Sorghum: Genetic Improvement for Biofuel

Yanna Liang

Department of Civil and Environmental Engineering, 1230 Lincoln Dr., Southern Illinois University Carbondale, Carbondale, IL 62901, USA

Corresponding author: *liang@engr.siu.edu

INTRODUCTION

During recent years, sorghum has drawn intensive attention from the scientific community and policy decision-makers. Three categories of sorghum, grain, forage, and sweet sorghum can be used differently for producing bioethanol and other kinds of biofuels. While grains from grain sorghum and juice from sweet sorghum are readily fermentable, the sorghum stalk which is a lignocellulosic material is recalcitrant. To utilize sorghum as a bioenergy crop, however, the biomass quality and yield must be enhanced. Highly fermentable and highly digestible lines are extremely desirable but research on this aspect is lagging behind those of other crops, such as maize. With numerous genes affecting sorghum quality being known, it is reasonable to expect that sorghum can be improved to serve as a biomass plant well through genetic and molecular breeding. Over the years, genetic modification of sorghum has been mainly conducted in three areas: 1) reducing lignin content by establishing brown midrib mutants and understanding the genetic mechanisms for lignin reduction; 2) developing cold tolerant progenies and identifying quantitative trait loci (QTL) related to cold tolerance; and 3) investigating sugar related traits with the aim of increasing contents of sugar and nonstructural carbohydrates. This review provides detailed information on how these studies are carried out. Future research directions are also recommended for transforming sorghum from a promising crop to a truly invaluable plant in the biofuel arena.

Keywords: bioenergy, cold tolerance, lignin content, sugar content

Abbreviations: bmr, brown midrib; CAD, cinnamylalcohol dehydrogenase; CIM, composite interval mapping; COMT, caffeic acid-O-methyltransferase; GAX, glucorono-arabinoxylan; ELP, expression level polymorphisms; QTL, quantitative trait loci; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphisms; RIL, recombinant inbred lines; SAM, S-adenosyl-methionine; SFP, single-feature polymorphisms; SIM, simple interval mapping; SNP, single nucleotide polymorphisms; SSR, small sequence repeat

CONTENTS

INTRODUCTION..............................................................1
MODIFYING LIGNIN CONTENT..............................................2
IMPROVING COLD TOLERANCE...........................................3
ENHANCING SUGAR CONTENT..........................................4
FUTURE RESEARCH RECOMMENDATION.........................7
ACKNOWLEDGEMENTS................................................7
REFERENCES.................................................................7

INTRODUCTION

Sorghum (Sorghum bicolor), a diploid C4 grass (2n = 2x = 20), is the fifth most cultivated cereal crop in the world (Paterson 2008). Several attributes make sorghum stand out as a potential bioenergy crop:

1. No strong competition between food and fuel exists (Yuan et al. 2008).
2. Sorghum is drought and heat tolerant (Tuck et al. 2006; Almodares and Hadi 2009; Ratnavathi et al. 2010). When water is available, sorghum grows rapidly. During period of water shortage, sorghum tends to stop growing. The high tolerance of drought is due to an extensive root system that can penetrate 1.5 to 2.5 m into the soil and extend 1 m away from the stalk (Vermerris et al. 2007). Corn roots, however, extend only 0.8 m into the soil and extend 0.5 m from the stem (Pellerin and Pagès 1996).
3. Removal of sorghum stalk from the land does not contribute to depletion of soil organic matter owing to the sizable root materials (Wilhelm et al. 2004).
4. Sorghum requires less fertilizer than corn (Prasad et al. 2007).
5. Sorghum can tolerate a wide span of soil conditions, from light sand to heavy clay soil with a broad pH range between 5.0 and 8.5. This tolerance enables its growth on marginal land not suitable for other crops (Smith and Fredericksen 2000).
6. Agricultural practices for corn and sorghum are identical, thus requiring only limited adaptations on the part of farmers (Vermerris et al. 2007).
7. Sorghum has a rich germplasm resource (Vermerris et al. 2007; Zhang et al. 2010).
8. Together with switchgrass (Panicum virgatum) and Miscanthus, sorghum is able to produce the highest net energy balance (NEB) (Venturi and Venturi 2003; Agrawal et al. 2007; Hill 2009).

Sorghum was found in archaeological excavations estimated to be over 6,000 years old and was domesticated in arid areas of north-eastern Africa (Rooney et al. 2007). Sorghum varieties have been cultivated in various regions during recent years.
of the US for more than 150 years. The first agronomic sorghum was a sweet sorghum, Chinese Amber, introduced from China in 1853 (Smith and Fredericksen 2000). Isaac Hedges called it the “Northern Sugar Plant” considering the high sugar content in the stalks (University of Illinois Extension 2010). Following this initial introduction, a number of forage and sweet types came to the US from China, Africa, and Australia (Ali et al. 2008). Currently, the Sorg- hum bicolor genus consists of both cultivated and wild species, with Sorghum bicolor (L.) Moench being the most domesticated one in this genus (Ritter et al. 2007). Based on morpho-
gical characteristics of the inflorescence, the most important agronomic form, S. bicolor spp. bicolor has five races: bicolor, caudatum, durra, guinea, and kafir (Harlan 1972). Based upon end product utilization, sorghum can be divided into three categories: sweet sorghum for sugars, grain sor-
ghum for starch, and forage sorghum for biomass (Wang et al. 2009).

Several review papers have covered various issues related to biofuel production from sorghum. A review by Liang focused on sweet sorghum and explains the process of generating ethanol, lipid, and hydrogen from either sor-
ghum juice, whole stalk, or sorghum bagasse (Liang 2010). A review chapter written by Saballos introduced basically everything about sorghum, the orient, biology, growth condition and characteristics, sorghum genome, breeding, etc. (Saballos 2008). Other reviews put sorghum in a bigger context and compare it with maize and sugarcane in terms of genetics and features (Vermerris et al. 2007; Carpita and McCann 2008; Vermerris et al. 2010). While genetic improvement of sorghum is a broad topic, this review emphasizes three areas: 1) modifying lignin content. This is applicable to all sorghum varieties should sorghum be sought for biofuel production; 2) improving cold tolerance which is essential to grow sorghum in large scales and ensure abundant supply of this crop for year-around biofuel production; and 3) enhancing sugar content. This appears critical for sweet sorghum but the results can be applied to grain and forage sorghum as well to increase the content of nonstruc-
tural carbohydrates in the stalks. At present, grain sorghum ranks third in production in the US, behind maize and wheat. Forage sorghum was grown on 6 million acres in the US in 2007 with a total production of 58 million tons of biomass (Dien et al. 2009). Hence, in view of sorghum’s productivity, genetic improvement on the aforementioned three aspects will provide a huge impact on sorghum’s future.

MODIFYING LIGNIN CONTENT

All plant cells have a primary cell wall which is formed during the late stage of cell division and is composed of cel-
lulose, glucoronarabinoxylans (GAX), and phenolic resid-
ues (Dien et al. 2009; Vermerris et al. 2010). Secondary cell walls are limited to specialized parts such as support tissue for structure and vascular tissue for water transport (Vermerris et al. 2007). Secondary cell walls are located between primary cell walls and the membrane and consist of secondary and lignin. The secondary cell wall is categorized as a type II wall which distinguishes itself from type I as: 1) use of the GAX instead of xylo-
glucan in the matrix to embed cellulose microfibrils and 2) having low content of pectin and lignin (Carpita and Gibeaut 1993).

Lignin is a heterogeneous, complex, and hydrophobic biopolymer of p-hydroxyphenyl, guaiacyl, and syringyl residues (Di-GAX, 2009). Correspondingly, higher plant lignin has three primary subtypes based on the monomers involved. H-lignin arises from p-coumaryl alcohol, G-lignin from coniferyl alcohol, and S-lignin from sinapyl alcohol (Boerjan et al. 2003). Lignin is synthesized via the concerted action of the shikimic acid and phenylpropanoid pathways and polymerized via an oxidative coupling mecha-
nism (Ralph et al. 2004). A generalized scheme for lignin biosynthesis in sorghum is presented by Palmer et al. (2008). In plant cell walls, lignin fills in the spaces between cellulose and hemicellulose and provides a physical barrier against pathogen attack and chemical cleavage (Buendgen et al. 1990; Bonello et al. 2003; Chen and Dixon 2007). In the process of convering lignocellulosic materials to bio-
fuels, lignin exerts three effects: 1) behaving as a physical barrier; 2) hampering swelling of cellulose fibers, and 3) binding to cellulase enzymes nonspecifically (Vermerris et al. 2007). Thus, for the purpose of improving yield of sugars from cellulose, a low content of lignin in cellulosic biomass is desired.

The brown midrib (bmr) phenotype has been useful for identifying mutants with modified lignin synthesis in grasses. The mutants can be recognized easily by the reddish-brown coloration of vascular tissue in the leaf blade and sheath. Spontaneous bmr mutants were first observed in maize (Zea mays) (Jorgensen 1931; Sattler et al. 2009) and were subsequently generated in sorghum through chemical mutagenesis of two grain sorghum line (“954114” and ‘954104’) using diethyl sulfate at Purdue University in 1978 (Porter et al. 1978). Since then, additional bmr lines from the mutagenized population and additional spontaneous brown midrib mutants have been identified by Vogler et al. (1994). To date, 28 bmr mutants have been identified.

Before the reported study by Saballos, all of these mutants were named according to the order of discovery of the individual mutants and do not reflect their allelic rela-
tions (Saballos et al. 2008). Genetic and biochemical analyses, however, reveal that all known sorghum bmr lines fall within four distinct allelic groups, bmr2, bmr6, bmr12, and bmr19. In this new nomenclature system, the most widely recognized allele for each allelic group becomes the ref-
ference allele for the group while the current number is the individual allele identifier. The bmr2 group which includes bmr2, bmr5, and bmr14 is characterized by a decrease in G subunits with bmr14 demonstrating the largest reduction in compounds derived from G-residues. The bmr19 mutant has a chemical composition that is similar to those in the bmr2 group, a noticeable indication of reduction in compounds developed from G-subunits. However, the reductions in G-residues are considerably less obvious than those in bmr2 group. Though bmr2 and bmr19 genes may have similar or overlapping functions, the genes are not considered redundant since a mutant phenotype can be observed when either one of the genes is mutated.

The bmr12 group includes bmr7, bmr12, bmr15, bmr18, bmr25, and bmr26. The lignin of this group contains less S-
residues. But the content of p-coumaric acid varies among different mutants. The bmr6 group is the largest one among the four allelic groups. Lignin of this group of mutants has a reduced concentration of 4-vinylphenol and compounds representative of S-units, especially, 2,6-dimethoxyphenol. Among the four groups, significant reduction of Klasson lignin concentration is demonstrated except for bmr19 (Sabal-
os et al. 2008). Accordingly, enzymatic saccharification of stover representing bmr2, bmr6, and bmr12 lines demonstr-
ates up to 25% of increased glucose yield compared with that of wild-type stover. Among all bmr mutants, two mutants, bmr6 and bmr12, have been investigated the most. bmr6 encodes cinnamyl-
 alcohol dehydrogenase (CAD, EC 1.1.1.95) that catalyzes the reduction of cinnamyl aldehydes (coniferyl, coumaryl, and sinapyl aldehyde) to their corresponding cinnamyl alco-
hol dehydrogenase (CAD, EC 1.1.1.95) as a cofactor, prior to their incorpora-
tion into the lignin polymer (Sattler et al. 2010). Plants carrying this mutation typically display reduced level of lignin and incorporation of cinnamylaldehydes into the lignin polymer (Palmer et al. 2008). Between two genetic backgrounds, ‘RTx430’ and ‘Weatland’, bmr6 near-iso-
genic lines have the greatest amounts of different free phe-
nolics and other aromatics while the CAD activity is lowered. As a result, bmr6 plants exhibit reduced levels of G-, S-, and H-lignin in stem cell walls and increased incor-
poration of G- and H-indene into lignin relative to the wild type. In addition, reduced or loss of CAD activity also shifts the flux...
of soluble aromatic intermediates toward ferulic acid (FA). But the overall incorporation of esterified FA into the cell wall is not affected. Analysis of the available sorghum genome (730 Mb, www.phytozome.net/sorghum) reveals the existence of 14 CAD-like genes at seven genomic locations (Paterson et al. 2009). Of the 14 CAD-like genes identified with homology searches, at least 10 genes are expressed based on presence of corresponding expressed sequence tags (ESTs) and expression analysis (Saballos et al. 2009). Among all expressed genes, SbCAD2 is the predominantly expressed one. Mutations in this gene are responsible for the phenotype of the bmr6 mutants and have been identified. bmr6-ref is likely a null allele. In bmr6-ref, the occurrence of a C-to-T transition changes amino acid 132 of the protein from Gin (CAG) to a stop codon (UAG) (Sattler et al. 2009). The nucleotide base transition is consistent with the mutagen diethyl sulfate used to treat the population from which bmr6 was isolated (Bignami et al. 1988). This nonsense mutation truncates the reading frame prior to the NADPH binding and C-terminal catalytic domains. Consequently, the encoded protein is expected to be nonfunctional (Sattler et al. 2010). In bmr6-3, the mutation (G191S) is missense. This gene produces a full-length protein with a single disruption in the SAM binding site (Saballos et al. 2009). In bmr6-27, a frame shift results in the truncation of the last 27 amino acids and a disrupted secondary structure outside of the active site (Saballos et al. 2009). In bmr6 tissues, though CAD activity is reduced significantly (15-50% of that in wild-type), this enzyme activity is still detectable. Therefore, it is possible that other CAD proteins are present in sorghum. But bmr6 encodes the main CAD protein that is in charge of converting monolignols to monomonomers (Sattler et al. 2010).

bmr6 encodes caffeic acid-O-methyltransferase (COMT) that transfers a methyl group from S-adenosylmethionine (SAM) to 5-hydroxycinnamyl aldehyde to form sinapaldehyde (Li et al. 2000; Barrière et al. 2004). In bmr12 plants, loss of COMT activity leads to greater incorporation of G- and 5-hydroxyguaiaacyl lignin into cell wall but a strong decrease in S-lignin. The bmr12 allele is also observed to interact with genetic backgrounds. Increased levels of soluble aromatic compounds are indicated in bmr12 and double mutant (DM, bmr6 and bmr12) ‘Wheatland’ plants, but not in the comparable ‘RT*430’ near isogenic lines. In addition, concentrations of G- and H-lignin monomers are higher in ‘RT*430’ bmr12 plants than those in ‘Wheatland’ bmr12 lines. Thus, the impact of the bmr trait is different in different genetic backgrounds. The bmr6 gene contains six distinct alleles. For four of the alleles, bmr12, bmr18, bmr25, and bmr26, the mutations are nonsense since the premature stop codons truncate the polypeptide prior to the SAM binding site of the enzyme (Bout and Vermerris 2003; Saballos et al. 2008; Xin et al. 2008). The other two mutations in bmr7 and bmr13 are caused by missense mutations. A point mutation at position 2139 in bmr7 gene results in a proline to leucine substitution at position 152 (P152 L) of the COMT enzyme. Substitution of proline with leucine is likely to affect the conformation of the enzyme and its kinetic properties since the incorporation of proline is known to induce a twist in a polypeptide chain (Saballos et al. 2008). Alteration at position 152 does not affect the catalytic activity of the enzyme directly. Compared to null mutants of bmr12, bmr18, and bmr26, bmr7 enzyme is still functional though with lower efficiency.

In light of reduced lignin content in bmr6 mutants, three mutants, bmr6, bmr12, and bmr18 have been incorporated into commercial lines as bioenergy crops (Saballos et al. 2008; Corredor et al. 2009). Near isogenic bmr6, bmr12, and bmr6/bmr12 double mutant are compared with wild type in terms of ethanol yield following pretreatment and enzymatic hydrolysis. As reported by Dien et al. (2009), bmr6 and bmr12 have nearly identical effects on reducing lignin content by 13% and 15%, respectively, though each affects independent steps in monolignol synthesis and alters lignin composition in different ways. These two numbers are similar to 17 and 13% reported by Saballos et al. (2008) for bmr6 and bmr12, respectively. Interestingly, both effects are additive for the double mutant which has a lignin content reduction of 27%. Surprisingly, differences in genotypes do not influence carbohydrate contents due to low lignin content of 14% in the forage sorghum studied. Following pretreatment with dilute sulfuric acid (2%, w/v) at 121°C for 1 h and enzymatic hydrolysis of the washed and recovered solids, glucose recovery for the wild-type biomass and the double mutant is 68 and 88%, respectively. Glucose yields for sorghum biomass are improved by 27, 23, and 34% for bmr6, bmr12, and double mutant, respectively. Reduced lignin content does not affect xylose yield since xylan is completely hydrolyzed during pretreatment step. Ethanol yields from material pretreated by dilute acid were improved by 22, 21, and 43% for bmr6, bmr12, and the double mutant, respectively. The highest ethanol yield of 54.5% is obtained from the double mutant. Higher ethanol efficiencies are observed from sorghum pretreated by dilute ammonium hydroxide at 170°C for 20 min. In this case, the overall hydrolyses are fermented without washing. The overall ethanol yield is increased to 116.0 and 129.9 mg/g wet and dry biomass for the double mutant. In terms of the wild type, the ethanol yields are 98.4 and 113.2 mg/g wet and dry biomass, respectively. Overall, yields of glucose and ethanol are negatively related to lignin content with R = -0.976 and -0.943, respectively, for degrained sorghum biomass.

**IMPROVING COLD TOLERANCE**

For vast majority of the US sorghum belt, many varieties of sorghum are vulnerable to cold temperature from April to May. In particularly, stand establishment and early-season vigor of sorghum are adversely affected by air and soil temperatures below 15°C during germination, emergence, and early seedling growth (Yu and Tuinstra 2001). Robust cold tolerance of sorghum is attractive considering: 1) the crop can be sown earlier to take advantage of spring moisture and 2) it may facilitate two cropping cycles (Tiryaki and Andrew 2001a; Tiryaki and Andrews 2001b; Cisse and Ejeta 2003; Yu et al. 2004; Burrow et al. 2010). In order to incorporate cold tolerance into hybrids along with other desirable characteristics, research on cold tolerance has been focused on dissecting the Quantitative Trait Loci (QTL) for cold tolerance and developing effective and efficient means to transmit the trait to elite parental sorghum lines.

Using a mapping population of 153 recombinant inbred lines (RILs) developed from a cold-tolerant Chinese landrace, ‘Shan Qui Red’ (SQR) and a cold sensitive ‘SRN39’ of African origin, a genetic linkage map is constructed using 132 markers and Mapmaker/EXP v. 3.0b software (Knoll et al. 2008). These markers include 56 small sequence repeats (SSRs), 67 random amplified polymorphic DNA (RAPD), and 17 restriction fragment length polymorphisms (RFLPs). After eliminating unlinked markers, a total of 126 linked markers are used to perform simple interval mapping (SIM) and composite interval mapping (CIM) through QTL Cartographer.

The genetic distance of the map is 2,128 centiMorgans (cM) which is larger than 1,406 cM and 1,713 cM reported by Bhattramakki et al. (2000) and Menz et al. (2002), respectively. Six QTL are identified in this study and include: 1) a QTL identified by both CIM and SIM for germination at both cold and optimal temperatures on linkage group SBI-03a. This QTL is surprisingly connected with ‘SRN39’; 2) a QTL identified by CIM only for germination at cold temperature on group SBI-07a. This QTL is associated with SQR genotype; 3) two QTL associated with SOR genotype and picked by SIM only for seedling vigor on group SBI-01a for both early and late planting; 4) one QTL detected by CIM only for early seedling emergence on SBI-
02, which favors the ‘SRN39’ allele; and 5) one QTL selected by SIM only for seedling vigor in early, but not later planting.

Considering the small sample size used in the aforementioned study and some unresolved issues in terms of the precise locations of some of the QTL, the same authors conducted a follow-up study to validate the QTL associated with cold tolerance in different populations using a similar approach as in Ye et al. (2002). Markers: Xtxp43, Xtrp51, and Xtrp 211, each representing a QTL, were selected for this purpose (Knoll and Ejeta 2008). Xtrp51 and Xtrp 211 are chosen since they are only significant by one of the mapping models used and their favorable alleles are from different parents. Xtxp43 is selected due to its technical simplicity, codominant inheritance, and high level of repeatability. In addition, all these three markers have distinct polymorphisms regarding the SQR and the other two parents, ‘Tx2794’ and ‘Wheatland’.

These markers are tested in two new populations consisting 394 ‘Tx2794’ (cold sensitive) × SQR F3 and 390 ‘Wheatland’ (moderate early season vigor) × SQR BC1F1. The positive effect of SQR allele of Xtxp43 is validated in the two populations under early-season planting with regard to seedling vigor and emergence. However, the effect on emergence is slightly negative for the ‘Wheatland’ background. Marker Xtxp51 has the unexpected effect in the ‘Tx2794’ background in terms of superior seedling vigor, 15-day emergence, final emergence, and stand biomass. The effect on ‘Wheatland’ background, however, is not clear. Apparently, this QTL seems to be highly affected by temperature in some genetic backgrounds. Marker Xtrp211 demonstrates relatively small, but statistically significant effects for both genetic backgrounds regarding seedling vigor and final emergence. Overall, this study indicates that the effects of the QTL-associated markers are more pronounced in the SQR × ‘Tx2794’ F1 population than in the SQR × ‘Wheatland’ BC1F1 population. Xtxp43 and Xtrp51 are useful for introgressing early-season cold tolerance from SQR into different population, at least to ‘Tx2794’. These results are in agreement with those reported by Cisse and Ejeta (2003) that SQR is an excellent genetic source for seedling vigor and stand establishment and the vigorous seedling growth in this line is heritable.

The most recent publication on cold tolerance is reported by Burrow et al. (2010). In this study, ‘RT-430’, a widely adapted inbred and an important pollinator/restorer line of sorghum and ‘PI610727’, a Chinese kaoliang, are selected as the parents. ‘PI610727’ is a semi-adapted photoperiod-insensitive landrace and is selected for early-season cold tolerance and is relatively resistant to many sorghum-implicated diseases. The 171 F1 RILs are planted at two locations in Texas, US and are screened for four traits: germinability at cold and optimal temperatures, field emergence, and seedling vigor. Genotyping is achieved through analysis of SSR markers by PCR. The linkage map constructed by software Mapmaker/EXP consists of 141 SSR markers and spans a total of 1,005 cM. QTL are identified using QTLCartographer v. 2.5. A total of 14 QTL associated with the four traits tested are uncovered in this study. Two and three QTL are detected for cold and optimal temperature germinability, respectively. These QTL are located on chromosome 2/9 and 1/1/2 for these two temperatures, respectively. Interestingly, two QTL from this group are found to co-localize under the same genome region and have the same flanking marker. QTL associated with field emergence are identified as four for each of the two locations. Regarding USDA_LBK site, the QTL are on chromosome 1, 7, and two on 9. In terms of ND_TAMU location, one QTL is identified on chromosome 1 and 4 and two on chromosome 9. For seedling vigor, only one QTL is located on chromosome 4. Thus, the QTL are located in five chromosomes, 1, 2, 4, 7, and 9. Chromosome 9 is special as it harbors four QTL for field emergence. These QTL have also been related to resistance to grain mould (Klein 2007).

The genetic map attained from this study is consistent with the sorghum genome sequence. The map distance (1,005 cM), though comparable to other studies (Huang and Wu 2007; Salas Fernandez et al. 2008), is lower than the map presented by Menz et al. (2002). This could be due to limited coverage of the genomic regions for several chromosomes, especially chromosome 5 and 8. Fewer polymorphisms for the SSR markers are found in these two chromosomes. The parental 630 could be used for breeding early season cold tolerant sorghum considering the high heritability values (h² = 0.62-0.83). For that purpose, it is strongly suggested that field emergence at early-season planting must be the main screening method to ensure efficient selection though germinability could serve as one of the indicating factors for cold tolerance.

ENHANCING SUGAR CONTENT

Sweet sorghum, also called sorgos, typically has low grain yields, but high sugar content in the stalk. Generally speaking, the stem juice contains 10-25% of sugars comprising sucrose, glucose, and fructose (Liang 2010). Sweet sorghum has been researched for biofuel production in the US for over 30 years (Lipinsky 1980). Recently, in light of the success with sugarcane in Brazil, investigation on sweet sorghum has regained momentum. Bioethanol production from sweet sorghum is straightforward, a process similar to that of producing ethanol from sugarcane. In addition, these sugars can be transformed to lipids through microalgal fermentation (Gao et al. 2010; Liang et al. 2010). Compared with grain sorghum whose production mainly relies on F1 hybrids, breeding effort for sweet sorghum has been much less (Wang et al. 2009).

While genotypic differences for sucrose enrichment have been described in sugarcane (Lingle and Wiegand 1997), the mechanism of sugar accumulation in sorghum is basically unknown. Apart from that, many enzymes associated with sugar accumulation in sugarcane, such as sucrose phosphate synthase and invertase, do not seem to play significant roles for sugar accumulation in sorghum (Lingle 1987; Tarpley et al. 1994). What is known is that sugar accumulation and the final sugar content in the parenchyma of the juicy stem are controlled by multiple genes and environmental conditions. Understanding the mechanisms controlling sugar traits is not only important for sweet sorghum, it is also crucial for grain and forage sorghum for the purpose of increasing content of nonstructural carbohydrates. Two research threads are ongoing with the goal of understanding sugar accumulation in sorghum. One is to study the differential gene expression of the sorghum genome. Another is to identify and map QTL for sugar-related traits.

Sorghum and sugarcane are both panicoid C4 cereal plant and diverged from each other around 8-9 million years ago (Guimarães et al. 1997; Jannoo et al. 2007). Using an Affymetrix Genechip originally created to study gene expression in sugarcane (Casu et al. 2007), Calvino et al. (2008) identified genes involved in sugar accumulation and lignocellulose synthesis by comparing genes in grain and sweet sorghum. The sugarcane array comprises 8,224 probe sets and is specifically developed with sequences obtained from several cDNA libraries representing distinct tissue types including stem from 15 sugarcane varieties. With sorghum RNA samples, more than 70% (5,900) of probes give positive signals. Between sweet sorghum line ‘Rio’ and grain sorghum ‘BT×623’, a total of 132 transcripts are down-regulated and 63 are up-regulated in ‘Rio’ when the cut-off value is two-fold. Considering the fact that some
probe sets identify the same gene, the number of genes that is down-regulated and up-regulated is 103 and 51, respectively. Obviously, down-regulated genes are twice of the up-regulated ones.

Among the transcripts that are up-regulated in ‘Rio’, a saposin-like type B gene displays the highest differential expression. Saposins are water soluble proteins that interact with the lysosomal membrane and are involved in the catalytic subunit 12 involved in cellulose synthesis. In addition, a series of genes encoding proteins in lignin synthesis are also down-regulated, for example, cinnamoyl CoA reductase, cinnamyl alcohol dehydrogenase, 4-coumarate coenzyme A ligase, caffeoyl-CoA O-methyltransferase, xyloglucan endotransglycosylase/ hydrolase, peroxidase, phenylalanine, and histidine ammonia-lyase.

The differentially expressed genes are also mapped to the sorghum genome using GenomeThreadder software. Out of 195 probes sets, 176 can be mapped to the genome based on their alignment with a sorghum gene. Six probe sets are mapped to the genome but do not agree with the current sorghum gene annotation. The left 13 probe sets do not fit into the sorghum genome. The mapping analysis shows that genes that are differentially expressed between sweet and grain sorghum are randomly distributed and do not cluster at any particular locations.

Following the above mentioned study on hybridization of RNAs derived from the stems of grain and sweet sorghum onto the sugarcane Affymetrix gene chip (Calviño et al. 2008), the same group of researchers used the sugarcane gene chip analysis to extend the cross-species concept in discovering single-feature polymorphisms (SFPs) (Calviño et al. 2009). Employing the recently developed software GeneSNP, SFPs are identified in 58 out of 154 differentials between sweet and grain sorghum. Chromosomes of 1, 2, and 3 have the highest number of genes displaying both expression level polymorphisms (ELPs) and SFPs while chromosomes 5 and 6 have the lowest number of ELPs and SFPs. Interestingly, chromosomes 8 and 9 are the most polymorphic ones according to measurement of number of SNPs per Kbp of sequence. Based on 87 single nucleotide polymorphisms (SNPs) in 21,612 bps of sequence obtained from both parental lines, it is calculated that an average of one SNP exists for every 248 bps of sequence between ‘Rio’ and ‘BT’+‘623’. Considering the size of sorghum genome of 730 M bps (Paterson et al. 2009), it is suggested that 2,938,800 SNPs may be present between ‘Rio’ and ‘BT’+‘623’. Between the two lines, at least 0.4% of the genome is polymorphic.

Fragments harboring SFPs with t values equal or larger than 7 from the 58 genes are cloned and sequenced. 30 of them have SFPs. Out of these 30, 19 genes (33%) have a validated SFP. Based on these validated SFPs, from 58 candidate genes, 18 allele-specific PCR markers are further developed using the single nucleotide amplified polymorphism (SNAP) technique (Drenkard et al. 2000). These markers will be invaluable for later mapping purposes. One marker within the gene Shbgg029170 coding for a putative ketol-acid reductoisomerase can discriminate grain sorghum from sweet sorghum line. This result is in line with others in using length polymorphism (AFLP) markers on chromosome 8 which unambiguously distinguish grain from sweet sorghum lines (Ritter et al. 2007).

Other researches focusing on improving traits related to sugar accumulation have taken a similar approach of genetic mapping, but have concentrated on characterizing the related QTL. QTL investigated include those associated with sugar components (Brix, glucose, sucrose, and total sugar content) and sugar-related agronomic traits (flowering date, plant height, stem diameter, tiller number per plant, fresh panicle weight and estimated juice weight) (Shiringani et al. 2010). Several studies have used the cross between sweet and grain sorghum to generate RILs. These RILs and series of markers are then used to construct genetic linkage maps adopting different software.

As shown in Table 1, the effort for identifying QTL for certain sorghum traits is reported as early as 1995 (Lin et al. 1995; Pereira and Lee 1995) and continued throughout the years (Rami et al. 1998; Crasta et al. 1999; Hart et al. 2001; Klein et al. 2001; Natoli et al. 2002; Feltus et al. 2006; Murray et al. 2008, 2009; Srinivas et al. 2009). Extensive studies have been conducted in recent years reflecting the urgent need for sugar feedstocks for biofuel production. Ritter et al. (2008) conducted an analysis of QTL for stem sugar-related and other agronomic traits using a population derived from sweet sorghum (‘R9188’) and grain sorghum (‘R9403463-2-1’). ‘R9188’ is a dwarf conversion of sweet sorghum line ‘Rio’ while ‘R9403463-2-1’ is a ‘DPI&F’ elite grain sorghum male parent. 184 F6 RIL populations are phenotyped for sugar-related traits. Using 228 SSR and AFLP markers, a genetic map is constructed. QTL are identified for all traits investigated and are found to co-locate on five chromosomes. For example, the QTL for sucrose content, sugar content, Brix, sucrose yield, fructose content, plant height, and flowering time are all located on a similar region on chromosome 6. This co-localization may be explained by two possible reasons: 1) genes in these genomic locations affect a number of traits (pleiotropy) and 2) these genomic regions may contain several genes, each of which affect a different trait and the co-localization of QTL for different traits is simply a result of linkage. The co-locating QTL for sugar related traits are excellent choices for marker-assisted selection in sweet sorghum.

Stem sugar is one of the most important parameters in sweet sorghum breeding. In this study, it was found that sugar content is strongly correlated with plant height and weakly correlated with delayed flowering. This observation hints that taller plants with delayed flowering produce more stem biomass and have more time to accumulate photosynthate. Ferraris and Charles-Edwards (1986) in view of the fact that the progeny phenotypes exceed the parental range for total dry matter, grain yield, glucose and fructose content, this study suggests that it is possible to enhance sugar related traits by combining QTL from ‘R9188’ and ‘R9403463-2-1’. Sweet sorghum ‘R9188’ contributes to increased sucrose content and sugar content by providing QTL alleles on chromosome 1, 5, and 6. QTL alleles for increased Brix and sugar content are also found on the ‘R9188’ and are identified on chromosome 5/6 and 3, respectively. QTL alleles from ‘R9403463-2-1’ is detected for enhanced sucrose content and sucrose yield on chromosome 10 and increased glucose content on chromosome 7. A study by Shiringani et al. (2010) is by far the most comprehensive one on detecting QTL for sugar related traits. In this investigation, QTL for 10 traits of sorghum planted at two field locations and two different times are identified using the F₃_s generation of 188 RILs from a cross between
Table 1: Chromosome location of QTL for different traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number of QTL</th>
<th>QTL chromosome location</th>
<th>% of phenotypic variation explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering dates</td>
<td>5</td>
<td>3, 4a, 6, 7, 8</td>
<td>33.4</td>
</tr>
<tr>
<td>plant height</td>
<td>17 (9)</td>
<td>1, 2, 4a, 4b, 6, 7, 8</td>
<td>7.5-26.1</td>
</tr>
<tr>
<td>Stem diameter</td>
<td>13 (9)</td>
<td>1, 3, 6, 7, 8</td>
<td>13.3</td>
</tr>
<tr>
<td>number of tillers per plant</td>
<td>14 (6)</td>
<td>1, 2, 3, 4a, 4b, 5, 6, 7, 9, 10</td>
<td>6.7-15.7</td>
</tr>
<tr>
<td>Fresh panicle weight</td>
<td>5</td>
<td>1, 2, 4a, 6</td>
<td>20.3</td>
</tr>
<tr>
<td>Stem juice weight</td>
<td>1</td>
<td>2</td>
<td>7.9</td>
</tr>
<tr>
<td>Brix</td>
<td>14 (4)</td>
<td>2, 4b, 7</td>
<td>21.9-22.3</td>
</tr>
<tr>
<td>Glucose content</td>
<td>11 (4)</td>
<td>1, 3, 7</td>
<td>13.7</td>
</tr>
<tr>
<td>Sucrose content</td>
<td>9 (4)</td>
<td>1, 2, 6, 10</td>
<td>42</td>
</tr>
<tr>
<td>Sugar content</td>
<td>17 (9)</td>
<td>1, 2, 6, 7, 8, 9</td>
<td>6.6-12.5</td>
</tr>
</tbody>
</table>

Reference: Shiringnani et al. 2010

<table>
<thead>
<tr>
<th>QTL Chromosome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering dates</td>
</tr>
<tr>
<td>plant height</td>
</tr>
<tr>
<td>Stem diameter</td>
</tr>
<tr>
<td>number of tillers per plant</td>
</tr>
<tr>
<td>Fresh panicle weight</td>
</tr>
<tr>
<td>Stem juice weight</td>
</tr>
<tr>
<td>Brix</td>
</tr>
<tr>
<td>Glucose content</td>
</tr>
<tr>
<td>Sucrose content</td>
</tr>
<tr>
<td>Sugar content</td>
</tr>
<tr>
<td>Dry matter yield</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>QTL Chromosome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering dates</td>
</tr>
<tr>
<td>plant height</td>
</tr>
<tr>
<td>Stem diameter</td>
</tr>
<tr>
<td>number of tillers per plant</td>
</tr>
<tr>
<td>Fresh panicle weight</td>
</tr>
<tr>
<td>Stem juice weight</td>
</tr>
<tr>
<td>Brix</td>
</tr>
<tr>
<td>Glucose content</td>
</tr>
<tr>
<td>Sucrose content</td>
</tr>
<tr>
<td>Sugar content</td>
</tr>
<tr>
<td>Dry matter yield</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>QTL Chromosome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering dates</td>
</tr>
<tr>
<td>plant height</td>
</tr>
<tr>
<td>Stem diameter</td>
</tr>
<tr>
<td>number of tillers per plant</td>
</tr>
<tr>
<td>Fresh panicle weight</td>
</tr>
<tr>
<td>Stem juice weight</td>
</tr>
<tr>
<td>Brix</td>
</tr>
<tr>
<td>Glucose content</td>
</tr>
<tr>
<td>Sucrose content</td>
</tr>
<tr>
<td>Sugar content</td>
</tr>
<tr>
<td>Dry matter yield</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>QTL Chromosome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering dates</td>
</tr>
<tr>
<td>plant height</td>
</tr>
<tr>
<td>Stem diameter</td>
</tr>
<tr>
<td>number of tillers per plant</td>
</tr>
<tr>
<td>Fresh panicle weight</td>
</tr>
<tr>
<td>Stem juice weight</td>
</tr>
<tr>
<td>Brix</td>
</tr>
<tr>
<td>Glucose content</td>
</tr>
<tr>
<td>Sucrose content</td>
</tr>
<tr>
<td>Sugar content</td>
</tr>
<tr>
<td>Dry matter yield</td>
</tr>
</tbody>
</table>


Legend:
a: Number in the parenthesis indicates number of significant QTL.
b: Bold numbers are the number of chromosome where major significant QTL are located.
c: Italicized number means percentage from one QTL, not from all of the QTL.

a female parent, sweet sorghum (‘SS79’) and a male parent, grain sorghum (‘M71’). A series of 102 AFLP, 49 SSR, and 6 EST-SSR markers are employed to construct the genetic map. As indicated in Table 1, QTL for almost all traits are located on several chromosomes. But one or two QTL for some traits has more significant contribution for phenotypic variation than those of others.

Using combined analysis, this study detects 24 significant digenic epistatic effects for 7 traits. Two locus interactions are found to be common in the entire genome. Several digenic QTL are identified: five regarding fresh panicle weight, four for plant height and glucose content, three for number of tillers, sucrose, and sugar content, and two for flowering date. Additionally, interactions between QTL and environment are analyzed. Significant QTL × environment interactions are observed for sugar content on chromosome 1, fresh panicle weight on chromosome 1 and 6, tiller number per plant on chromosome 1 and 7, and flowering date on chromosome 6. Furthermore, some QTL are observed to influence more than one trait. One major QTL with multiple effects is discovered in linkage group 6 with a support interval of 12-18 cM. This QTL has been detected for all traits except for glucose and number of tillers per plant and was significant for five out of six traits. Moreover, this study is able to determine hotspots which are QTL clusters that have at least one QTL explaining more than 5% of the phenotypic variance with both paternal and maternal effects. Hotspots are claimed on chromosome 1, 2, 6, 7, and 9 since all traits co-localized on those linkage groups and more than one QTL has been found per trait.

While most of the researches on improving sugar content aim at sweet sorghum, one group has attempted to study whether there is a tradeoff between sugar content and starch content in grain sorghum in consideration of the fact that both sugars and starch are valuable feedstocks for biofuel production (Murray et al. 2008). The purpose of this investigation is to help maximize the total usable energy that can be produced and stored in sorghum. Using a mapping population of 176 F₂:5 RILs derived from a cross between sweet sorghum ‘Río’ and grain sorghum ‘BT×623’, QTL controlling yield and composition of sugars in the stalk as well as yield and composition of starch, fat, protein, fiber, and phosphorous in grain have been identified. The genetic map contains a total of 259 SSR and AFLP markers. The total genetic distance represented on the map is 1,836 cM. Significant discoveries from this study are: 1) the majority of favorable alleles for QTL are derived from the anticipated parent; 2) for most traits, at least one positive QTL across locations is contributed by the unexpected parent; 3) for many traits, QTL co-localize either within or between
locations, in particular for the major height and/or flowering time genes; 4) QTL co-localization clusters observed on chromosome 4, 6, 7, and 9 correspond to height, flowering time, or stand density-tilling QTL; and 5) QTL for Brix and stem sugar concentration maps to near identical positions on chromosome 3. The researchers conclude that it is feasible to improve grain starch and stem sugar simultaneously in both grain and sweet sorghum types under ideal agronomic conditions. This is supported by other studies which show that grain yield is not strongly negatively correlated with the stem sugars. In another word, stem sugar storage does not strongly affect final grain yield (Ritter et al. 2008). However, the tradeoffs will increase with environmental stress. Considering the accuracy of QTL detection which can be affected by population size, trait heritability, and recombination, more sweet and grain sorghum lines need to be evaluated to confirm this conclusion.

FUTURE RESEARCH RECOMMENDATION

Lignin content modification

With so many mutants available for people who work in bioenergy and plant breeding fields, one question is obvious: which mutant should be used for further study to reduce lignin content and enhance sugar yield from sorghum biomass. While this seems an easy question, there is no short answer in view of different recommendations made by different research groups. However, it should be stressed that establishing a superior hybrid with desired properties is possible. One needs to keep in mind, however, that the interaction between bmr gene and the genetic background varies (Palmer et al. 2008). Thus, it is strongly suggested that end-users of brown midrib sorghum should always evaluate the performance of specific hybrid/gene combinations (Sattler et al. 2010). In addition, more research needs to be conducted to elucidate the effects of reduced lignin content on sorghum yield, lodging susceptibility, and disease susceptibility as reviewed by Sattler et al. (2010).

Improving cold tolerance and enhancing sugar content

Over the years, different QTL for different traits have been reported. But putting all results together as listed in Table 1, no significant agreement can be found for any traits. It appears that QTL for different traits are distributed randomly in different sorghum populations. The possible explanation for this wide distribution may be: 1) different sorghum crop has been used for study. At the time of writing, no two studies have adopted the same sorghum line; 2) for most of these studies, several hundreds of markers are tested. But compared to the genome size of sorghum, a saturation coverage of the whole genome will require more than 100 times more markers, which is practically impossible. The sorghum breeding community really needs to come up with a consensus on what elite sorghum lines should be focused on testing. By doing so, effort can be concentrated and rapid progress can be made to understand the genes and proteins involved in cold tolerance and sugar accumulation in sorghum. With this knowledge, future work can be initiated for establishing the best sorghum crop with beneficial traits.

Future bioenergy crop should have characteristics like higher yield, greater NEB, lower recalcitrance, and improved ecological benefits (carbon fixation, water and soil conservation) (Yuan et al. 2008). For sorghum, it would be a revolutionary accomplishment if certain sorghum lines can be developed incorporating all desirable properties: reduced lignin content, improved cold tolerance, and enhanced sugar content. Combining all needed genes into one elite sorghum population will be a tremendous task. But seeing how bio-technologies progress in recent years, it is envisioned that such a crop will be developed in the near future if enough resources are dedicated to this effort.

ACKNOWLEDGEMENTS

Financial support from Burroughs Welcome Fund (BWF) Foundation is greatly appreciated. Most of this manuscript was written while the author was participating in a workshop sponsored by BWF Foundation.

REFERENCES


Calviño M, Bruggmann R, Messing J (2008) Screen of genes linked to high-sugar content in stems by comparative genomics. Rice 1, 166-176


Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal 3, 1-30


between inter- and intra-specific Sorghum populations. Theoretical and Applied Genetics 112, 1295-1305


Hille R (2006) Use of sweet sorghum juice for costs and benefits of transportation biofuel production from food-and lignocellulose-based energy crops: A review. Sustainable Agriculture Part I, 125-139


Janzen VC, Biotechnology 291, 1061-3623


Jorgenson LR (1931) Brown midrib in maize and its linkage relations. Agronomy Journal 23, 540-557


Yousef GGJ, John A (2002) Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. *Crop Science* 42, 96-104


