stalk volume (dm^3)	0.1 a	0.2 b	0.2	0.00	0.1	0.3
Total biomass weight (kg)	1.9 a	5.7 b	4.1	0.09	2.5	6.3
Vegetative biomass weight (kg)	1.4 a	5.1 b	3.4	0.08	1.9	5.6

Parent´s genotypic means followed by the same letter are not significantly different (P=0.05, LSD)
VR (visual ratings) = a scale from 1 to 9 was adopted, with 1 representing an interval of 0% to 10% of juicy stalk cross section area, 9 representing 90% to 100% of juicy stalk cross section area, and each interval in between representing a 10% increase for juicy area percentage (juiciness)

Results

Assessment of parents and RIL phenotypes

Most traits were significantly different between the parental lines BTx3197 and Rio based on genotypic means (BLUEs) at single and combined-environment analyses (Table 4). Averages of juice ratings, bagasse moisture, and stalk moisture were statistically similar between the parents in both environments as well in the combined-environment level. Likewise, diameter of 1st internode, juice rating of 3rd internode, and digital juice area of 3rd internode did not differ between the parents in Weslaco, 2014.

At the combined-environment level, parents showed statistically similar performance for diameter of the 3rd internode, as well as for digital and visual assessments of juice area percentage for the 3rd, 5th, and 7th internodes. As observed in previous QTL linkage mapping studies (Felderhoff, et al., 2012, Murray, et al., 2008), the sweet sorghum parent Rio consistently showed significantly higher values for height, days to flowering, and biomass yields than the grain type parent, which is this study was BTx3197. Stalk-related traits such as diameter of the 3rd internode, sum of juice ratings, stalk length, and stalk volume were consistently distinct between the parents. The parents also showed significantly contrasting performance for flowering, plant height, biomass weights, and juice yield for individual and combined analyses. Range of BLUEs within and across-environments revealed the presence of transgressive segregation for all traits in at least one of the two environments. This phenomenon occurred even in cases where the performances of the two parents were statistically equal as observed for the average of juice ratings, bagasse moisture, stalk moisture, and stalk juice ratings (Table 4). Increases

in phenotypic performance of many traits, especially for those traits directly influenced by flowering, were observed for both parental and recombinant inbred lines when planted in College Station, 2014.

Variance components and trait heritability

Mixed model analysis revealed a significant genotypic effect for all traits within and across environments (Table 5). Within environment, genetic variation explained a relatively high proportion of the phenotypic variation of many traits, which is why fairly high broad sense heritability estimates were obtained for such characters. Yet, less than 25% of the total model variation was due to genotypic effects for diameters of the 1st and 3rd internodes, and bagasse moisture in College Station and Weslaco, respectively. Because of that, these traits had the smallest heritability estimates within each location. At the multi-environment level, environmental effects were significant for all traits; although, these effects were very small for some traits (e.g., average of juice ratings, juice ratings of the 3rd and 5th internodes, and digital assessment of juice areas for the 3rd, 5th, and 7th internodes). Effects from genotype by environment interaction were significant for all traits except bagasse moisture, diameter of the 3rd and 5th internodes, and stalk volume. The relative influence of genotype by environment interaction was greatest for flowering time, and stalk juiciness. Heritability estimates obtained from the combined analysis were the lowest for days to flowering. Over all analyses however, heritability estimates were the highest for plant height and stalk length with values greater than 90%. All remaining traits were intermediate to highly heritable with values ranging from ~51% to 90%.

Table 5 Variance components presented in percentage of total variation and heritability (H^2) estimates are shown for all traits as assessed from individual (CS14 = College Station, 2014 and WE14 = Weslaco, 2014) and combined-environment analyses

Environment	Effects	ajr	bmt	d1	d3	dy	ht	jr3	jr5	jr7	jrs
CS14	σ^2_g	63.1***	34.4***	19.0*	23.3*	58.3***	79.0***	63.6***	59.1***	59.7***	54.6***
	σ^2_{error}	30.9	46.8	75.2	70.4	33.7	13.7	34.6	35.5	38.4	45.4
	$\boldsymbol{H^2}$	**80.3**	**59.5**	**33.6**	**39.9**	**77.6**	**92.0**	**78.6**	**76.9**	**75.7**	**70.6**
WE14	σ^2_g	68.6***	22.5*	49.3***	52.6***	61.4***	83.3***	38.8***	53.1***	64.3***	50.4***
	σ^2_{error}	29.7	68.6	50.3	44.6	26.1	13.6	61.2	40.6	34.8	49.1
	$\boldsymbol{H^2}$	**82.2**	**39.7**	**66.2**	**70.2**	**82.5**	**92.4**	**55.9**	**72.4**	**78.7**	**67.2**
Combined	σ^2_e	5.0***	27.6***	55.5***	56.0***	11.0**	21.4***	1.9**	1.9*	17.9***	59.3***
	σ^2_g	48.0***	21.5***	11.6***	13.6***	20.2***	53.6***	36.6***	31.4***	25.9***	15.2***
	$\sigma^2_{g\ x\ e}$	16.1***	0.0	1.9	0.8	32.2***	11.8***	14.1**	24.0***	25.8***	6.2**
	σ^2_{error}	27.5	44.9	29.5	27.9	27.8	10.0	46.9	37.8	29.1	19.2
	$\boldsymbol{H^2}$	**76.2**	**65.7**	**58.2**	**64.9**	**46.7**	**86.5**	**66.1**	**59.4**	**56.2**	**65.8**

Table 5 Continued

Environment	Effects	jar3	jar5	jar7	jyd	slg	smt	svl	tbw	vbw
CS14	σ^2_g	63.6***	64.3***	37.7***	52.3***	80.2***	42.5***	41.3***	57.4***	65.0***
	σ^2_{error}	34.7	28.5	54.7	39.9	13.8	50.6	53.9	33.0	26.3
	$\boldsymbol{H^2}$	**78.6**	**81.9**	**58.0**	**72.4**	**92.1**	**62.7**	**60.5**	**77.6**	**83.2**
WE14	σ^2_g	44.6***	35.9***	44.2***	48.5***	82.8***	53.5***	70.4***	77.3***	79.8***
	σ^2_{error}	50.8	63.0	47.4	32.1	13.8	34.3	25.7	16.4	15.0
	$\boldsymbol{H^2}$	**63.7**	**53.3**	**65.1**	**75.2**	**92.3**	**75.7**	**84.5**	**90.4**	**91.4**
Combined	σ^2_e	10.9***	0	12.5***	50.7***	13.7***	29.0***	58.3***	53.4***	58.1***
	σ^2_g	25.4***	25.7***	25.2***	14.0***	62.2***	26.8***	15.5***	17.6***	15.3***
	$\sigma^2_{g\ x\ e}$	18.2***	19.1***	14.5***	12.1***	17.1***	8.3*	4.5	10.3***	12.5***
	σ^2_{error}	41.7	49.9	47.8	19.7	2.4	30.2	19.8	14.0	10.4
	$\boldsymbol{H^2}$	**56.5**	**53.8**	**56.8**	**56.1**	**87.2**	**69.7**	**75.9**	**67.0**	**63.3**

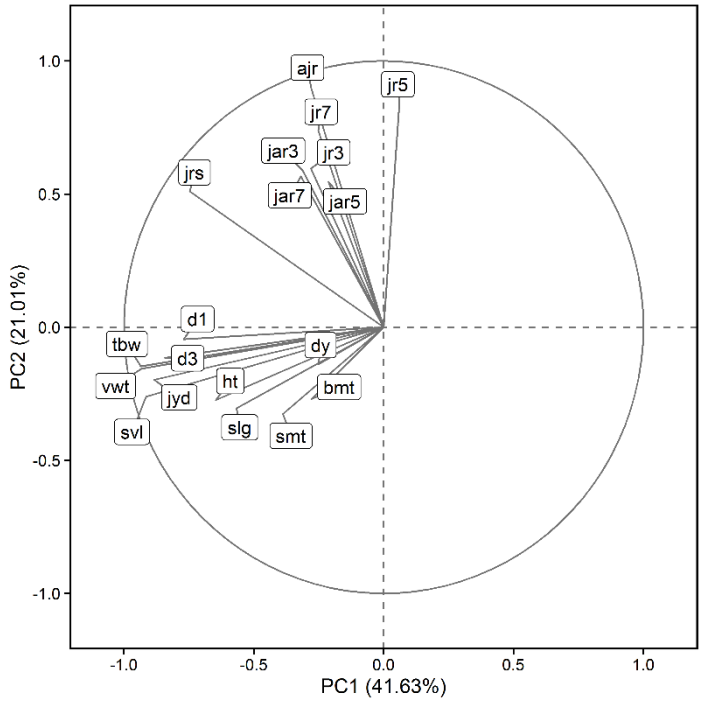
ajr = average of juice ratings, bmt = bagasse moisture, d1 = diameter of 1st internode, d3 = diameter of 3rd internode, dy = days to mid-anthesis, ht = plant height, jr3 = juice rating of 3rd internode, jr5 = juice rating of 5th internode, jr7 = juice rating of 7th internode, jrs = sum of juice ratings, jar3 = digital juice area of 3rd internode, jar5 = digital juice area of 5th internode, jar7 = digital juice area of 7th internode, jyd = juice yield, slg = stalk length, smt = stalk moisture, svl = stalk volume, tbw = total biomass weight, vbw = vegetative biomass weight

* significant at 5%; ** significant at 1%; *** significant at 0.1%; ns = non-significant at 5%

Associations among traits

Trait associations were investigated based on correlation principal component analysis and are depicted in Figure 6 and Table A-4. From these, there is evidence that the first component contrasted visual juice rating of the 5th internode with stalk volume, biomass weights and juice yield. This suggests that the increased juice rating of the 5th internode tends to be associated with reduced performance for stalk volume, biomass weight and juice yield.

Figure 6 Correlation circle of variables measured in two Texas environments (College Station, 2014 and Weslaco, 2014) in the first two principal components of the PCA



ajr = average of juice ratings, bmt = bagasse moisture, d1 = diameter of 1st internode, d3 = diameter of 3rd internode, dy = days to mid-anthesis, ht = plant height, jr3 = juice rating of 3rd internode, jr5 = juice rating of 5th internode, jr7 = juice rating of 7th internode, jrs = sum of juice ratings, jar3 = digital juice area of 3rd internode, jar5 = digital juice area of 5th internode, jar7 = digital juice area of 7th internode, jyd = juice yield, slg = stalk length, smt = stalk moisture, svl = stalk volume, tbw = total biomass weight, vbw = vegetative biomass weight

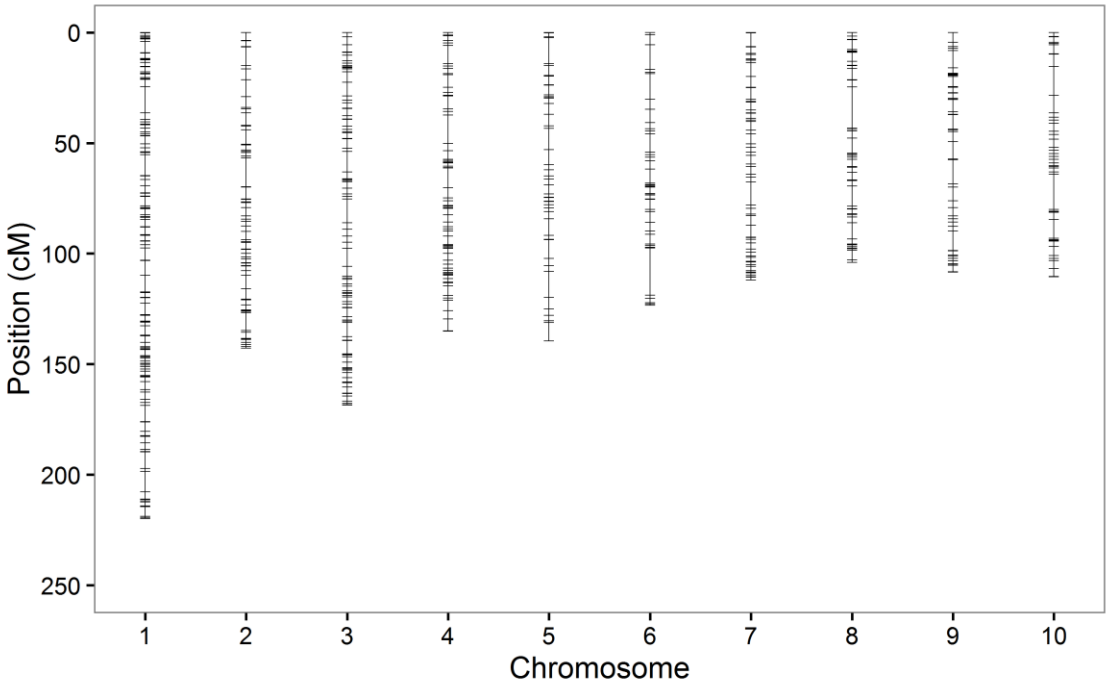
The second component contrasted the different assessments of stalk juiciness, principally average of juice ratings, with stalk and bagasse moistures. This suggests an inverse relationship between increased stalk juiciness and bagasse moisture. Strong positive associations were observed for juice yield with total biomass weight, vegetative biomass weight, and stalk volume. Likewise, the diameter of the third internode and plant height were closely and positively associated with juice yield, vegetative and biomass weights, and stalk volume; however, with slightly smaller magnitude. Most visual and digital assessments of stalk juiciness were tightly associated, except for the sum of juice ratings and juice rating of the 5^{th} internode which had weaker associations with the remaining assessments of stalk juiciness. Although the single component contrast suggests an inverse relationship among some traits as previously described, the joint contrasts of PC1 and PC2 revealed that some of these relationships were weak, suggesting independence between some traits (i.e. bagasse and stalk moisture vs. stalk juiciness; stalk juiciness vs. juice yield).

Linkage map

A total of 741 markers were scored through 90 $F_{2:3}$ individuals resulting in a linkage map consisting of 10 linkage groups. The order of the markers mostly agreed with the expected marker order based on the sorghum physical map (Paterson, et al., 2009). The genomic coverage of this linkage map spanned 1364 cM similar to the map lengths described in studies of Boivin, et al. (1999), Childs, et al. (1997), Hart, et al. (2001), and Mace, et al. (2008). An average distance of 1.84 cM between markers was observed

(Figure 7). A total of 30.07% of the loci were heterozygous, 5% higher than the theoretical 25.00% heterozygosity that would be expected for an $F_{2:3}$ population. For the construction of the linkage map heterozygous loci were ignored. The distribution of alleles from each parent was similar with 34.42% for Tx3197 and 33.93% for Rio. All 741 marker data was included in map construction.

Figure 7 The genetic map showing the 10 sorghum chromosomes and positions of markers used in this study



Single trait QTL analysis, College Station 2014

In College Station (CS14), nine significant QTLs were detected across the genome, influencing different traits with different magnitudes (Table 6). Seven of these underlined traits related to stalk juiciness. Accordingly, two QTLs for juice rating of the 7th internode, namely JR7-7 and JR7-9, were mapped at positions (QTL position corresponds to – $\log_{10}$(p) peak for a given QTL) 52.04 and 108.33 cM of chromosomes 7 and 9, respectively. JR7-7 explained 20.67% of genetic variance for visual juice rating of the 7th internode while JR7-9 explained 13.71% of the genetic variance for the same trait. For both JR7-7 and JR7-9, the high parent alleles (i.e., alleles increasing the phenotype) were donated by the grain type parent BTx3197 with additive effects of 0.66 and 0.54 (in a 1 to 10 scale), respectively.

Four QTLs were found for digital assessments of juicy areas in distinct internodes. Accordingly, one QTL was detected for digital juice area of the 3rd internode (JAR3), one for digital juice area of the 5th internode (JAR5), and two for digital juice area of the 7th internode (JAR7). The QTLs detected for digital juicy area of internodes 3 and 5, namely JAR3-7 and JAR5-6, were mapped at positions 98.00 and 72.85 cM on chromosomes 7 and 6, respectively. JAR3-7 and JAR5-6 explained 21.68% and 12.38% of the genetic variance, respectively. High parent alleles were donated by the parent Rio at both QTL loci. For juice area of the 7th internode two QTLs were mapped on different chromosomes, one on chromosome 3 (JAR7-3) and one on chromosome 7 (JAR7-7). JAR7-7 explained a slightly larger percentage of the genetic variance at 18.45% compared to 13.59% for JAR7-3. Similar additive effects were observed for both QTLs, but the high parent alleles

Table 6 Estimates of QTL effects of single-trait analysis for multiple stalk and agronomic-related traits

Environment [a]	QTL	QTL position (cM) [b]	Closest markers	Additive effects	High value allele [c]	%GV	Avse
CS14	JR7-7	52.04	ATM7.38 – ATM7.35	0.66	BTx3197	20.67	(0.15)
	JR7-9	108.33	ATM9.77 – ATM9.82	0.54	BTx3197	13.71	(0.14)
	JAR3-7	98.00	ATM7.78 – ATM7.80	0.97	Rio	21.68	(0.29)
	JAR5-6	72.85	ATM6.49 – ATM6.58	1.40	Rio	12.38	(0.42)
	JAR7-3	10.38	ATM3.19 – ATM3.23	1.25	Rio	13.59	(0.34)
	JAR7-7	52.04	ATM7.38 – ATM7.35	1.46	BTx3197	18.45	(0.35)
	JRS-10	102.17	ATM10.99 – ATM10.98	4.95	Rio	16.77	(1.37)
	SLG-7	27.54	ATM7.19 – ATM7.23	14.12	BTx3197	18.29	(3.79)
	SMT-5	32.07	ATM5.21 – ATM5.26	1.31	BTx3197	15.30	(0.36)
WE14	AJR-6	69.85	ATM6.49 – ATM6.55	0.44	Rio	18.12	(0.11)
	BMT-1	149.75	ATM1.212 – ATM1.210	1.07	Rio	15.54	(0.29)
	D3-1	132.70	ATM1.178 – ATM1.185	0.006	Rio	27.84	(0.001)
	D3-6	118.90	ATM6.81 – ATM6.99	0.005	BTx3197	17.02	(0.001)
	DY-5	93.59	ATM5.51 – ATM5.54	1.04	BTx3197	15.93	(0.28)
	JR5-6	49.97	ATM6.22 – ATM6.37	0.58	Rio	20.64	(0.15)
	JR7-6	69.04	ATM6.48 – ATM6.55	0.86	Rio	19.18	(0.22)
	JAR7-6	72.98	ATM6.57 – ATM6.62	3.48	Rio	15.46	(1.05)
	JYD-5	93.59	ATM5.51 – ATM5.54	10.62	BTx3197	17.72	(2.76)
	SMT-1	148.68	ATM1.199 – ATM1.209	1.19	Rio	17.48	(0.26)
	SMT-3	82.47	ATM3.86 – ATM3.92	1.04	BTx3197	13.35	(0.28)
	SMT-5	14.08	ATM5.6 – ATM5.8	1.35	BTx3197	22.27	(0.26)
	TBW-5	93.73	ATM5.52 – ATM5.55	0.24	BTx3197	15.93	(0.07)
	VBW-5	93.73	ATM5.52 – ATM5.55	0.20	BTx3197	16.13	(0.06)

Table 6 Continued

Environment [a]	QTL	QTL position (cM) [b]	Closest markers	Additive effects	High value allele [c]	%GV	Avse
	AJR-6	69.85	ATM6.49 – ATM6.55	0.35	Rio	17.26	(0.09)
	BMT-1	155.50	ATM1.228 – ATM1.230	0.90	Rio	15.86	(0.24)
	D3-1	132.70	ATM1.178 – ATM1.185	0.005	Rio	19.37	(0.001)
Combined	JAR5-7	52.04	ATM7.38 – ATM7.35	2.40	BTx3197	17.61	(0.59)
	JAR5-9	102.28	ATM9.61 – ATM9.64	2.18	BTx3197	14.49	(0.58)
	SMT-1	147.38	ATM1.200 – ATM1.207	1.27	Rio	25.11	(0.25)
	SMT-5	23.76	ATM5.14 – ATM5.17	1.14	BTx3197	19.93	(0.25)

Average standard errors (Avse) for each effect are shown between parentheses. QTL are coded as average of juice ratings (AJR), bagasse moisture (BMT), diameter of 3rd internode (D3), days to mid-anthesis (DY), visual juice rating of 5th internode (JR5), visual juice rating of 7th internode (JR7), digital juice area of 3rd internode (JAR3), digital juice area of 5th internode (JAR5), digital juice area of 7th internode (JAR7), juice yield (JYD), stalk moisture (SMT), total biomass weight (TBW), vegetative biomass weight (VBW). Number following hyphen at each QTL name indicates the chromosome where they were mapped. %GV stands for the percentage of genetic variance that is explained by a given QTL

[a] CS14: College Station, TX, 2014, WE14: Weslaco, TX, 2014, Combined: multiple environment analysis

[b] QTL position corresponds to $-\log_{10}(p)$ peak for a given QTL

[c] Allele that increases the phenotype

at each QTL came from distinct parents. At chromosome 10, one QTL was found for sum of juice ratings at position 102.17 cM that explained ~17% of the genetic variance for this trait. At this locus, the Rio allele contributes to increase the phenotype with an additive effect of 4.95.

A single QTL for stalk length was mapped on chromosome 7 at position 27.54 cM that explained 18.29% of the genetic variance for this trait. This locus had an additive effect of ~14 cm with superior alleles coming from the parent BTx3197. Similarly, a unique stalk moisture QTL was also mapped on chromosome 5. This QTL, SMT-5, explained about 15% of the trait's genetic variance with an additive effect of 1.31%, and in this case high parent alleles were donated by the parent BTx3197.

Single trait QTL analysis, Weslaco 2014

Compared to College Station 2014, fewer QTLs were found for traits related to stalk juiciness (Table 6). Nonetheless, all QTL found for stalk juiciness were mapped at relatively close positions on chromosome 6 (AJR-6 @ 69.85 cM, JR5-6 @ 49.97 cM, JR7-6 @ 69.04 cM, and JAR7-6 @ 72.98 cM). These QTLs explained a considerable portion of the genetic variance ranging from 15.46% for JAR7-6 to 20.64% for JR5-6. Moreover, the superior alleles at all QTL loci were donated by the parental line Rio with additive effect values that ranged from 0.44 to 3.48 (on a 1 to 10 scale). One QTL was detected for bagasse moisture (BMT-1) and one for stalk moisture (SMT-1) at chromosome 1 at 149.75 and 148.68 cM, respectively. Superior alleles were donated by Rio at both loci. Coincidently, both QTLs explained similar portion of the genetic variance for bagasse and

stalk moisture (15.54% and 17.48%) with similar additive effects of 1.07 and 1.19%. In addition, two distinct QTLs for stalk moisture were mapped on chromosomes 3 (SMT-3) and 5 (SMT-5) at 82.47 and 14.08 cM, respectively. Although the additive effects and the variance explained by SMT-3 and SMT-5 were similar to SMT-1, the alleles increasing stalk moisture at these loci were donated by the parental line BTx3197.

Two QTLs controlling the diameter of the 3rd internode were mapped on chromosomes 1 (D3-1) and 6 (D3-6). Both QTLs had practically the same additive effect of 5 mm. However, D3-1 explained a larger proportion of the genetic variance for the trait at 27.84% with positive alleles from Rio, while for D3-6 the high value alleles were donated by BTx3197. Only one QTL was found for days to mid-anthesis (DY-5) at 93.59 cM on chromosome 5 which collocated with a juice yield QTL (JYD-5) at the exact same position. Furthermore, a QTL for vegetative biomass yield (VBW-5) and total biomass yield (TBW-5) also co-localized with the flowering and juice yield QTLs at position 93.73 cM. At all four QTL loci (DY-5, JYD-5, VBW-5, TBW-5) the high value allele came from the parent BTx3197. The percentage of genetic variation explained by these four QTLs were consistent ranging from 15.93 to 17.72%. The effect of replacing one allele from Rio with one allele from BTx3197 at DY-5 was a one day delay to flowering. The substitution effect at JYD-5 caused an increase of 10 g of juice per plant, at TBW-5 an increase of 2.4 kg of total biomass weight per sample basis (10 plants), and at VBW-5 an increase of 2.0 kg of vegetative biomass weight per sample basis (10 plants).

Single trait QTL analysis, combined

QTL analysis using the combined analysis BLUEs revealed a QTL for average of juice ratings which mapped on chromosome 6 at the same position for AJR-6 found in Weslaco 2014 (WE14) (Table 6). Even though the high value allele for AJR-6 found from the combined analysis was donated by the parental line Rio, both the additive effect and the genetic variance explained by this QTL were slightly smaller in WE14 with values equal to 0.35 and 17.26%, respectively. A similar phenomenon was observed for BMT-1, D3-1, SMT-1, and SMT-5; however, small variations were observed between the positions where these QTLs were mapped for Weslaco 14 and for the combined analysis as well as for the additive effects and genetic variance explained by these QTLs in each of these environments (Weslaco 2014 vs. Combined). Two QTLs distinct from those mapped in both College Station 2014 and Weslaco 2014 were mapped by using the BLUEs from the combined analysis. Two QTLs for digital juicy area of the 5^{th} internode, JAR5-7 and JAR5-9, were detected at 52.04 and 102.28 cM on chromosomes 7 and 9, respectively. High value alleles at both loci were donated by the parental line BTx3197. JAR5-7 had an additive effect of 2.40 (on a 1 to 10 scale) and explained 17.61%, while the JAR5-9 additive effect was equal to 2.18 and explained 14.49% of the genetic variance. A summary of QTL positions detected from single and multi-environment analysis is depicted in Figure 8.

Table 7 Estimates of QTL effects of multi-trait analysis for multiple stalk and agronomic-related traits

Environment [a]	QTL	Chrs	QTL position (cM) [b]	Closest markers	Additive effects [c]	High value allele [d]	%GV	Avse
Combined	d3	1	148.70	ATM1.207 - ATM1.212	0.47	Rio	22.0	0.11
WE14	d3	1	148.70	ATM1.207 - ATM1.212	0.50	Rio	25.4	0.10
Combined	smt	1	148.70	ATM1.207 - ATM1.212	0.51	Rio	25.5	0.10
WE14	smt	1	148.70	ATM1.207 - ATM1.212	0.50	Rio	25.3	0.10
Combined	bmt	1	148.70	ATM1.207 - ATM1.212	0.47	Rio	22.0	0.11
WE14	bmt	1	148.70	ATM1.207 - ATM1.212	0.47	Rio	18.9	0.11
WE14	d3	5	14.08	ATM5.6 - ATM5.8	0.25	BTx3197	6.0	0.11
Combined	smt	5	14.08	ATM5.6 - ATM5.8	0.44	BTx3197	19.0	0.10
WE14	smt	5	14.08	ATM5.6 - ATM5.8	0.51	BTx3197	25.5	0.10
Combined	bmt	5	14.08	ATM5.6 - ATM5.8	0.51	BTx3197	9.5	0.11
WE14	bmt	5	14.08	ATM5.6 - ATM5.8	0.51	BTx3197	7.6	0.11
WE14	d3	8	54.68	ATM8.46 - ATM8.59	0.25	BTx3197	6.5	0.10
Combined	smt	8	54.68	ATM8.46 - ATM8.59	0.31	BTx3197	9.9	0.10
WE14	jr5	6	69.85	ATM6.49 - ATM6.55	0.35	Rio	12.1	0.11
WE14	jr7	6	69.85	ATM6.49 - ATM6.55	0.54	Rio	28.7	0.11
Combined	ajr	6	69.85	ATM6.49 - ATM6.55	0.43	Rio	18.0	0.11
WE14	ajr	6	69.85	ATM6.49 - ATM6.55	0.43	Rio	18.7	0.11
CS14	jar5	6	69.85	ATM6.49 - ATM6.55	0.35	Rio	12.5	0.11
WE14	jar7	6	69.85	ATM6.49 - ATM6.55	0.43	Rio	18.6	0.12

Average standard errors (Avse) for each effect are shown between parentheses. QTL are coded as average of juice ratings (AJR), bagasse moisture (BMT), diameter of 3rd internode (D3), days to mid-anthesis (DY), visual juice rating of 5th internode (JR5), visual juice rating of 7th internode (JR7), digital juice area of 3rd internode (JAR3), digital juice area of 5th internode (JAR5), digital juice area of 7th internode (JAR7), juice yield (JYD), stalk moisture (SMT), total biomass weight (TBW), vegetative biomass weight (VBW). Number following hyphen at each QTL name indicates the chromosome where they were mapped. %GV stands for the percentage of genetic variance that is explained by a given QTL

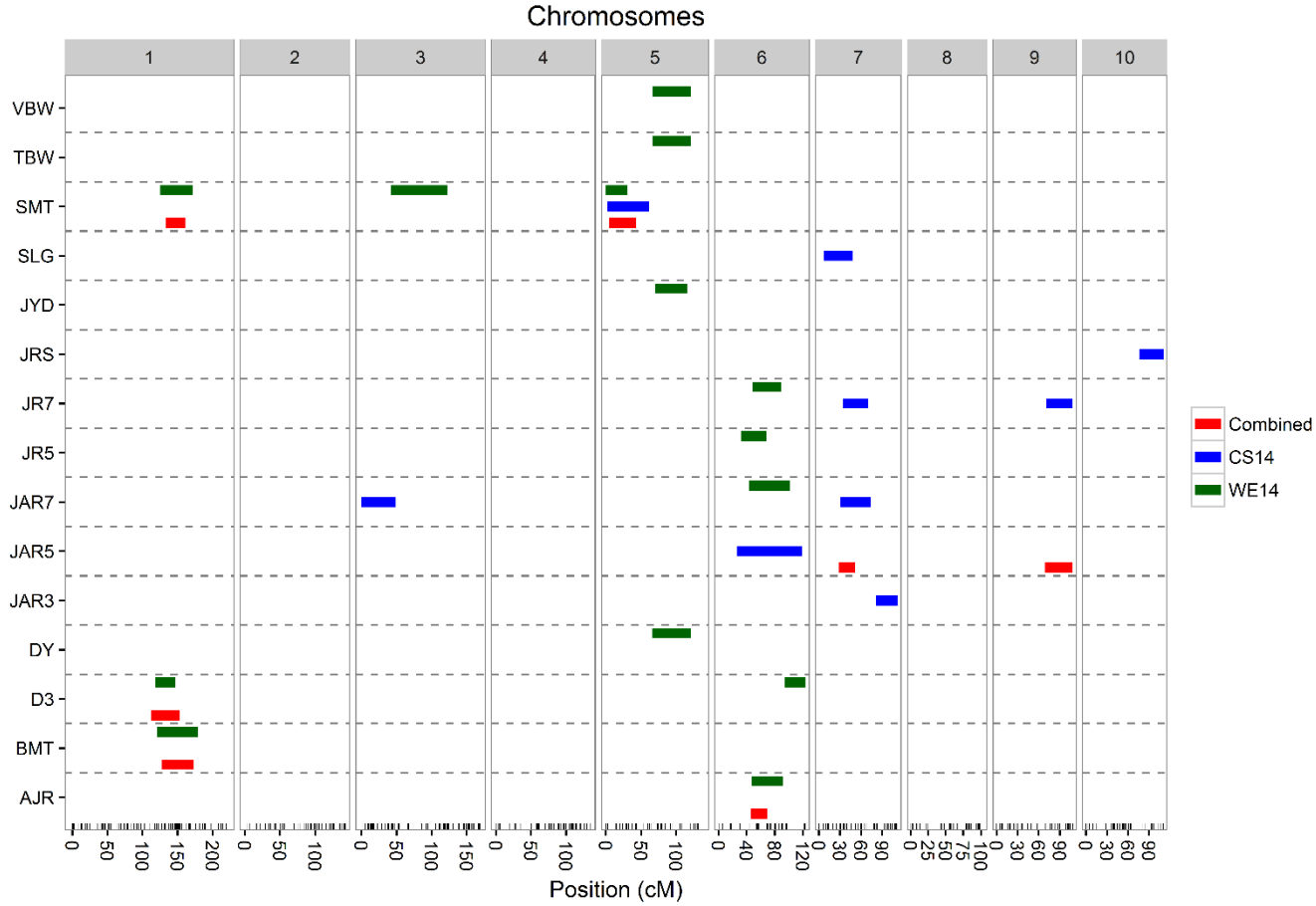
[a] CS14: College Station, TX, 2014, WE14: Weslaco, TX, 2014, Combined: multiple environments analysis

[b] QTL position corresponds to $-\log_{10}(p)$ peak for a given QTL

[c] Standardized phenotypic effects (measurement units are meaningless)

[d] Allele that increases the phenotype

Figure 8 Estimated locations of quantitative trait loci (QTLs) using best unbiased estimators (BLUEs) for individual trials (College Station 2014 = CS14 and Weslaco 2014 = WE14) and combined analysis (Combined) estimated from the BTx3197 x Rio $F_{3:4}$ population using composite interval mapping (CIM). The width of the bar estimates the QTL confidence interval estimated according to Weller and Soller (2004)



ajr = average of juice ratings, bmt = bagasse moisture, d1 = diameter of 1st internode, d3 = diameter of 3rd internode, dy = days to mid-anthesis, ht = plant height, jr3 = juice rating of 3rd internode, jr5 = juice rating of 5th internode, jr7 = juice rating of 7th internode, jrs = sum of juice ratings, jar3 = digital juice area of 3rd internode, jar5 = digital juice area of 5th internode, jar7 = digital juice area of 7th internode, jyd = juice yield, slg = stalk length, smt = stalk moisture, svl = stalk volume, tbw = total biomass weight, vbw = vegetative biomass weight

Multi-trait QTL analysis

Multi-trait QTL analysis revealed three distinct QTLs affecting diameter of the 3rd internode, stalk moisture, and bagasse moisture as assessed in Weslaco 2014 and from the combined analysis which appears to be pleiotropic (i.e. multiple traits were affected by the same QTL) (Table 7). The first QTL was detected at 148.70 cM on chromosome 1 and explained from 18.9 to 25.5% of the genetic variance for these traits with superior alleles coming from the parental line Rio. The second QTL affecting these traits was mapped at 14.08 cM on chromosome 5. This QTL explained 6.0% of the genetic variance for diameter of the 3rd internode in Weslaco 2014, between 19.0 to 25.5 % for stalk moisture, and between 7.6 to 9.5% for bagasse moisture when using BLUEs from the combined analysis or from Weslaco 2014, respectively. However, the superior alleles at the second QTL was donated by the parent BTx3197. The third QTL was mapped on chromosome 8 at 54.68 cM and had a significant effect only on diameter of the 3rd internode when evaluated in Weslaco 2014 and on stalk moisture as assessed in the combined analysis. This third QTL with high value alleles coming from BTx3197 explained 6.5 and 9.9% of the genetic variance for diameter of the 3rd internode and stalk moisture, respectively.

Finally, another QTL was mapped at 69.85 cM on chromosome 6. This QTL jointly affected juice ratings of the 5th and 7th internodes in Weslaco 2014 and College Station 2014, as well as the average of juice ratings as assessed in Weslaco 2014 and in the combined analysis. At this locus, the superior alleles for all traits came from the parent Rio, and the variance explained by it ranged from 12.1% for juice rating of the 5th

internode in Weslaco 2014 to 28.7% for juice rating of the 7th internode at the same environment.

Discussion

The RIL population evaluated in this study was derived from the inbred lines BTx3197 and Rio. BTx3197 is a 3-dwarf, photoperiod-insensitive line, while Rio is a 2-dwarf, intermediately photoperiod sensitive line. BTx3197 is a grain type line with early flowering and good general combining ability (King, et al., 1961), while Rio is highly resistant to leaf anthracnose and rust, and shows high sucrose yield (Broadhead, 1972). Yet, both lines presented similar expression for stalk pithiness. Because these lines are important sources of key agronomic and energy traits, the identification of QTL related to juice yield was intentionally undertaken in a RIL population that is highly relevant for breeding purposes aiming to develop superior energy sorghum cultivars.

The long day conditions in College Station favored the intermediate expression of photoperiod sensitivity in some RILs as suggested by Murray, et al. (2008), lengthening the duration of their vegetative phase. This phenomenon likely explains the greater average performance of the RIL population for most traits evaluated in College Station 2014. One of the exceptions was stem pithiness with expression being considerably higher in Weslaco 2014 based on digital image analysis. Pithiness development in the stem increases with longer periods of dehydration (Aloni and Pressman, 1981). Because the average precipitation in Weslaco 2014 was ~4.0-fold lower than in College Station 2014 (110 mm vs 449 mm), it is logical to hypothesize that the limited availability of water in

Weslaco 2014 could have increased the severity of pithiness. Flowering has also been associated with higher expression of pithiness in surgarcane (Eksteen, et al., 2014, Rao and Naresh Kumar, 2003). Although the exact mechanism driving this phenomenon is poorly understood in sorghum, the shorter photoperiod observed in Weslaco 2014 could act synergistically with the effects of drought stress to increase pithiness expression.

Multi-trait QTL analysis revealed a QTL affecting multiple measurements of stalk juiciness (inverse of stalk pithiness) in both environments and in the combined analysis at position 69.85 cM on chromosome 6. This peak position was located 145,284 bp away from one of the markers (bpsb069) flanking a stem moisture QTL (Han, et al., 2015), suggesting the QTL found for stem pithiness herein and for stem moisture in the Han, et al. (2015) study are likely the same. There are apparently multiple candidate genes related to cellular biological process, cellular metabolic process-related (Han, et al., 2015), and the gene known to confer mid rib type in sorghum, the *d* gene (Hart, et al., 2001), in this genomic region. Although one or many of these genes might influence the expression of stem pithiness in this population, the design of this study does not allow to certainly associate each specific gene or genes are the causative variant(s) of this phenotype. In a previous study performed by Felderhoff, et al. (2012) using a larger version of this population, no consistent QTL for stem juiciness based on phenotypic ratings was mapped on chromosome 6. In that study, tillering could have posed a large confounding effect on stalk pithiness affecting both the trait expression and phenotypic measurements; therefore, reducing the power to identifying such QTL. According to Aloni and Pressman (1981), pithiness in stem of tomatoes initiates in the upper part of the stem close to the plant´s

apex upon water stress proceeding down the stem as the stress was prolonged. Herein, no QTL for stem pithiness was found at this locus for internode 3. From internodes 5 to 7, there was an increase in the percentage of genetic variance explained by this locus, which suggests a heterogeneous expression of stem pithiness across internodes similar to the pattern observed in tomatoes.

Two other QTLs related to stem juiciness, one for juice rating of the 7th internode, and one for digital juicy area of the 7th internode were mapped on chromosome 7 in College Station 2014. Likewise, another QTL for juice rating of the 7th internode was also mapped on chromosome 9 in the same environment. These QTLs on chromosomes 7 and 9 also co-localized with a QTL found for digital juicy area of the 5th internode in the combined analysis. Altogether these results suggest that stem pithiness is not a monogenic trait in accordance with the findings of Felderhoff, et al. (2012). More importantly, some QTLs underlying the trait were specific to the College Station 2014 environment where there was higher quantity of water available and longer day lengths were observed. Although speculative, this indicates that adaptive genes might play a role modulating the expression of the gene or genes underlying stem pithiness on chromosome 6 according to the availability of water to plants and/or variation in photoperiod. However, because these QTLs were apparently mapped at these chromosomes for the first time, further QTL analysis in a larger number of environmental conditions is needed to validate the significance of such QTLs.

In accordance with Felderhoff, et al. (2012) there was no clear evidence at both the phenotypic and genetic level that pithiness expression influences moisture content within

the stalk. However, an apparent pleiotropic QTL affecting stem moisture, bagasse moisture, and stem diameter was detected on chromosome 1 in both Weslaco 2014 and in the combined analysis. Because stem moisture and diameter are major traits affecting both biomass and juice yield, marker assisted selection at this locus would allow improving two traits whose phenotypic association is rather weak. Although it is quite challenging to explain the genetic basis of such a relationship, the QTLs for such unrelated traits might be actually distinct. A flowering QTL had been identified at a similar location in the same population by Felderhoff, et al. (2012). Although the co-localization between this flowering QTL and the pleiotropic QTL found herein is not perfect, it is possible that the latter actually corresponds to a flowering locus which typically shows a large phenological impact on many different traits. Likewise, multi-trait QTL analysis revealed a second pleiotropic QTL affecting stem moisture, stalk moisture, and stem diameter. This QTL was consistent across environments, but in this case the QTL analysis suggested independence between this locus and a flowering locus mapped on the same chromosome.

The only QTLs found for biomass and juice yields were mapped at the exact same position as a flowering QTL on chromosome 5 from both single and multi-trait QTL analysis. Delayed flowering favors sorghum biomass production, a major trait underlying juice yield. In the studies of Felderhoff, et al. (2012), and Murray, et al. (2008), QTLs for biomass and juice yields were mapped in close proximity to distinct flowering loci. Although the QTLs found herein and in these other studies were mapped on distinct chromosomes, all of these studies suggest a pleiotropic role of flowering loci in affecting biomass and juice yields. Since phenotypic correlation analysis also revealed a strong

association between total biomass weight, vegetative biomass weight, and juice yield; there is solid evidence that both biomass and juice yield QTLs are actually conditioned by the flowering locus located on chromosome 5.

Although the alleles increasing stem juiciness at the pleiotropic QTL on chromosome 6 came from Rio (a sweet sorghum line), alleles from the grain parent BTx3197 improved stalk juiciness at QTLs mapped on chromosomes 7 and 9. Likewise, the same pattern of allelic inheritance was observed for other traits such as stalk moisture. Accordingly, there is now evidence that sorghum germplasm originally developed for grain production could possess alleles that enhances energy traits influencing juice yield. Because sorghum hybrid production requires seed parents with economic seed yield, combining suitable female grain lines with male sweet lines could be a fast and cheap strategy to developing superior sweet or dual purpose hybrids, at least initially.

Overall, there was little agreement between QTLs found in this study and those found in previous studies conducted in Texas, except for few QTLs found for highly heritable traits such as plant height and flowering time. Most of the traits overlapping between this current research and the studies of Felderhoff, et al. (2012), and Murray, et al. (2008) are quantitative. QTL analysis is greatly influenced by environmental signals (Bernardo, 2010), management, and phenotyping techniques. Each one of these studies were conducted in distinct years with particular weather patterns, especially for characteristics such as precipitation and temperature. This study was also the first of its kind where both population density and tillering were thoroughly controlled. Moreover, the generations of the populations used in each study were different, dominance effects

was not estimated herein, QTL analytical methods were distinct, and the population examined by Murray et al. (2008) was a cross between BTx623 x Rio. Finally, different population sizes and sampling were used in each study which also have a direct effect on QTL mapping results. After taking all these facts together, the inconsistency between QTLs are actually not surprising. This suggests that the genetic control of juice yield, and most of its component traits is governed by adaptive QTLs, with presence and magnitude of these QTLs varying greatly between experiments.

Due to this apparent complex genetic architecture, breeding for improved stem features underlying biomass and juice yields should target specific environments. In this context, marker assisted selection should facilitate the genetic manipulation of adaptive genetic loci whose effects are small and directly influenced by environmental signals. Despite the possible influence of drought and flowering on stem pithiness in sorghum being purely speculative at this point, further investigation is needed to better understand this phenomenon. Because, higher expression of stem pithiness implies less water per total plant biomass, this trait might be important in conferring resistance to drought and key when developing energy germplasm for high dry biomass content in sorghum. The real impact of stem pithiness on juice yield is still unknown, mainly due the confounding background effect of many other traits on juice yield. Therefore, development of near-isogenic lines and hybrids with dry and juicy stems would be key to reveal and quantify the real impact of stem pithiness on juice yield.

CHAPTER IV

CONCLUSIONS

There are several advantages for using sweet sorghum as a dedicated bioenergy feedstock. However, the first step to expand its use as such relies in improving its economic feasibility. Although, techniques to improving ethanol yield in sorghum lag behind sugarcane, there is a great potential to enhance the rate of genetic gain of sorghum ethanol yield by closely studying one of its major underlying traits, juice yield. This research revealed, via a coefficient path analysis, that stalk weight and stalk volume have the greatest total effect on juice yield, followed by stalk diameter, stalk length and plant height. The direct effects for stalk diameter on stalk volume and juice yield were positive and significantly greater than the direct effect for plant height on the same traits. According to the findings herein, juice yield could be successfully predicted by jointly assessing stalk weight and stalk moisture or stalk diameter, plant height, and stalk juiciness. All these traits were moderately to highly repeatable within and across environments. Taken together, this suggests that prediction equations or index selection models can be potentially used to accelerate phenotyping for juice yield which will improve the selection process. Directly selecting cultivars with higher stalk weight or with thicker and longer stalks should be a primary goal when improving both juice and biomass yields in sorghum.

Multiple QTLs were found for multiple stem-related traits including stem pithiness, stem diameter, and stem moisture. Overall, these QTLs were mapped at different

regions across the sorghum genome, and in some occasions they co-localized with known genes or QTLs for traits such as stem moisture, stem-mid rib type, and flowering time. Except for few QTLs underlying plant height and flowering time, there was little agreement between the QTLs found in this study and those found in previous studies conducted by investigators using the same RIL population or genetically related mapping populations. Therefore, results from this research suggest that the genetic control of juice yield, and a great number of its component traits is governed by adaptive QTLs, whose presence and magnitude vary greatly between experiments. Due to this apparent complex genetic architecture, breeding for improved stem features underlying biomass and juice yields should target specific environments. More importantly, it should take advantage of tools such as marker assisted selection to facilitate the genetic manipulation of adaptive genetic loci whose effects are relatively small and directly influenced by environmental signals.

REFERENCES

Alimi, N.A., M.C. Bink, J.A. Dieleman, J.J. Magan, A.M. Wubs, A. Palloix, et al. 2013. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theor Appl Genet 126: 2597-2625. doi:10.1007/s00122-013-2160-3.

Aloni, B. and E. Pressman. 1981. Stem pithiness in tomato plants: the effect of water stress and the role of abscisic acid. Physiologia Plantarum 51: 39-44. doi:10.1111/j.1399-3054.1981.tb00876.x.

Asif, M. and T. Muneer. 2007. Energy supply, its demand and security issues for developed and emerging economies. Renewable and Sustainable Energy Reviews 11: 1388-1413. doi:10.1016/j.rser.2005.12.004.

Audilakshmi, S., A.K. Mall, M. Swarnalatha and N. Seetharama. 2010. Inheritance of sugar concentration in stalk (brix), sucrose content, stalk and juice yield in sorghum. Biomass and Bioenergy 34: 813-820. doi:10.1016/j.biombioe.2010.01.025.

Bell, M., A. Garside, N. Halpin and J. Berthelsen. 2004. Interactions between stalk number and stalk weight and the implications for cane yield. Proceedings of the Australian Society of Sugar Cane Technologists 26: 12.

Bernardo, R.N. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury, MN.

Boer, M.P., D. Wright, L. Feng, D.W. Podlich, L. Luo, M. Cooper, et al. 2007. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data

using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177: 1801-1813.

Boivin, K., M. Deu, J.-F. Rami, G. Trouche and P. Hamon. 1999. Towards a saturated sorghum map using RFLP and AFLP markers. Theor Appl Genet 98: 320-328. doi:10.1007/s001220051076.

Bollen, K.A. 1989. Structural equations with latent variables. 2nd ed. John Wiley & Sons, New York.

Broadhead, D.M. 1972. Registration of Rio sweet sorghum. Crop Science 12: 716.

Broadhead, D.M. and K.C. Freeman. 1980. Stalk and sugar yield of sweet sorghum as affected by spacing. Agron J 72: 523-524. doi:10.2134/agronj1980.00021962007200030026x.

Brown, A., J. Daniels and B. Latter. 1969. Quantitative genetics of sugarcane: II. Correlation analysis of continuous characters in relation to hybrid sugarcane breeding. Theor Appl Genetics 39: 1-10.

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.

Bull, T.A. and K.T. Glasziou. 1963. The evolutionary significance of sugar accumulation in *Saccharum*. Australian Journal of Biological Sciences 16: 737-742. doi:10.1071/BI9630737.

Burks, P.S., C.M. Kaiser, E.M. Hawkins and P.J. Brown. 2015. Genomewide association for sugar yield in sweet sorghum. Crop Science 55: 2138-2148. doi:10.2135/cropsci2015.01.0057.

Calviño, M. and J. Messing. 2012. Sweet sorghum as a model system for bioenergy crops. Current Opinion in Biotechnology 23: 323-329. doi:10.1016/j.copbio.2011.12.002.

Caravetta, G.J., J.H. Cherney and K.D. Johnson. 1990. Within-row spacing influences on diverse sorghum genotypes: I. Morphology. Agron J 82: 206-210. doi:10.2134/agronj1990.00021962008200020007x.

Childs, K.L., M.-M. Cordonnier-Pratt, L.H. Pratt and P.W. Morgan. 1992. Genetic regulation of development in *Sorghum bicolor* VII. *ma3*^{R} flowering mutant lacks a phytochrome that predominates in green tissue. Plant Physiology 99: 765-770.

Childs, K.L., F.R. Miller, M.-M. Cordonnier-Pratt, L.H. Pratt, P.W. Morgan and J.E. Mullet. 1997. The sorghum photoperiod sensitivity gene, *Ma3*, encodes a phytochrome B. Plant Physiology 113: 611-619.

Corn, R.J. 2009. Heterosis and composition of sweet sorghum. Texas A&M University, College Station, TX.

Davila-Gomez, F.J., C. Chuck-Hernandez, E. Perez-Carrillo, W.L. Rooney and S.O. Serna-Saldivar. 2011. Evaluation of bioethanol production from five different varieties of sweet and forage sorghums (*Sorghum bicolor* (L) Moench). Industrial Crops and Products 33: 611-616. doi:10.1016/j.indcrop.2010.12.022.

Eksteen, A., E. Halse, E. Simwinga and D. Sutherland. 2014. An investigation into factors influencing flowering and pithing of sugarcane in southern africa. Proc S Afr Sug Technol Ass 87: 493-496.

Felderhoff, T.J., S.C. Murray, P.E. Klein, A. Sharma, M.T. Hamblin, S. Kresovich, et al. 2012. QTLs for energy-related traits in a sweet × grain sorghum [*Sorghum bicolor* (L.) Moench] mapping population. Crop Science 52: 2040-2049. doi:10.2135/cropsci2011.11.0618.

Ferraris, R. and D. Charles-Edwards. 1986. A comparative analysis of the growth of sweet and forage sorghum crops. I. Dry matter production, phenology and morphology. Australian Journal of Agricultural Research 37: 495-512. doi:10.1071/AR9860495.

Gerik, T.J. and C.L. Neely. 1987. Plant density effects on main culm and tiller development of grain sorghum. Crop Science 27: 1225-1230. doi:10.2135/cropsci1987.0011183X002700060027x.

Godoy, J.G.V. and T.T. Tesso. 2013. Analysis of juice yield, sugar content, and biomass accumulation in sorghum. Crop Science 53: 1288-1297. doi:10.2135/cropsci2012.04.0217.

Gravois, K.A., S.B. Milligan and F.A. Martin. 1990. The role of pith, tube, and stalk density in determining sugarcane sucrose content and stalk weight. Theor Appl Genet 79: 273-277. doi:10.1007/BF00225963.

Grof, C.P.L. and J.A. Campbell. 2001. Sugarcane sucrose metabolism: scope for molecular manipulation. Functional Plant Biology 28: 1. doi:10.1071/pp00039.

Guan, Y., H. Wang, L. Qin, H. Zhang, Y. Yang, F. Gao, et al. 2011. QTL mapping of bio-energy related traits in *Sorghum*. Euphytica 182: 431-440. doi:10.1007/s10681-011-0528-5.

Hammer, G. 2006. Pathways to prosperity: breaking the yield barrier in sorghum. Agricultural Science 19: 16-22.

Han, Y., P. Lv, S. Hou, S. Li, G. Ji, X. Ma, et al. 2015. Combining next generation sequencing with bulked segregant analysis to fine map a stem moisture locus in sorghum (*Sorghum bicolor* L. Moench). PloS one 10: e0127065.

Hart, E.G., F.K. Schertz, Y. Peng and H.N. Syed. 2001. Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103: 1232-1242. doi:10.1007/s001220100582.

Hilson, G.R. 1916. A note on the inheritance of certain stem characters in sorghum. Agric J India 11: 150-155.

International Energy Agency. 2015. Global energy trends to 2040. Business-as-usual or a brave new world? World Energy Outlook 2015. France. p. 53-109.

Jackson, P.A. 2005. Breeding for improved sugar content in sugarcane. Field Crops Research 92: 277-290. doi:10.1016/j.fcr.2005.01.024.

Jansen, R.C. and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136: 1447-1455.

Kang, M.S., O. Sosa and J.D. Miller. 1989. Path analyses for percent fiber, and cane and sugar yield in sugarcane. Crop Science 29: 1481-1483. doi:10.2135/cropsci1989.0011183X002900060032x.

Kim, M. and D.F. Day. 2011. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. Journal of Industrial Microbiology & Biotechnology 38: 803-807. doi:10.1007/s10295-010-0812-8.

King, J.G., J.R. Quinby, J.C. Stephens, N.W. Kramer and K.A Lahr. 1961. An evaluation of parents of grain sorghum hybrids.TAES, TX.

Kosambi, D.D. 1943. The estimation of map distances from recombination values. Ann Eugenic 12: 172-175.

Lander, E.S. and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Li, J. and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95: 221-277. doi:10.1038/sj.hdy.6800717.

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.

Liu, D.L. and K.R. Helyar. 2003. Simulation of seasonal stalk water content and fresh weight yield of sugarcane. Field Crops Research 82: 59-73. doi:10.1016/s0378-4290(03)00019-4.

Lv, P., G. Ji, Y. Han, S. Hou, S. Li, X. Ma, et al. 2013. Association analysis of sugar yield-related traits in sorghum [*Sorghum bicolor* (L.)]. Euphytica 193: 419-431. doi:10.1007/s10681-013-0962-7.

Mace, E.S., L. Xia, D.R. Jordan, K. Halloran, D.K. Parh, E. Huttner, et al. 2008. DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC genomics 9: 26. doi: 10.1186/1471-2164-9-26

Makanda, I., P. Tongoona and J. Derera. 2009. Quantification of genotypic variability for stem sugar accumulation and associated traits in new sweet sorghum varieties. African Crop Science Conference Proceedings 9: 391-398.

Malosetti, M., J.M. Ribaut and F.A. van Eeuwijk. 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4: 44. doi:10.3389/fphys.2013.00044.

Malosetti, M., J.M. Ribaut, M. Vargas, J. Crossa and F.A. Van Eeuwijk. 2008. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 161: 241-257.

Miller, J. and N. James. 1974. The influence of stalk density on cane yield. Proc Int Soc Sugar Cane Technol 15: 177-183.

Miller, J.D. 1977. Combining ability and yield component analyses in a five-parent diallel cross in sugarcane. Crop Science 17: 545-547. doi:10.2135/cropsci1977.0011183X001700040016x.

Monk, R.L., F.R. Miller and G.G. McBee. 1984. Sorghum improvement for energy production. Biomass 6: 145-153. doi:10.1016/0144-4565(84)90017-9.

Morishige, D.T., P.E. Klein, J.L. Hilley, S.M. Sahraeian, A. Sharma and J.E. Mullet. 2013. Digital genotyping of sorghum - a diverse plant species with a large repeat-rich genome. BMC genomics 14: 448. doi:10.1186/1471-2164-14-448.

Mullet, J.E. and W.L. Rooney. 2013. Method for production of sorghum hybrids with selected flowering times. US Patent Application No. 13/886, 130.

Mullet, J.E., W.L. Rooney, P.E. Klein, D. Morishige, R. Murphy and J.A. Brady. 2012. Discovery and utilization of sorghum genes (*MA5/MA6*). US Patent 8,309,793.

Murphy, R.L., R.R. Klein, D.T. Morishige, J.A. Brady, W.L. Rooney, F.R. Miller, et al. 2011. Coincident light and clock regulation of *pseudoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum. Proceedings of the National Academy of Sciences 108: 16469-16474.

Murphy, R.L., D.T. Morishige, J.A. Brady, W.L. Rooney, S. Yang, P.E. Klein, et al. 2014. *Ghd7* (*Ma6*) represses sorghum flowering in long days: *Ghd7* alleles enhance biomass accumulation and grain production. The Plant Genome 7(2). doi: 0.3835/plantgenome2013.11.0040

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell and S. Kresovich. 2009. Sweet sorghum genetic diversity and association mapping for brix and height. The Plant Genome 2: 48-62.

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

Packer, D.J. and W.L. Rooney. 2014. High-parent heterosis for biomass yield in photoperiod-sensitive sorghum hybrids. Field Crops Research 167: 153-158. doi:10.1016/j.fcr.2014.07.015.

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, et al. 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457: 551-556. doi:10.1038/nature07723.

Pereira, M.G. and M. Lee. 1995. Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90: 380-388. doi:10.1007/bf00221980.

Pfeiffer, T.W., M.J. Bitzer, J.J. Toy and J.F. Pedersen. 2010. Heterosis in sweet sorghum and selection of a new sweet sorghum hybrid for use in syrup production in Appalachia. Crop Science 50: 1788-1794. doi:10.2135/cropsci2009.09.0475.

Prasad, S., A. Kumar and K.S. Muralikrishna. 2013. Assessment of ethanol yield associated characters in sweet sorghum. Maydica 58: 299-303.

Quinby, J. and R. Karper. 1945. Inheritance of three genes that influence time of floral initiation and maturity date in milo. Journal of the American Society of Agronomy.

Quinby, J.R. 1966. Fourth maturity gene locus in sorghum. Crop Science 6: 516-518. doi:10.2135/cropsci1966.0011183X000600060005x.

Quinby, J.R. 1972. Influence of maturity genes on plant growth in sorghum. Crop Science 12: 490-492. doi:10.2135/cropsci1972.0011183X001200040030x.

Quinby, J.R. 1967. The maturity genes of sorghum. In: A. G. Norman, editor Advances in Agronomy. Academic Press. p. 267-305.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statitical Computing, Vienna, Austria.

Rao, P.N.G. and K. Naresh Kumar. 2003. Effect of flowering on juice quality and fibre content in sugarcane. Sugar Tech 5: 185-187. doi:10.1007/bf02943633.

Reddy, B.V., S. Ramesh, P.S. Reddy, B. Ramaiah, P. Salimath and R. Kachapur. 2005. Sweet sorghum – A potential alternate raw material for bio-ethanol and bioenergy. SAT eJournal 1: 1-8.

Ritter, K.B., D.R. Jordan, S.C. Chapman, I.D. Godwin, E.S. Mace and C. Lynne McIntyre. 2008. Identification of QTL for sugar-related traits in a sweet x grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. Molecular Breeding 22: 367-384. doi:10.1007/s11032-008-9182-6.

Rooney, W.L. and S. Aydin. 1999. Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Science 39: 397-400. doi:10.2135/cropsci1999.0011183X0039000200016x.

Rooney, W.L., J. Blumenthal, B. Bean and J.E. Mullet. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioproducts and Biorefining 1: 147-157. doi:10.1002/bbb.15.

Rutto, L.K., Y. Xu, M. Brandt, S. Ren and M.K. Kering. 2013. Juice, ethanol, and grain yield potential of five sweet sorghum (*Sorghum bicolor* [L.] Moench) cultivars. Journal of Sustainable Bioenergy Systems 03: 113-118. doi:10.4236/jsbs.2013.32016.

Salas Fernandez, M.G., P.W. Becraft, Y. Yin and T. Lübberstedt. 2009. From dwarves to giants? Plant height manipulation for biomass yield. Trends Plant Sci 14: 454-461. doi:10.1016/j.tplants.2009.06.005.

Schaffert, R. 1992. Sweet sorghum substrate for industrial alcohol. In: M. I. Gomez, L. R. House, L. W. Rooney and D. A. V. Dendy, editors, Utilization of sorghum and millets. Patancheru, A. P. 502 324, India: International Crops Research Institute for the Semi-Arid Tropics. p. 131-137.

Schaffert, R.E. and L.M. Gourley. 1982. Sorghum as an energy source. Sorghum in the eighties: proceedings of the international symposium on sorghum, Pantacheru, A.P., ICRISAT.

Schneider, C.A., W.S. Rasband and K.W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9: 671-675. doi:10.1038/nmeth.2089.

Shiringani, A.L. and W. Friedt. 2011. QTL for fibre-related traits in grain x sweet sorghum as a tool for the enhancement of sorghum as a biomass crop. Theor Appl Genet 123: 999-1011. doi:10.1007/s00122-011-1642-4.

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.

Siirola, J.J. 2014. Speculations on global energy demand and supply going forward. Current Opinion in Chemical Engineering 5: 96-100. doi:10.1016/j.coche.2014.07.002.

Smith, G.A., M.O. Bagby, R.T. Lewellan, D.L. Doney, P.H. Moore, F.J. Hills, et al. 1987. Evaluation of sweet sorghum for fermentable sugar production potential. Crop Science 27. doi:10.2135/cropsci1987.0011183X002700040037x.

Smith, G.A. and D.R. Buxton. 1993. Temperate zone sweet sorghum ethanol production potential. Bioresource Technology 43: 71-75. doi:10.1016/0960-8524(93)90086-q.

Stephens, J.C. and J.R. Quinby. 1939. The *D Rs P* linkage group in sorghum. Journal of Agricultural Research 59: 725-730.

Stickler, F.C. and A.W. Pauli. 1961. Influence of date of planting on yield and yield components in grain sorghum. Agron J 53: 20. doi:10.2134/agronj1961.00021962005300010007x.

Swanson, A.F. and J.H. Parker. 1931. Inheritance of smut resitance and juiciness of stalk. The Journal of Heredity 22: 51-56.

Upadhyaya, H.D., Y.-H. Wang, C.L.L. Gowda and S. Sharma. 2013. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. Theor Appl Genet 126: 2003-2015. doi:10.1007/s00122-013-2113-x.

Van Ooijen, J.W. 2006. JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Wageningen, The Netherlands: Kyazma B. V..

Vasilakoglou, I., K. Dhima, N. Karagiannidis and T. Gatsis. 2011. Sweet sorghum productivity for biofuels under increased soil salinity and reduced irrigation. Field Crops Research 120: 38-46. doi:10.1016/j.fcr.2010.08.011.

VSN International. 2014. GenStat for Windows 17th Edition. VSN International, Hemel Hempstead, UK.

Wang, L., S. Jiao, Y. Jiang, H. Yan, D. Su, G. Sun, et al. 2013. Genetic diversity in parent lines of sweet sorghum based on agronomical traits and SSR markers. Field Crops Research 149: 11-19. doi:10.1016/j.fcr.2013.04.013.

Wang, M.L., C. Zhu, N.A. Barkley, Z. Chen, J.E. Erpelding, S.C. Murray, et al. 2009. Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor Appl Genet 120: 13-23. doi:10.1007/s00122-009-1155-6.

Weller, J.I. and M. Soller. 2004. An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. Theor Appl Genet 109: 1224-1229. doi:10.1007/s00122-004-1664-2.

Woods, J. 2001. The potential for energy production using sweet sorghum in southern Africa. Energy for Sustainable Development 5: 31-38. doi:10.1016/s0973-0826(09)60018-1.

Worley, J.W., J.S. Cundiff, D.H. Vaughan and D.J. Parrish. 1991. Influence of sweet sorghum spacing on stalk pith yield. Bioresource Technology 36: 133-139. doi: 10.1016/0960-8524(91)90170-O.

Wu, X., S. Staggenborg, J.L. Propheter, W.L. Rooney, J. Yu and D. Wang. 2010. Features of sweet sorghum juice and their performance in ethanol fermentation. Industrial Crops and Products 31: 164-170. doi:10.1016/j.indcrop.2009.10.006.

Yang, S., R.L. Murphy, D.T. Morishige, P.E. Klein, W.L. Rooney and J.E. Mullet. 2014. Sorghum phytochrome B inhibits flowering in long days by activating expression of *SbPRR37* and *SbGHD7*, repressors of *SbEHD1*, *SbCN8* and *SbCN12*. PLoS One 9: e105352. doi:10.1371/journal.pone.0105352.

Yu, J., T. Zhang, J. Zhong, X. Zhang and T. Tan. 2012. Biorefinery of sweet sorghum stem. Biotechnol Adv 30: 811-816. doi:10.1016/j.biotechadv.2012.01.014.

Yuan, J.S., K.H. Tiller, H. Al-Ahmad, N.R. Stewart and C.N. Stewart, Jr. 2008. Plants to power: bioenergy to fuel the future. Trends Plant Sci 13: 421-429. doi:10.1016/j.tplants.2008.06.001.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Zhang, C., G. Xie, S. Li, L. Ge and T. He. 2010. The productive potentials of sweet sorghum ethanol in China. Applied Energy 87: 2360-2368. doi:10.1016/j.apenergy.2009.12.017.

Zhao, Y.L., A. Dolat, Y. Steinberger, X. Wang, A. Osman and G.H. Xie. 2009. Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. Field Crops Research 111: 55-64. doi:10.1016/j.fcr.2008.10.006.

Zheng, L.Y., X.S. Guo, B. He, L.J. Sun, Y. Peng, S.S. Dong, et al. 2011. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biology 12: R114. doi:10.1186/gb-2011-12-11-r114.

Zou, G., S. Yan, G. Zhai, Z. Zhang, J. Zou and Y. Tao. 2011. Genetic variability and correlation of stalk yield-related traits and sugar concentration of stalk juice in a

sweet sorghum (*Sorghum bicolor* L. Moench) population. Australian Journal of Crop Science 5: 1232-1238.

APPENDIX

Table A-1 Summary information of genotypes evaluated. Entry number, pedigree, type and use class are shown

Entry	Pedigree	Maturity	Type	Use class
1	B.Tx623	PI	Inbred	Grain
2	Wray	PI	Inbred	Sweet
3	Della	PI	Inbred	Sweet
4	Rio	MPS	Inbred	Sweet
5	R.07007	PI	Inbred	Biomass
6	B.Tx635	PI	Inbred	Grain
7	B.TX3197	PI	Inbred	Grain
8	R.Tx436	PI	Inbred	Grain
9	B.Tx631	PI	Inbred	Grain
10	M81E	MPS	Inbred	Sweet
11	R.11434	PS	Inbred	Biomass
12	R.11461	PS	Inbred	Biomass
13	R.10135	PS	Inbred	Biomass
14	GRASSL	MPS	Inbred	Biomass
15	ATx631/R07007	PS	Hybrid	Biomass
16	ATx631/Tx2909	PS	Hybrid	Forage / Biomass
17	R.10030	PS	Inbred	Biomass
18	R.11436	PS	Inbred	Biomass
19	R.11450	PS	Inbred	Biomass
20	R.11438	PS	Inbred	Biomass

PI = photoperiod insensitive (flowering is independent of day length); PS = photoperiod sensitive (flowering is induced at day lengths equal or shorter than 12 h 15 min); MPS = moderately photoperiod sensitive (flowering is induced at day lengths between 13 h and 12 h and 15 min)

Table A-2 Pearson´s phenotypic correlation coefficients between eighteen traits analyzed across three Texas environments (lower diagonal) with respective significance test (upper diagonal)

Traits	JYD	SWT	SVL	D1	D3	DN	SLG	HT	SMT	SJR	AJR	SDS	JAR3	JAR5	JAR7	VJR3	VJR5	VJR7
JYD		**	**	**	**	**	**	**	**	**	*	**	**	**	**	**	ns	ns
SWT	0.88		**	**	**	**	**	**	**	**	ns	**	*	ns	ns	*	ns	ns
SVL	0.82	0.89		**	**	**	**	**	**	**	ns	ns	ns	ns	ns	*	*	*
D1	0.73	0.71	0.85		**	**	**	**	**	**	ns	**	**	*	ns	**	ns	ns
D3	0.76	0.77	0.87	0.97		**	**	**	**	**	ns	ns	**	ns	ns	**	ns	ns
DN	0.61	0.78	0.74	0.64	0.71		**	**	**	**	ns	ns	*	ns	ns	*	ns	ns
SLG	0.58	0.63	0.76	0.46	0.51	0.30		**	ns	**	**	**	ns	**	**	ns	**	**
HT	0.61	0.66	0.73	0.47	0.50	0.24	0.84		ns	**	**	**	ns	**	**	ns	**	**
SMT	0.57	0.44	0.39	0.52	0.53	0.52	-0.03	0.00		**	**	ns	**	**	**	**	**	**
SJR	0.56	0.52	0.41	0.51	0.51	0.27	0.26	0.27	0.43		**	*	**	**	**	**	**	**
AJR	0.17	0.07	-0.11	0.10	0.10	0.04	-0.29	-0.28	0.42	0.71		*	**	**	**	**	**	**
SDS	0.27	0.36	0.00	-0.21	-0.11	0.10	0.19	0.21	0.10	0.18	0.19		ns	*	*	ns	ns	ns
JAR3	0.27	0.17	0.14	0.31	0.29	0.18	-0.10	-0.08	0.47	0.48	0.61	0.05		**	**	**	**	**
JAR5	0.21	0.07	-0.06	0.15	0.13	-0.01	-0.25	-0.22	0.47	0.58	0.85	0.16	0.81		**	**	**	**
JAR7	0.20	0.04	-0.10	0.10	0.08	-0.03	-0.28	-0.27	0.43	0.64	0.89	0.19	0.69	0.95		**	**	**
VJR3	0.25	0.17	0.15	0.35	0.32	0.19	-0.12	-0.10	0.45	0.48	0.58	-0.03	0.92	0.74	0.61		**	**
VJR5	0.09	-0.03	-0.16	0.06	0.03	-0.06	-0.33	-0.32	0.36	0.52	0.85	0.09	0.71	0.95	0.91	0.68		**
VJR7	0.09	-0.03	-0.18	0.02	0.00	-0.06	-0.32	-0.33	0.32	0.59	0.86	0.14	0.58	0.85	0.94	0.51	0.88	

** significant at 1%, * significant at 5%, ns = non-significant at 5%

Table A-3 Path analysis: unfolding of phenotypic correlations into components of direct and indirect effects, effect significance, and residual effects of five traits measured in twenty diverse sorghum genotypes across three environments in Texas. Effects of stalk volume (SVL), and visual juice rating of seventh internode (VJR7) on juice yield (JYD) as well as the effect of diameter of third internode (D3), and plant height (HT) on stalk volume (SVL) are shown

Models	Trait	Mode of action	Coefficient	Significance
Model A-1	*SVL*	Direct effect on *JYD*	0.86	***
		Indirect effect via *VJR7*	-0.04	*
		Total effect (direct and indirect)	0.82	***
	VJR7	Direct effect on *JYD*	0.25	***
		Indirect effect via *SVL*	-0.15	*
		Total effect (direct and indirect)	0.09	***
		Residual	0.27	
Model A-2	*HT*	Direct effect on *SVL*	0.37	***
		Indirect effect via *D3*	0.34	***
		Total effect (direct and indirect)	0.71	***
	D3	Direct effect on *SVL*	0.69	***
		Indirect effect via *HT*	0.18	***
		Total effect (direct and indirect)	0.87	***
		Residual	0.13	

*** significant at 0.1%, ** significant at 1%, * significant at 5%, ns = non-significant at 5%

Table A-4 PCA of traits evaluated. Correlation of the variables with the first two components are shown.

Trait	PCA1	PCA2
ajr	-0.28	0.90
bmt	-0.28	-0.27
d1	-0.77	-0.05
d3	-0.84	-0.12
dy	-0.25	-0.14
ht	-0.65	-0.27
jr3	-0.28	0.60
jr5	0.06	0.84
jr7	-0.25	0.74
jrs	-0.75	0.51
jyd	-0.88	-0.20
slg	-0.57	-0.30
smt	-0.39	-0.33
svl	-0.91	-0.26
tbw	-0.94	-0.15
vwt	-0.93	-0.16
jar3	-0.31	0.59
jar5	-0.21	0.55
jar7	-0.32	0.57

ajr = average of juice ratings, bmt = bagasse moisture, d1 = diameter of 1st internode, d3 = diameter of 3rd internode, dy = days to mid-anthesis, ht = plant height, jr3 = juice rating of 3rd internode, jr5 = juice rating of 5th internode, jr7 = juice rating of 7th internode, jrs = sum of juice ratings, jar3 = digital juice are of 3rd internode, jar5 = digital juice area of 5th internode, jar7 = digital juice area of 7th internode, jyd = juice yield, slg = stalk length, smt = stalk moisture, svl = stalk volume, tbw = total biomass weight, vbw = vegetative biomass weight

Figure A-1 Stalk cross sections of multiple internodes of distinct genotypes. Images were captured by using a flatbed scanner and analyzed with ImageJ. White and green areas correspond to pithy (dry) and juicy tissue areas, respectively. Numbers on the bottom of each cross section represents the juiciness visual ratings. In this case 1 represents an interval of 0% to 10% of juicy stalk cross section area, 9 representing 90% to 100% of juicy stalk cross section area, and each interval in between representing a 10% increase for juicy area percentage (juiciness)

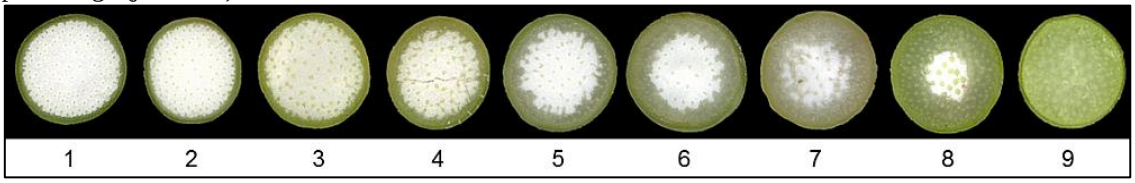


Figure A-2 Individual visual juice ratings versus correspondent juicy area estimates as assessed through digital image analysis using ImageJ. Summarized distribution of stem juicy area (%) are shown. A scale from 1 to 9 was adopted, with 1 representing 0% to 10% of juicy area within the internode cross section, 9 representing 90% to 100% of juicy area, and each interval in between representing a 10% increase for juiciness. Range of values, first and third quartile, and median are represented by the length of solid vertical black lines, bottom and top boxes' edges, and horizontal black line, respectively. N = 1260

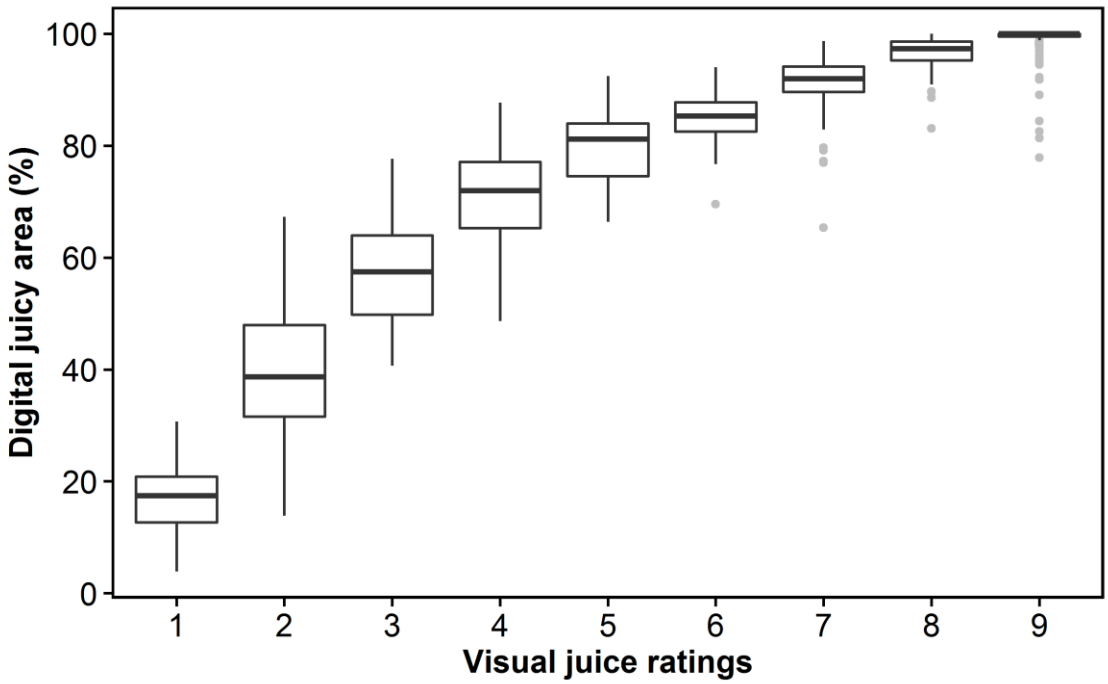


Figure A-3 Frequency of R^2 values **(a)** and average prediction errors **(b)** calculated after fitting two models 1000 times each upon resampling of training and testing sets. Model L1 corresponds to juice yield = stalk weight + stalk moisture, and model L2 corresponds to juice yield = diameter of third internode + plant height + visual juice rating of seventh internode. Training and testing sets sizes were equal to 75 and 25 percent of total data points collected, respectively. N = 139. Purple highlighting indicates intersection values between both models

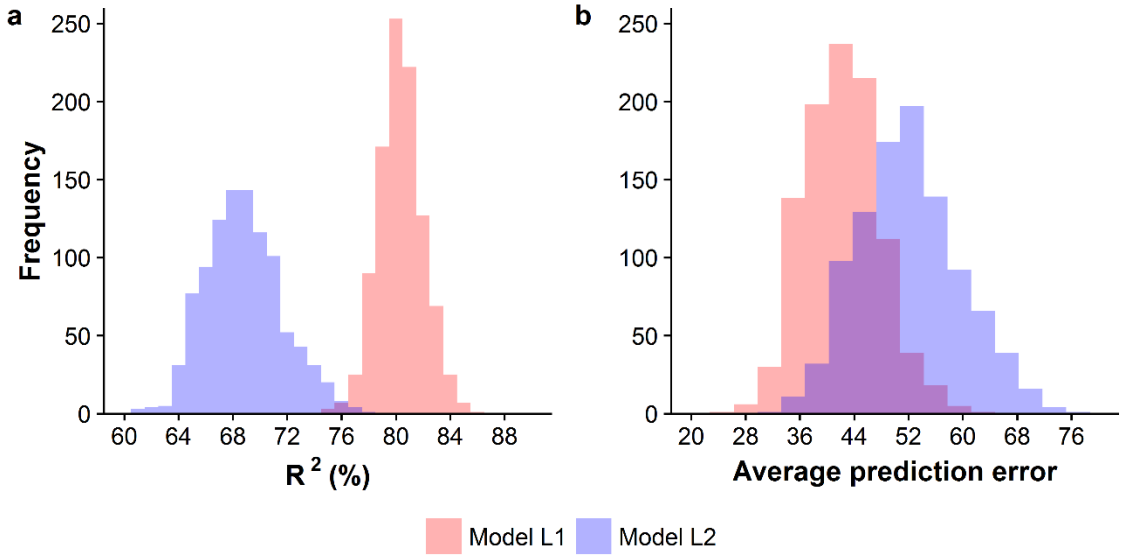


Figure A-4 Scatter plot of R^2 versus average prediction errors of models L1 **(a)** and L2 **(b)** after fitting both models 1000 times upon resampling of training and test sets. Model L1 corresponds to juice yield = stalk weight + stalk moisture, and model L2 corresponds to juice yield = diameter of third internode + plant height + visual juice rating of seventh internode. Training and testing sets sizes were equal to 75 and 25 percent of total data points collected, respectively. N = 139. Horizontal and vertical dashed red lines depict the mean value for R^2 and average prediction error, respectively