DESIGNING AN IDEAL ENERGY CROP:

THE CASE FOR *SORGHUM BICOLOR*

A Dissertation

by

SARA NICOLE OLSON

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

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Biochemistry

Designing an Ideal Energy Crop: The Case for *Sorghum bicolor*

DESIGNING AN IDEAL ENERGY CROP:

THE CASE FOR *SORGHUM BICOLOR*

A Dissertation

by

SARA NICOLE OLSON

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	John Mullet
Committee Members,	Tatyana Igumenova
	Gary Kunkel
	Terry Thomas
Head of Department,	Greg Reinhart

August 2012

Major Subject: Biochemistry

ABSTRACT

Designing an Ideal Energy Crop: The Case for *Sorghum bicolor*.

(August 2012)

Sara Nicole Olson, B.S., Texas A&M University

Chair of Advisory Committee: Dr. John Mullet

Following the passage of the United States Energy Independence and Security Act in 2007, significant progress has been made in replacing liquid fossil fuels with biofuels from lignocellulosic biomass. Many crops have been examined as potential energy crops and any ideal energy crop will meet at least three requirements. First, an ideal energy crop must generate a large amount of biomass. Second, a crop must be able to accumulate biomass from minimal inputs like water and nitrogen fertilizer. Third, an ideal energy crop will generate biomass with a composition that is ideal for refinement into biofuels.

Bioenergy hybrid genotypes of *Sorghum bicolor* represent an ideal energy crop. These plants generate very large amounts of biomass over the course of an extremely long duration of vegetative growth. This accumulation of biomass is achieved without the requirement of additional fertilizer beyond a standard application level, and the biomass of *S. bicolor* has a composition that is ideal for generation of biofuels.

This dissertation demonstrates the genetic yield potential of bioenergy hybrid *S. bicolor*. In addition, it is shown that *S. bicolor* is able to grow with very high nitrogen use efficiency throughout its long duration of vegetative growth. Many genetic loci are identified which modulate plant size traits in *S. bicolor*, including stem length, leaf area, and total biomass yield. Finally, the genetic position of *Ma2*, an important maturity locus, is identified. These results together make the case the *S. bicolor* is an ideal energy crop.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mullet, and my committee members, Dr. Igumenova, Dr. Kunkel, and Dr. Thomas, for their guidance and support throughout the course of my graduate career. Thanks are also due to Dr. Rooney and the members of his lab and field crew, without whom all of my field trials and harvests would have been impossible. In addition, I owe great thanks to Susan Hall and the undergraduate student workers who have tirelessly assisted me in the harvests that have been such a large part of my dissertation.

My thanks go to Dr. Wild, whose encouragement gave me the push I needed to apply to the doctorate program in the first place. I would also like to thank the many faculty members of the Biochemistry department who have provided great support over the years, as well as the staff of the Biochemistry department for maintaining a friendly, helpful atmosphere in our building. I would like to extend special thanks to Daisy Wilbert, whose help behind the scenes has been truly invaluable to me. I would also like to thank the Biochemistry Graduate Association for the privilege of serving as your Vice President. To the many students I have taught over the years, both in the classroom and at the bench, thank you for the opportunity to share my knowledge with you and for being patient with me as I learned to be a better teacher.

On a personal note, I would like to thank the many people who have been there for me throughout this graduate school process. To my wonderful parents and sister, thank you for patiently tolerating my “Rudolph” moments over the years and for helping me keep things in perspective through the ups and downs of this experience. To my brilliant, loving fiancé, thank you for tirelessly supporting me as we have both separately struggled to complete our research and our dissertations; with your help I become a better scientist every day. And finally, to the remarkable ladies and gentlemen of the University Writing Center Dissertation Writing Group, thank you for debating the nuances of my writing with me as we all worked together to make me a stronger writer.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

LIST OF TABLES

Page

Page

CHAPTER I

INTRODUCTION

Fossil fuel past, biofuel present

As consumption of fossil fuels increases across the globe, so does the demand for alternative sources of fuel. Enacted in 2007, the United States Energy Independence and Security Act (EISA) required that United States production of biofuels must increase to 36 billion gallons annually by 2022 (Rahall, 2007). In 2007, the production was approximately 5 billion gallons annually and by 2010 production reached 12 billion gallons (United States Energy Information Agency, 2009), nearly all from corn ethanol. To achieve the requirements set forth in EISA, biofuel production will need to increase by nearly 200 percent in the coming years (Table 1).

Generation of biofuel is accomplished by processing biomass into fuel. Biomass is a general term used to describe any number of plant products including grain, lignocellulosic plant matter such as culms and leaves, woody materials, and agricultural residues (Rooney, et al., 2007). Lignocellulosic biomass is so-named because of its high concentration of lignin, cellulose, and other structural, long-chain carbohydrates. These and other similar molecules can yield significant quantities of liquid fuel when processed (Bouton, 2007). There are a variety of processes that can be used to generate liquid fuel from any of these biomass sources, each with its own benefits and drawbacks

_____

This dissertation follows the style of Plant Physiology.

(Rooney, et al., 2007). In order to bridge the gap between current production and required future production of biofuels, advances in fuel generation strategies and advances in the ability to generate high quality biomass in large quantities will both be of critical importance (National Research Council (U.S.). Committee on Economic and Environmental Impacts of Increasing Biofuels Production., et al., 2011).

Table 1: Global liquid fuel demand for 2007 and projected global liquid fuel demand for 2030. Projections for 2030 are estimates based on requirements set forth in The Energy Independence and Security Act of the United States.

	Fuel Type		
	Conventional	Biofuels, others	Total
2007 (mbpd)	81.8	4.4	86.2
2030 (mbpd)	93	12.4	104.5
% growth	14%	181%	22%

Initial efforts aimed at generating biofuels have focused on extraction of energy from the starch molecules contained in grain biomass, specifically corn grain (Byrt, et al., 2011). This process is not ideal for multiple reasons. First, if the entire crop of corn grain produced in the United States in one year was converted to fuel, it would generate less than one fourth of the total fuel consumed annually (US Environmental Protection Agency, 2010). Secondly, corn grain is a widely used feed for livestock as well as for human consumption (International Crops Research Institute for the Semi-Arid Tropics, 2011). As such, use of this form of biomass to generate fuel has created competition

between food production and fuel production, raising food prices around the globe (Raneses, et al., 1999). This unintended result demonstrates that corn grain and other grain sources are not ideal energy crops.

The issues that have emerged as a result of using corn for fuel have brought to light the fact that dedicated energy crops will be necessary to generate the quantities of biofuel required by the EISA (Perlack, et al., 2005). In order to be an ideal energy crop, there should be minimal competition between food and fuel uses for biomass generated by that crop. This means that lignocellulosic biomass sources offer an attractive alternative to grain sources as vegetative plant material does not tend to be a direct part of human food supply systems (Rooney, et al., 2007). While no single crop may be able to generate all of the biomass needed to fulfill biofuel demand, development of dedicated, ideal energy crops can ease the pressure on grain supplies and simultaneously increase potential biofuel yield (Byrt, et al., 2011).

Multiple previous studies have demonstrated the phenomenal biomass generation abilities of various plant species that may be considered as ideal energy crops (Byrt, et al., 2011; Bennett and Anex, 2009; Carpita and McCann, 2008; Dohleman and Long, 2009; Dohleman, et al., 2009; Lewandowski and Schmidt, 2006; Jorgensen, 2011; Sanderson, et al., 1996b; Vermerris, 2011). Most of these potential energy crops are C4 grass species, including *Panicum virgatum* (switchgrass), *Miscanthus giganteus* (miscanthus), and *Sorghum bicolor* (sorghum) (Byrt, et al., 2011). While biomass yields have been reported

for various genotypes of each of these crops, more research is needed to determine the attributes of each species that could contribute to its status as an ideal energy crop. This dissertation examines *S. bicolor* as a potential energy crop on the basis of the many attributes that contribute to a crop's appropriateness for use in biomass generation. This research makes that case that *S. bicolor* has many attributes of an ideal energy crop.

***Sorghum bicolor* background**

Sorghum bicolor is the fifth most cultivated cereal crop in the world (International Crops Research Institute for the Semi-Arid Tropics, 2011). The global importance of *S. bicolor* can be attributed to its multiple, varied uses, including as a source of grain, a forage crop for livestock feed, and more recently, as a source of cellulosic biomass for biofuel generation (Smith and Frederiksen, 2000). Originating in sub-Saharan East Africa, *S. bicolor* is widely distributed throughout eastern and southern Africa, southern Asia, and Australia and is also cultivated in North and South America (Dewet and Huckabay, 1967). Owing in part to its wide distribution, *S. bicolor* exhibits significant within-species variation for many traits. Historically, *S. bicolor* was divided into 52 species (Snowdon, 1936). However, this division was later revised due to the lack of reproductive barriers between Snowdon's species, and all 52 groups are now included in the *S. bicolor* classification and are grouped into five races based on phenotypic variation: bicolor, kafir, caudatum, durra, and guinea (Carena, 2009; Murty and Govil, 1967).

The genome of *S. bicolor* is simple and relatively small in comparison to many other grasses. The diploid genome contains ten chromosomes (2n=20) totaling approximately 730 million base pairs (Mbp) (Paterson, et al., 2009). There are approximately 34,500 protein-coding loci annotated in the genome, encoding just under 37,000 transcripts (Goodstein, et al., 2012). The largest of the chromosomes, chromosome 1, is 119 Mbp long. The chromosomes are numbered in order of decreasing length, through chromosome 10, which is 69 Mbp long (Kim, et al., 2005).

S. bicolor is a member of the family Poaceae, also called Gramineae, which includes true grasses (Smith and Frederiksen, 2000). Zea mays, which diverged from *S. bicolor* approximately 11.9 million years ago (mya) (Swigonova, et al., 2004), is included in this family, as well as Oryza sativa, which diverged approximately 50 mya (Wolfe, et al., 1989). As with all members of the Poaceae, *S. bicolor* is a monocotyledonous plant (Smith and Frederiksen, 2000). Sorghum genotypes with annual lifecycles and perennial lifecycles are known. *S. bicolor* is readily rattooned and will re-grow nearly indefinitely if kept in conditions that are favorable for growth (Vanderlip, 1993).

The specific photosynthetic biochemistry employed by *S. bicolor*, C4 NADP-ME photosynthesis, is responsible for a significant portion of its extremely high water use efficiency (WUE) (Smith and Frederiksen, 2000; Allen, et al., 1998; Heaton, et al., 2004; Mastrorilli, et al., 1995; Schmitt and Edwards, 1981; Zhu, et al., 2010). This mechanism divides the process of carbon dioxide uptake

and fixation between mesophyll and bundle sheath cells, respectively, to maximize carbon fixation while minimizing water loss through stomata (Fig. 1)

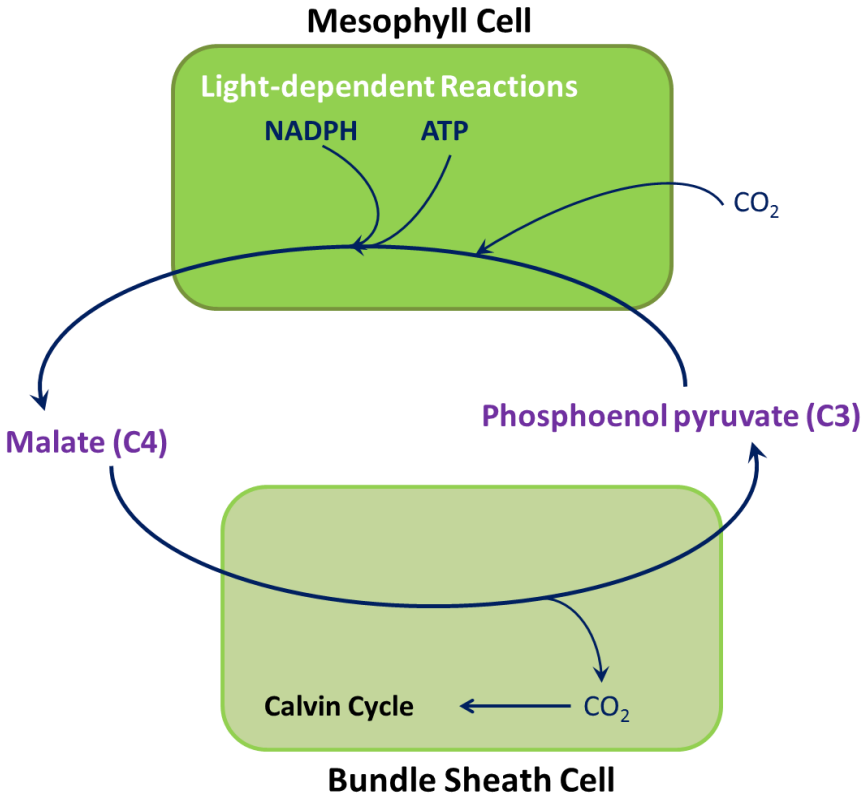

Figure 1: Simplified schematic depicting C4 photosynthesis paradigm. Carbon dioxide is taken into the mesophyll cells through stomata where it is fixed into malate (a C4 acid). Malate is then transported into bundle sheath cells where the newly-fixed carbon dioxide is released to create an artificially high concentration of carbon dioxide in the bundle sheath cell to facilitate carbon fixation by RuBisCO.

and uses the NADP-dependent isoform of malic enzyme to decarboxylate malate during carbon assimilation and fixation (Sage and Monson, 1999).

Growth of Sorghum bicolor

The growth of *S. bicolor* can be broken into three main stages, numbered one through three (Vanderlip, 1993; Gerik, et al., 2003). Growth stage one (GSI) refers to the vegetative growth stage of the plant, starting with emergence of the seedling and including growth of leaves and culm. GSI encompasses all growth from the emergence of the coleoptile leaf through the transition of the shoot apical meristem from vegetative to flowering (Fig. 2a). Leaves develop within the whorl and once mature, are positioned on alternating sides of the culm. Each leaf blade is connected to the culm by its own leaf sheath, and each leaf sheath joins the culm at a node. The portion of culm between two nodes is called an internode. Thus, the culm is made up of an alternating series of nodes and internodes. This first stage is the most variable in length in *S. bicolor*, and the duration is controlled by the actions of many genes. Photoperiod sensitive bioenergy hybrid genotypes of *S. bicolor* that have been developed grow in GSI for extremely long durations when grown in conditions where the length of exposure to light each day exceeds 12 hours and 20 minutes (long days) (Rooney and Aydin, 1999).

GSII begins when flowering is initiated in the shoot apical meristem and carries through anthesis of the panicle, including booting and exsertion (Vanderlip, 1993; Gerik, et al., 2003). There is minimal variation in the duration of this growth stage when compared with GSI. During GSII, it is possible to anticipate emergence of the panicle by looking for the flag leaf. This is the final

leaf produced by the plant, and its leaf sheath surrounds and protects the developing panicle until it emerges. As the immature panicle develops within the leaf sheath of the flag leaf, the peduncle also increases in length. This increase in peduncle length is called booting, and this booting is the main source of increases in culm height observed in *S. bicolor* plants once GSI is complete and flowering has been initiated. The last several internodes of the plant will expand during this time as well. GSII ends when the panicle, pushed up by the growing peduncle, emerges from the leaf sheath of the flag leaf, a process called exsertion. Plants in this stage of growth are typically highly sensitive to water limitation. Any water deficit experienced during GSII can have a drastic impact on the eventual grain yield and quality (Smith and Frederiksen, 2000).

The final stage of growth, GSIII, covers the time from anthesis through grain maturity (Vanderlip, 1993; Gerik, et al., 2003). First, the panicle will release pollen (anthesis). Following this release and subsequent pollination, whether from selfed or outcrossed pollen, seeds begin to develop (Fig. 2b). Seed maturation on the panicle culminates with the development of a dark spot on the seed at the point where it joins the panicle. This spot, called black layer, is the indication that the seed is mature and will be viable if removed from the panicle (Smith and Frederiksen, 2000).

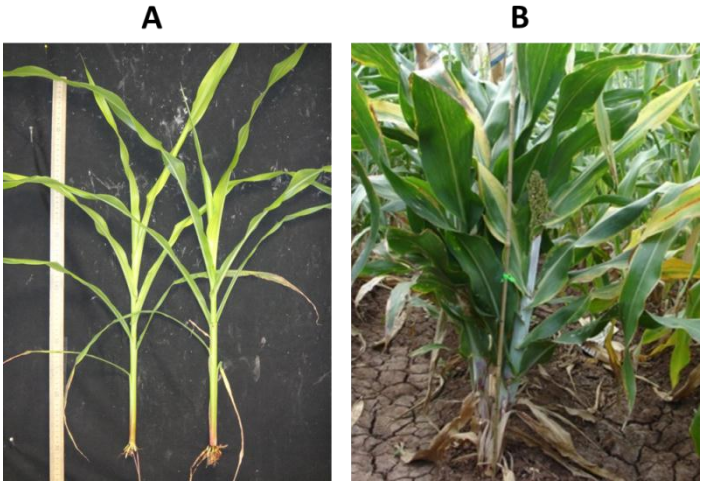

Figure 2: Growth stages of *S. bicolor*. (A) *S. bicolor* plant in GSI, where entire plant is still vegetative. (B) *S. bicolor* plant in GSIII, where anthesis of panicle has begun and panicle is fully exserted from flag leaf sheath.

Sorghum bicolor *varieties and hybrid development*

The range of phenotypic variation within *S. bicolor* is vast, owing in part to the diverse conditions under which it has been grown, both wild and domestic (Dewet and Huckabay, 1967; Teshome, et al., 1997). There are four main categories of *S. bicolor* varieties that are grown domestically, each with specific downstream uses. These are grain, forage, sweet, and bioenergy types (Smith and Frederiksen, 2000). While no single variety is appropriate for all possible uses, the variation that is available within the species makes *S. bicolor* a highly adaptable and therefore highly useful crop, especially in terms of its utility in biomass production for biofuels.

Grain-producing varieties

Grain type *S. bicolor* varieties are characterized by short culms, relatively few leaves, and large, robust panicles at grain maturity (Fig. 3a) (Gerik and Neely, 1987). In general, these varieties have been bred to mature relatively early. Grain *S. bicolor* is ideal for harvesting by combine, making growth and harvesting of this grain highly efficient. Yields reported in 2005 indicate that *S. bicolor* can produce an average of 1.31 MT ha^{-1} of grain worldwide (International Crops Research Institute for the Semi-Arid Tropics, 2011). Starch from *S. bicolor* grain has previously been used to generate biofuel, but fuel conversion efficiency from this grain is slightly lower than that of corn grain, possibly due to decreased digestibility of the starch present in *S. bicolor* grain (Wu, et al., 2007). As such, grain from *S. bicolor* grain varieties is not an ideal source of biomass for biofuel.

***Sorghum bicolor* forage varieties**

Forage varieties of *S. bicolor* are generally grown for livestock feed. These varieties tend to have a higher number of leaves and taller culms than grain varieties, though these plants still tend to be relatively short (Fig. 3b). Culm diameter in these varieties is variable, and some commercial hybrids have very thick culms (Rooney, 2004). Having been developed for use in making hay, silage, and for grazing by livestock, forage varieties have many valuable traits and can generate very high yields of biomass but with low culm-to-leaf ratios.

This makes forage varieties of *S. bicolor* less ideal for generation of biomass for biofuel.

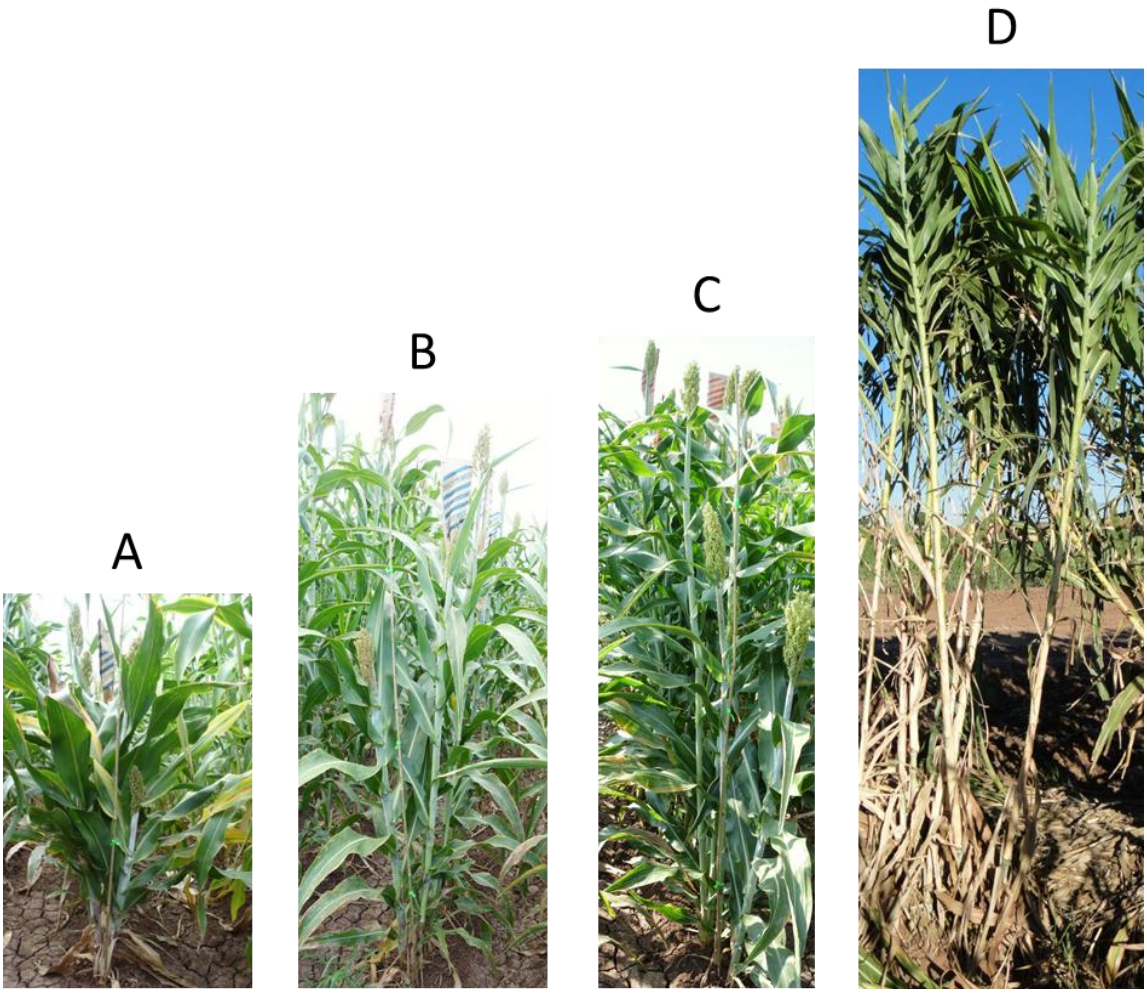

Figure 3: Wide variation in *S. bicolor* morphology. (A) Grain sorghum is characterized by very short stature. (B) Forage sorghum is characterized by a slender stem and many leaves. (C) Sweet sorghum is characterized by a stem with high sugar content. (D) Bioenergy hybrid sorghum is characterized by very tall, thick stems and extremely late flowering. For size reference, green ties on bamboo stakes in photos are 50 cm apart. All photos were taken in the field in College Station, Texas.

Sweet *Sorghum bicolor* varieties

A third group of *S. bicolor* varieties, the sweet sorghums have been used in a variety of ways, including for production of molasses as well as for generation of biofuel (Bennett and Anex, 2009; Rooney, 2004). These varieties are characterized by relatively tall, juicy culms (Fig. 3c). The juice is, as the name would indicate, very sweet, and is reduced into syrup for downstream use. The high sugar content is composed mainly of sucrose, with some glucose and fructose content being typical as well. The juice of these varieties is the desired final product for downstream uses like biofuel generation, unlike grain or forage types (Byrt, et al., 2011). Once the juice is removed, however, the remaining dry material from the culm, called bagasse, is available for conversion to fuel as well (Vermerris, 2011). This dual-phase process significantly increases the utility of sweet *S. bicolor* varieties for use in generation of biomass for biofuels.

The specific sugar content of sweet sorghum juice is highly variable between genotypes and this variation makes it difficult to design a highly efficient downstream fuel production system for the juice (Bennett and Anex, 2009). Unfortunately, the heritability of sugar content in the culms of sweet varieties of *S. bicolor* has been reported to be very complex, meaning that it will require a significant investment of time and effort to generate varieties that do not vary for culm sugar content (Rooney, 2004). As such, sweet varieties of *S. bicolor* have the potential to become an ideal energy crop but still require development to maximize their potential utility in biofuel generation.

Bioenergy hybrid *Sorghum bicolor* varieties

Bioenergy hybrid *S. bicolor* varieties are being developed as a way to capitalize on the capacity of *S. bicolor* to accumulate high biomass yields and the available genetic diversity within the *S. bicolor* germplasm (Fig. 3d). Previous breeding efforts have attempted to include the beneficial attributes of grain-producing varieties with sweet varieties, and the resultant hybrids have generated high yields of biomass (Zhao, et al., 2009). Another strategy, which is unique in its approach to yield improvement, involves the creation of photoperiod-sensitive (PS) hybrid varieties of *S. bicolor* (Rooney, et al., 2007; Rooney and Aydin, 1999). These hybrids were created to increase yield by increasing the duration of vegetative growth. An additional benefit of the delayed flowering of PS hybrid varieties is an increase in drought tolerance. As previously described, the reproductive phase of the *S. bicolor* growth cycle (GSII) is the most sensitive to limited water (Vanderlip, 1993). If the plant never enters GSII, there is less likelihood that water deficit will cause crop failure.

These promising PS bioenergy hybrid *S. bicolor* varieties were developed as an unanticipated product of a cross between two photoperiod-insensitive (PI) genotypes (Rooney and Aydin, 1999). Such hybrids are the main focus of the experiments described in this dissertation. This research shows how bioenergy hybrid *S. bicolor* genotypes are an ideal group of energy crops, for a multitude of reasons, which will be described in the following pages. In addition, experiments described herein identify genes and genomic regions of high importance for

future improvements to the utility of bioenergy hybrid *S. bicolor* varieties in the generation of lignocellulosic biomass for biofuel.

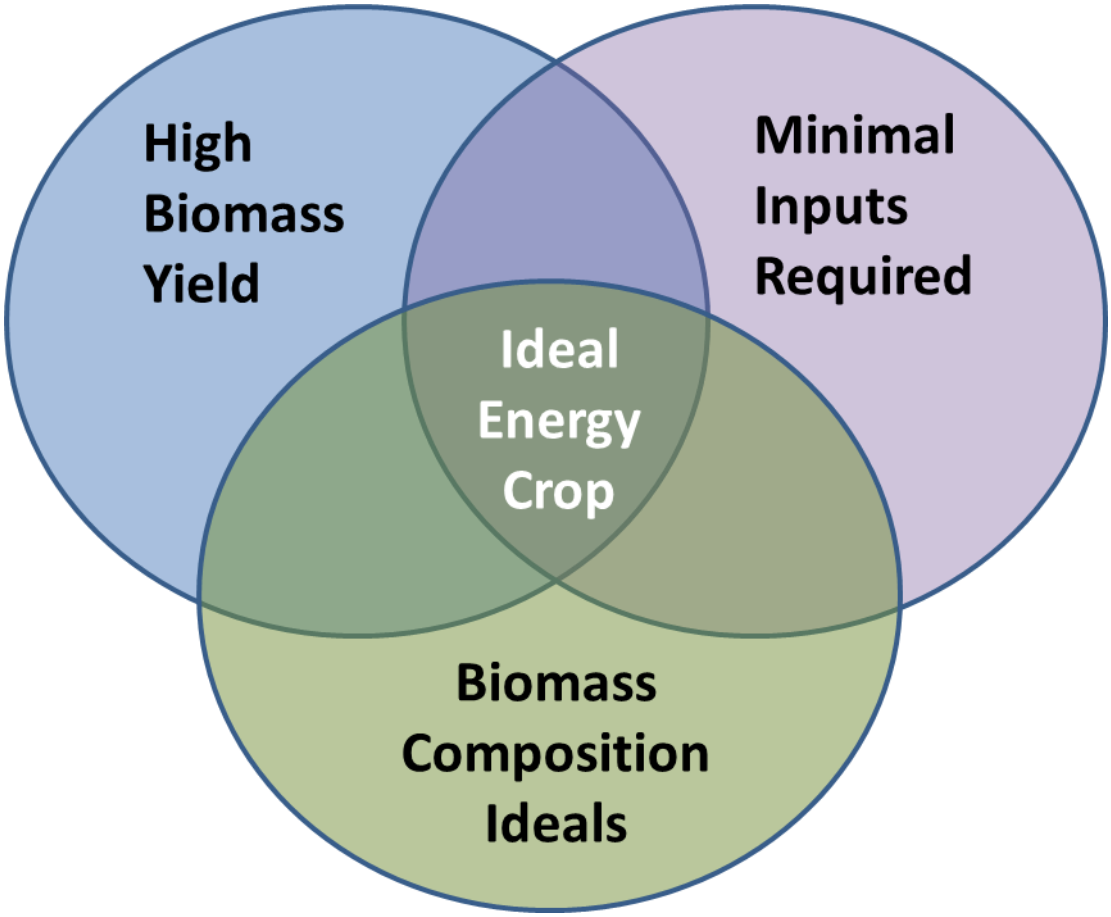

Figure 4: Attributes that make a crop ideal for use in generation of biomass for biofuels. Each of these attributes will be addressed as they pertain to *S. bicolor*. This dissertation will make the case that *S. bicolor* is an ideal candidate crop for use in generation of biomass for biofuels.

Attributes of an ideal energy crop

Cellulosic ethanol is generated by processing vegetative plant material, such as culms and leaves, into liquid fuel (Byrt, et al., 2011; Sang, 2011;

Schmer, et al., 2008). The yield of fuel from such processes is variable and highly dependent on the quantity and quality of input plant material. There are many attributes of a crop that may play a role in its ability to generate cellulosic ethanol. The most important of these attributes are generation of a large quantity of biomass, requiring minimal inputs for growth, and the composition of the biomass that is generated (Fig. 4) (Rooney, et al., 2007; Vermerris, 2011; Wang, et al., 2008). Each of these attributes must be optimized in order to develop an ideal energy crop.

Factors affecting biomass yield

Optimizing biomass yield requires consideration of multiple factors. First, it is important to maximize the rate of generation of biomass and mature plant size. It is also important, however, to manage the duration of vegetative growth in order to maximize biomass generated.

Rate of biomass generation, genetics of plant size

The rate of generation of biomass for a given crop can be affected by both environmental and genetic factors (Meyer, et al., 2007; Bhandari, et al., 2011). The environmental factors affecting the rate of biomass generation for a crop are many and varied. The availability of resources, such as water and nitrogen, is an important determinant of plant growth rate and will be discussed in depth later (Hons, et al., 1986; Clifton-Brown and Lewandowski, 2000; Lawlor, 2002). Plant spacing within a field situation can also have an effect on rate of plant growth, as the shade avoidance response will cause a plant to grow taller

at a faster rate if shaded by other plants (Robinson, et al., 1964; Hedge, et al., 1976; Caravetta, et al., 1990). Rate of canopy development by a crop can also affect the rate of biomass generation; in the case of *S. bicolor* for example, biomass accumulation has been demonstrated to reach its maximum rate following the closure of the leaf canopy (Muchow and Sinclair, 1994; van Oosterom, et al., 2010). The rate of biomass generation, whether in favorable or unfavorable conditions, is important to consider and an ideal energy crop will be able to generate biomass rapidly regardless of its environmental conditions. This research demonstrates how environmental factors can affect growth rates of divergent *S. bicolor* genotypes including grain types and bioenergy hybrids.

Aside from environmental factors contributing to variation in biomass generation rate, there are also genetic factors to consider. These factors can affect not only rate of biomass generation, but also eventual plant size. A measure of a plant's ability to use incident sunlight for growth, radiation use efficiency (RUE), is an important determinant of growth rate (Bégué, 1993; Sinclair and Muchow, 1999; Stockle and Kemanian, 2008; Stockle and Kemanian, 2009). The higher a crop's RUE, the more biomass it can generate from a given quantity of incident light. Specific plant morphological traits relating to eventual plant size can also be studied to yield clues about genetic control of plant growth (Gerik, et al., 2003; Meyer, et al., 2007; Quinby, 1974; Foster, et al., 1994). Many individually measured traits have been shown to correlate well with total yield, including culm length, culm diameter, and leaf area. An ideal energy

crop must have a genetic makeup that allows for rapid generation of large quantities of biomass as well as a predisposition to generate large plants. This dissertation identifies regions of the *S. bicolor* genome that are associated with plant growth traits that correlate with eventual plant size.

Duration of vegetative growth

Many monocot plant species, including *S. bicolor*, follow a distinctive growth pattern where vegetative growth ceases with the induction of the flowering process (Smith and Frederiksen, 2000; Quinby, 1972). This means that, in order to maximize duration of vegetative growth, and thereby maximize biomass yield, it is necessary to delay induction of flowering. Timing of flowering is a complicated process in plants and includes inputs from many different pathways (Fig. 5). The flowering time regulatory pathway integrates information about day length and irradiance, temperature, internal nutrient status, and phytohormones like gibberellins (Quinby, 1972; Quinby and Karper, 1945; Murphy, et al., 2011; Blázquez, 2000).

In terms of floral induction, *S. bicolor* is a short-day plant. This means that, when grown in conditions where the duration of light exposure each day is less than 12 hours, *S. bicolor* will be induced to flower once the juvenile phase has been completed (10-30 DAE) (Smith and Frederiksen, 2000). This induction occurs when the output from an internal molecular clock during its evening phase occurs in darkness and a signal generated by the gene FT is sent to the shoot apical meristem, initiating development of the inflorescence meristem

(Murphy, et al., 2011). The molecular clock mechanism is not well understood in *S. bicolor* beyond the knowledge that the sorghum orthologs of the core clock genes TOC1 and CCA1 show oscillating expression in a circadian manner as found in other plants (Rooney and Aydin, 1999; Murphy, et al., 2011).

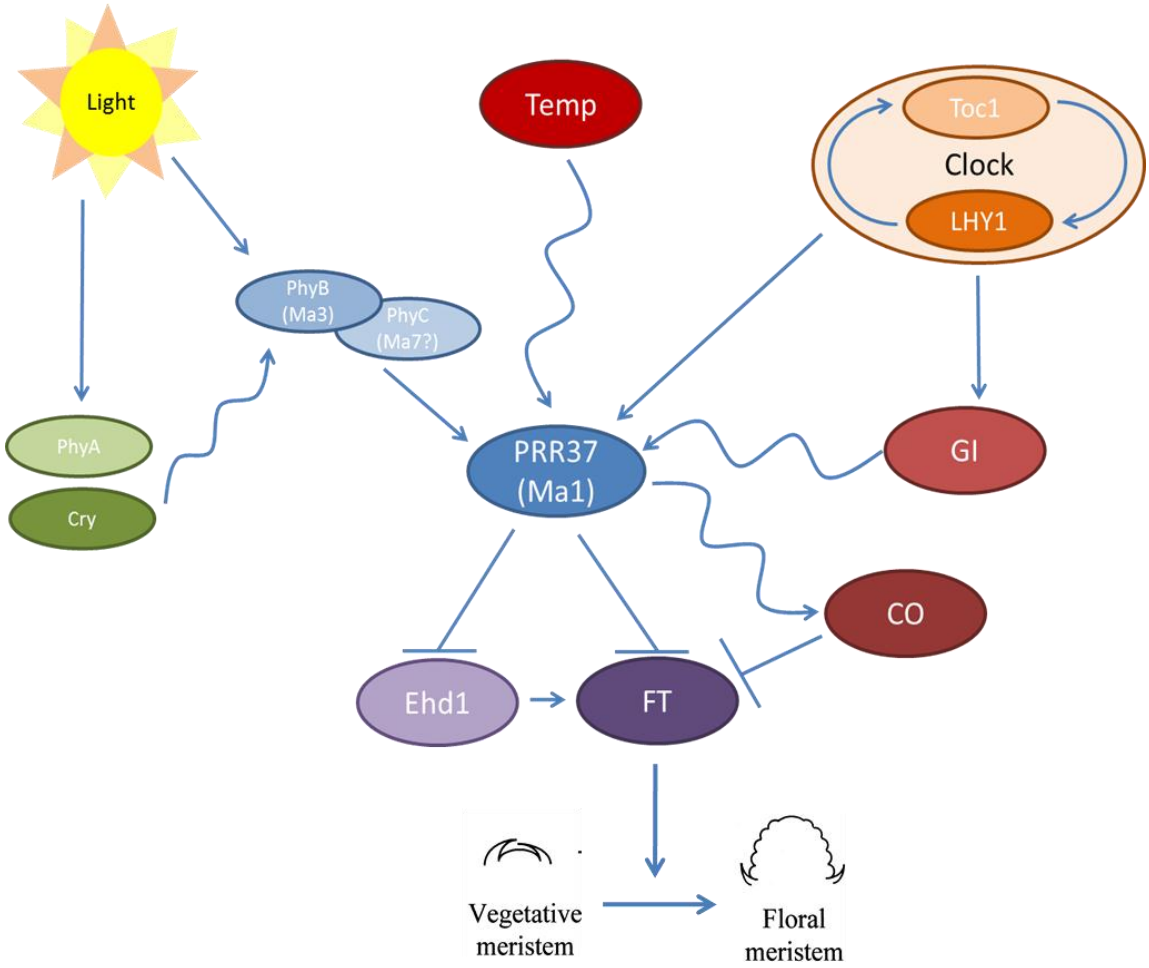

Figure 5: Control of flowering time in *S. bicolor*. This figure depicts contributions to flowering time control by environmental cues and the intrinsic circadian clock. Curved lines indicate a putative relationship, while straight lines indicate an experimentally verified relationship between factors/genes.

Within the floral regulatory pathway that mediates responses to photoperiod in *S. bicolor*, there are six loci, named maturity locus 1 (*Ma1*) through maturity locus 6 (*Ma6*), which are known to have effects on floral induction (Rooney and Aydin, 1999; Quinby and Karper, 1945; Quinby, 1966). Some of these loci have been previously identified through genetic and expression studies (Table 2) (Murphy, et al., 2011; Childs, et al., 1997). Others have not yet been identified. One of the loci remaining to be identified is *Ma2*. This research has identified the gene responsible for the activity of the *Ma2* locus, which contributes to efforts aimed at maximizing the duration of vegetative growth, and hence biomass yield, of *S. bicolor*.

Table 2: The maturity (Ma) loci in sorghum. Loci for which candidate genes have been identified are noted. The identity of Ma2 is investigated in this dissertation.

Ma Locus	Identity	Citation
Ma1	PRR37	Murphy et al., 2011
Ma2	Unknown	-
Ma3	Phytochrome B	Childs et al., 1997
Ma4	Unknown	-
Ma5	Unknown	-
Ma6	GHD7	Unpublished data

Minimal input requirements for growth

Another important attribute of an ideal energy crop is the requirement of minimal inputs such as nitrogen, phosphate and water for optimal growth. That

is, the less input required to generate maximum biomass output, the more ideal an energy crop. The most vital external inputs for plant growth are water and fertilizer, specifically nitrogen fertilizer (Lewandowski and Schmidt, 2006; Hons, et al., 1986; Lawlor, 2002; Bowman, 1991; Foyer, C.H., Hanma, Z., 2011; Tamang, et al., 2011). As is the case in any agronomic situation, the cost of these inputs must be taken into account when assessing the value of a crop's yield. Growing any crop must necessarily include these inputs at some level, and an ideal energy crop will be one that can generate large quantities of biomass with the fewest possible inputs.

Water

Water is of course a vital part of plant growth and development and, hence, biomass generation. Crop productivity is significantly reduced when rainfall or irrigation is inadequate (Apariciotejo and Boyer, 1983; Chaves, et al., 2003). Plant responses to water limiting environments can be grouped into three categories: drought escape, short-term dehydration avoidance, and long-term dehydration tolerance. Drought escape is a phenomenon wherein a plant will modify its lifecycle in order to reach maturity and set seed without being affected by a lack of water (Tuinstra, et al., 1997). As the name indicates, the plant "escapes" drought by maturing before the effects of minimal water availability can be felt. Dehydration avoidance responses are those used by the plant to minimize water loss in cases of limited water (Mutava, et al., 2011). Often, these responses can be observed in a plant prior to any measurable

decrease in the plant's water status but can have measurable effects on the plant's growth. One such short-term response is the closure of stomata on the leaves (Smirnoff, 1993). While this action minimizes water loss through the leaves, it also precludes normal uptake of carbon dioxide. This, in turn, can lead to decreased carbon fixation, decreased biomass generation, and an increase in generation of damaging reactive oxygen species (Chaves, et al., 2003).

Biomass losses due to short-term water deficits are often minimal when the deficit is quickly replenished (Chaves, et al., 2003; Lawlor and Cornic, 2002). If water remains limiting, it becomes necessary for the plant to shift into long-term water deficit management strategies, referred to as dehydration tolerance responses. Such strategies, while able to increase plant survival, are often accompanied by significant metabolic shifts from biomass generation to maintenance. Intentionally limiting shoot growth is one such response (Buchanan, et al., 2005). A smaller plant requires less water for survival and thus can survive longer in a limited water situation. In conjunction with limiting shoot growth, water stressed plants may also increase root growth to maximize the ability to retrieve water from a limited environment (Dugas, et al., 2011). In terms of biomass generation by an energy crop, these responses to limited water would be detrimental to biomass yield.

As it has already been demonstrated, both biomass generation rate and duration of vegetative growth are important determinants of a plant's ability to generate biomass for biofuel. The above described responses to limited water

directly affect both of these attributes in a negative way. Any plant considered to be an ideal energy crop must necessarily be able to grow normally under relatively low water conditions and must also be able to recover normal growth rapidly if water is returned following a period of deficit (Nguyen, 2004). This dissertation examines the effects of water limitation on the biomass generation ability of a bioenergy hybrid *S. bicolor* genotype. While *S. bicolor* is reported to be highly drought tolerant, there is little data available as to its specific ability to generate biomass under conditions of limited water. This data makes it possible to determine an optimal watering strategy for maximizing biomass generation by *S. bicolor*.

Nitrogen

In addition to water, nitrogen-containing fertilizer is a very important input for optimal plant growth. Many studies suggest a close link between nitrogen uptake and water status for plants (Heaton, et al., 2004; Sage and Pearcy, 1987b). This relationship only increases the importance of nitrogen in maximizing biomass yield from energy crops. Application of nitrogen through nitrate and/or ammonia containing fertilizer is a necessary but very costly part of biomass generation for biofuel production (Sainju, et al., 2006). All crop plants require some level of fertilization, and an ideal energy crop will yield high levels of biomass without requiring concomitant high levels of fertilization. The cost considerations of fertilization are both monetary and environmental. It has been estimated that, of the average costs associated with production of biofuel from

cellulosic sources, approximately 15% of the total cost comes directly from supplying nitrogen to the growing plants through fertilization (Wang, 1996). In an economic situation where biofuels are in direct competition with fossil fuels, minimizing the cost of this input is imperative. An ideal energy crop will be one that can generate above-average biomass without requiring above-average fertilization. Environmentally, production of fertilizer is expensive as well. Current methods for fertilizer production involve burning large quantities of fossil fuels to generate ammonia (Frink, et al., 1999). This process increases air pollution as well as increases the demand for fossil fuels, which serves to exacerbate rather than mitigate the effects of the limited global fuel supply.

Nitrogen is necessary for photosynthesis and plant growth. The central molecule for photosynthesis, chlorophyll, contains four atoms of nitrogen per molecule (Sage and Pearcy, 1987a). Besides carbon, there is no other element so critical for the survival of an organism as nitrogen. In addition to being necessary for chlorophyll synthesis and function, nitrogen is an important part of ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO), a vital part of the photosynthetic process and the single most abundant protein in the world (Ghannoum, et al., 2005). Reported plant tissue nitrogen content varies somewhat, and it has been reported that the critical nitrogen content for live plant tissue is 3% of total weight (Borrell and Hammer, 2000).

In order to adequately quantify a crop's ability to generate biomass based on its level of fertilization, a measure has been developed that relates biomass

to nitrogen content. This measure is nitrogen use efficiency (NUE), which is reported here in terms of total biomass generated per mass of nitrogen in the biomass tissue (Ranjith and Meinzer, 1997; Gaju, et al., 2011). This calculation provides a measure of how efficiently a plant can generate biomass with the nitrogen it has taken up. This measure assumes that the plants are growing in conditions of sufficient nitrogen availability. An alternative method of calculating NUE is on the basis of biomass per N applied to the field (Cui, et al., 2009). This measure of NUE includes the ability of a plant to take up, assimilate and reuse N for biomass production at a given N application rate (which is directly related to the cost of the N input). The latter method of calculation allows for differentiation between crops in conditions of limited nitrogen availability. The higher a crop's NUE, the more biomass it can generate from a given quantity of nitrogen, and thus the lower its input requirement for biomass generation. NUE can vary greatly between species as well as between different genotypes of a single species (Lewandowski and Schmidt, 2006; Jorgensen, 2011; Heaton, et al., 2004; Tamang, et al., 2011; Beale and Long, 1997).

There are many factors that can affect a plant's NUE including leaf size and thickness, leaf number, and culm-to-leaf ratio (Ranjith and Meinzer, 1997; Gaju, et al., 2011). In addition, there may be variation in the physical distribution of nitrogen within a plant's leaves. It is also important to consider that, in the case of grass species like *S. bicolor*, the culm tends to have significantly less

photosynthetic activity than the leaves and as such, culm nitrogen content may have a significant influence on total plant NUE (van Oosterom, et al., 2010).

It has been suggested that, in the case of extremely long-growing plants, the productivity of biomass generation will decrease over time as nitrogen is removed from the soil for use in construction of new plant tissues (van Oosterom, et al., 2010). Concerns have been raised that, in order to grow photoperiod sensitive *S. bicolor* hybrid plants to maturity, a second application of fertilizer may be necessary at the halfway point of the growing season in order to see continued growth until the plants mature (Yamoah, et al., 1998). An ideal energy crop will not require additional fertilizer applications and will have a high NUE, resulting in a large quantity of biomass being generated from minimal input nitrogen. This research quantifies the nitrogen needs and NUE of bioenergy hybrid *S. bicolor* genotypes and the partitioning of nitrogen between different organs of the *S. bicolor* plant to facilitate development of optimal fertilization and harvesting strategies for maximizing biomass output from a minimal fertilizer input.

Composition considerations for biofuel generation

Ideal biomass composition

There are many different strategies that may be employed to generate fuel from biomass. Each process has different steps, yielding a different product through a specific method. These processes are known as conversion technologies, and while each process begins with biomass input and ends with

fuel output, the products are very different from one technology to the next (Dale, 1987). For each conversion technology, the chemical and physical processes are different, resulting in different types of fuel being generated. Naturally, the ideal input for each of these technologies is different. An ideal energy crop will generate biomass that is optimized for one or all of these technologies, or will be easily manipulated to create changes in the composition of generated biomass.

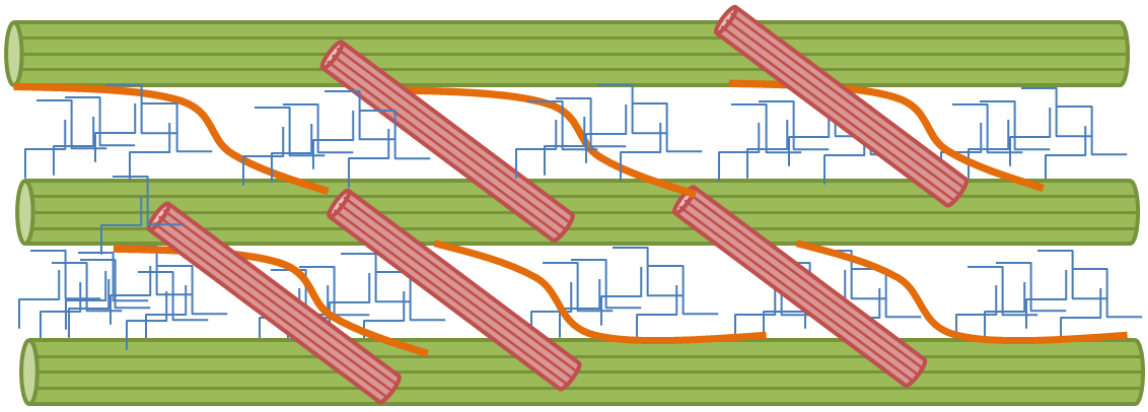

Figure 6: Schematic representation of cell wall architecture in *S. bicolor*. Cell wall matrix is a network of many polymers. Cellulose (green) forms microfibrils that are the basic structure of the cell wall. Structural proteins (pink) act as stiff cross-linking molecules that hold cellulose microfibrils together. Xylan (orange) fibers are flexible joiners that create a stretchy network with the stiffer cellulose microfibrils. Finally, lignins (blue) create a dense network of phenolic molecules that give the cell wall strength.

Plant tissue, like animal tissue, is made up of cells. Unlike animal cells however, plant cells are surrounded by a rigid cell wall (Wilson, et al., 1993). This cell wall is made up of a complex web of macromolecules which provide

strength and protection for the cellular contents while also helping the whole plant to maintain its rigid, upright structure. The main cell wall components in *S. bicolor*, for example, are cellulose, hemicellulose (xylan), and lignin (Fig. 6) (Boerjan, et al., 2003). There are other parts of the plant that must be considered when assessing biomass composition. Cell wall composition is important on a micro scale, but there are macro scale composition traits to consider. The vasculature of a plant carries fluids and soluble molecules throughout the plant (Murray, et al., 2008b). One of the most important classes of molecules being transported in this vasculature is soluble sugars. Not every cell in a plant is able to fix carbon and generate sugar at a sufficient level for its own survival. As such, some portion of the sugar molecules synthesized in actively photosynthesizing cells must be transported throughout the plant in the vasculature (Byrt, et al., 2011). There is wide variation with respect to the sugar content of the fluid circulating in a plant's vasculature. This circulating fluid can be extracted from culms by crushing and pressing the tissue to yield juice (Bennett and Anex, 2009; Zhao, et al., 2009; Tamang, et al., 2011). Both plant cell walls and this juice are relevant to downstream biofuel conversion technologies.

Conversion technologies

Of all conversion technologies as yet developed, a select few are most appropriate for use with lignocellulosic biomass like that generated by bioenergy hybrid *S. bicolor* genotypes. There are three main conversion technologies that

will be considered here; fermentation, conversion, and pyrolysis. Fermentation is a process used on biomass with a high concentration of fermentable sugars like glucose and sucrose (Dale, 1987). These sugar molecules are fermented enzymatically to create ethanol. Conversion is a process by which long chain biopolymers like cellulose and hemicellulose are first hydrolyzed, and then fermented (Wu, et al., 2007). Cellulose and hemicellulose are converted to glucose and xylose, respectively. Following hydrolysis, the resultant sugars can be fermented to create ethanol as described in the direct fermentation method (Dale, 1987). Pyrolysis is a conversion technology that uses heat and pressure to yield biodiesel from the lignins and phenolic compounds in lignocellulosic biomass (Boerjan, et al., 2003; Weng, et al., 2008). This process extracts energy from phenolic compounds in the plant tissue rather than from carbohydrate-based molecules as is the case in the first two conversion technologies described here.

Strategies for improvement of energy crops

Sorghum bicolor *domestication*

Directed improvement of domesticated crops is not a new phenomenon. Since the dawn of agriculture, practices have been aimed at increasing yield, ease of harvest, and the quality of the resulting products (Doebley, et al., 2006). A variety of techniques have been employed over the years, with technological advances playing a significant role in the rate of advancement of such techniques (Gepts, 2002). At the most basic level, crop improvement is

achieved by identification of desirable traits followed by selection and advancement of individuals exhibiting those attractive or desirable properties (Dillon, et al., 2007). During the initial domestication process of a crop, such properties may include, among others, ideal timing of maturity, apical dominance, and in the case of grain crops, minimizing grain shatter (Dewet and Huckabay, 1967). Once a crop has been successfully domesticated, improvement efforts shift in focus to crop improvement including ease of harvest and yield (Rooney, 2004).

In the case of *S. bicolor* improvement, breeding programs have focused on many traits including; development of high yield, both in terms of grain and biomass, improved quality of grain, juice, and whole biomass for downstream applications, and increased tolerance of stresses such as limited water and nutrients (Monk, et al., 1984). The first *S. bicolor* varieties grown in the United States were typically tall, had low seed yields, and tended to lodge (fall over) (Smith and Frederiksen, 2000). Initial improvements were made by farmers, selecting either random mutants or outcrossed individuals that appeared different within their fields (Rooney, 2004). Continual improvements, including application of advanced technological methods have led to the development of the *S. bicolor* varieties in use today.

Technology-driven advances in crop improvement

Technological advances have improved selection abilities for crop improvement. Methods are no longer limited to visual selection of individuals

exhibiting desirable traits for advancement. One important realm of advancement is improved phenotype measurement ability (Xin, et al., 2008). Many desirable traits of high agronomic value are actually complex conglomerations of many individual component traits (Hart, et al., 2001). Direct measurement of such complex traits for breeding is often difficult. Technological advances that facilitate measurement of the individual component traits contribute to advances in selecting for improvement in those complex but high value traits (Casa, et al., 2008). Efficiency of conversion of lignocellulosic biomass into biofuel is an example of a highly complex trait that can be more easily selected for by using technological advances (Dale, 1987; Himmel, et al., 2007). In order to predict a genotype's conversion efficiency, its biomass must be assessed in terms of individual components. Use of technology like near-infrared spectroscopy (NIR) to determine biomass composition allows for selection of varieties based on a wholly complex and previously immeasurable phenotype (Sanderson, et al., 1996a; Owen Reece, 1999; Roberts, et al., 2011).

Another area of technological advancement that has contributed in a significant way to energy crop improvement has been DNA sequencing and genotyping technology (Goodstein, et al., 2012; Xin, et al., 2008). While conventional breeding programs involve phenotype measurement and selection alone, advanced breeding programs apply genotyping techniques to select for desirable traits more accurately than would otherwise be possible through phenotype measurement alone (Bouton, 2007; Nguyen, 2004; Asins, 2002). In

order to use genotyping to facilitate crop advancement, it is necessary to first identify a DNA sequence variant, or marker, that is associated with plants exhibiting a favorable phenotype but not associated with plants exhibiting an unfavorable phenotype. Initial forays into this type of strategy have been carried out using a process called marker assisted selection (MAS). In MAS programs, a trait is measured in a group of plants and DNA sequencing is carried out on those plants to identify a genetic marker that is associated with the observed variation for the measured trait (Collard, et al., 2005). Once a marker has been identified, it is used in place of phenotype measurements to identify desirable individuals for advancement. DNA from other individuals is sequenced to determine presence or absence of the identified marker. Lines which carry the favorable marker are assumed to carry the favorable phenotype as well and are selected for advancement without requiring phenotype measurement. Though MAS has some limitations, it is a useful strategy for improving crops with respect to any trait that is expensive or difficult to measure, as such cumbersome phenotypic measurements can be replaced with relatively simple and cost-effective DNA sequencing once a marker has been identified (Rooney, 2004).

Crop improvement on a genetic level

MAS is a very useful tool, but many previous studies have indicated that gains from MAS programs are not as significant as expected, indicating that there are additional factors affecting phenotypes that are not controlled for when using a traditional MAS approach (Collard, et al., 2005). Typically, MAS

identifies a marker sequence but does not necessarily generate any information about the gene or genes responsible for the phenotypic variation that has been shown to correlate with that marker. This shortcoming has led to yet another advanced strategy for crop improvement: quantitative trait locus (QTL) mapping. QTL mapping strategies draw on concepts employed by MAS, expanding the genetic aspect to include the entire genome (Hackett, 2002; Rice, et al., 2001). In this strategy, portions of the genome (loci) are identified that are tightly correlated with phenotypic variation for a given trait. While follow-up activities are often required to conclusively identify the gene or genes acting to control a given trait, QTL maps are highly useful tools for identifying the cause (or causes) of variation in a trait of interest (Asins, 2002).

Often, traits of agronomic interest are controlled by the actions of genes that are part of large, complex pathways within the plant (Murray, et al., 2008b; Murray, et al., 2008a). This means that manipulation of a single gene known to have an effect on a given phenotype can often have unintended consequences for the remainder of the pathway that said gene is involved in. By carrying out QTL mapping it is possible to identify all portions of the genome that play a role in determining the phenotype for a given trait. This practice can contribute to improvement by giving an accurate representation of all of the factors affecting a trait of interest. When used in combination with other practices like fine mapping, haplotype sorting, and DNA sequencing, it is possible to identify the individual gene or genes that are acting to control phenotype values for a trait

(Murphy, et al., 2011). This information is often invaluable as it can indicate whether alteration of the phenotype, through alteration of the genotype, is likely to have unintended consequences for other traits.

In practice, QTL mapping strategies for *S. bicolor* begin with construction of a recombinant inbred line (RIL) population or other types of populations (F2, F3, etc.). This is done by first crossing two divergent *S. bicolor* lines (parental lines), then self-pollinating the offspring for multiple generations to reduce heterozygosity and generate RILs (Anderson, et al., 1993). The trait of interest is measured for each of these resultant RILs, and DNA is isolated from each RIL for genotyping. DNA from each parental line and each RIL is sequenced, and a genetic map is generated for the entire genome.

The final step in the process is the integration of the genetic map with the measured phenotypes to create the QTL map (Asins, 2002; Collard, et al., 2005). Such a map shows the statistical likelihood of a correlation between genotype and phenotype for every possible location within the genome (Churchill and Doerge, 1994). Using a measure called the log of odds (LOD), it is possible to score each location and determine a minimum threshold above which a correlation can be considered to be significant (Rice, et al., 2001). QTL mapping strategies are used frequently in this thesis to identify genomic regions, and occasionally individual genes, that are involved in observed phenotypic variation for a variety of traits (Fig. 7). Such traits include control of flowering

time, regulation of plant size and vegetative growth, and control of biomass composition.

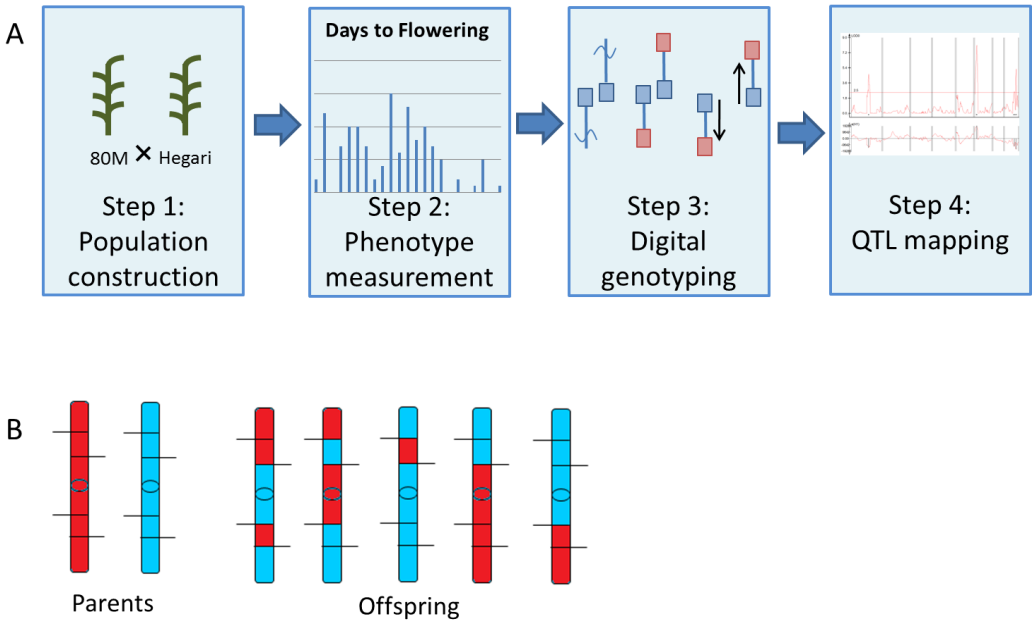

Figure 7: QTL mapping workflow. This begins with measurements of phenotypes and ends with identification of discrete QTL for each trait. (A) The workflow can be broken up into four stages. First, a population is constructed based on a measurable difference in a quantitative trait. Second, the population is grown out and the phenotypes of the offspring are recorded. Third, DNA is extracted from each of the offspring in the population and this DNA is analyzed using digital genotyping. Fourth and finally, QTL mapping is carried out through combination of genotype and phenotype data. (B) Idealized representation of a genetic map that would result from carrying out the QTL mapping workflow described in this chapter.

Roadmap for this dissertation

Each of the chapters of this dissertation examines one or more of the described attributes of an ideal energy crop as it applies to *S. bicolor*. Taken together, these studies provide a comprehensive analysis of the attributes of *S.*

bicolor that are ideal for a dedicated energy crop. In addition, the research reported herein contributes significantly to the understanding of the genes and genomic loci that control a multitude of biomass generation related traits in *S. bicolor*.

Chapter two describes the results of a growth cycle analysis of both grain and bioenergy hybrid *S. bicolor* genotypes. This experiment provides an understanding of the growth and development of *S. bicolor* in terms of the rate of biomass accumulation of diverse genotypes, the duration of vegetative growth of diverse genotypes, and the effects of water deficit on biomass accumulation in a bioenergy hybrid *S. bicolor* genotype.

Chapter three reports on the NUE of a bioenergy hybrid *S. bicolor* variety. The experiments described in this chapter elucidate nitrogen content of *S. bicolor* tissue as well as the removal of nitrogen from the soil by *S. bicolor*. These data allow for the calculation of NUE for these hybrid genotypes. In addition, nitrogen partitioning into different tissues is reported on a per-leaf and per-stem section basis for mature bioenergy hybrid *S. bicolor* plants. The data presented in this chapter provide a detailed description of the NUE of bioenergy hybrid *S. bicolor* varieties, an important attribute of an ideal energy crop.

Genes and genomic loci that control various traits in *S. bicolor* are reported in chapter four. This chapter considers not only the growth, size, and composition phenotypes measured in *S. bicolor* populations, but also reports on QTL identified for these traits. This chapter demonstrates that the size, biomass

accumulation patterns, and composition of biomass generated by bioenergy hybrid *S. bicolor* varieties makes them ideal for conversion to biofuel. Identifying the gene or genes controlling each of these traits contributes to the understanding of growth rate and regulation of biomass accumulation. These genes have a direct effect on rate of biomass accumulation and individual plant size, important traits that make *S. bicolor* an ideal energy crop. These data provide information as to how ideal biomass from bioenergy hybrid *S. bicolor* plants is for downstream conversion to biofuel using any of the previously described conversion strategies. The biomass was separated into its component tissues prior to phenotype measurements and composition analysis. As such, the appropriateness of each tissue for in terms of biomass accumulation as well as conversion into biofuel is reported independently.

Chapter five details the set of experiments carried out in pursuit of identifying the gene in *S. bicolor* that are heretofore referred to as *Ma2*. Identification of this gene was a multi-year process that required QTL mapping and DNA sequencing efforts involving multiple generations of three different *S. bicolor* populations. The contributions of each of these studies to the final gene discovery are reported here. By increasing understanding of the genes that control flowering time in *S. bicolor*, including *Ma2*, it is possible to manipulate flowering time to increase duration of vegetative growth, a key attribute of an ideal energy crop.

In chapter six, conclusions and future directions are reported. Many of the results presented in this dissertation have more weight when considered through the lens of all of the chapters together. As such, this final chapter will elucidate some of the more broad-reaching conclusions that can be drawn from this entire dissertation. In addition, the sixth chapter includes suggested future experiments that would advance the findings presented herein.

CHAPTER II

GROWTH CYCLE ANALYSIS OF BIOENERGY HYBRID *SORGHUM BICOLOR*

Background and introduction

Global biofuel demand and United States legislation

World energy consumption is projected to increase by 57% between 2002 and 2025 (National Research Council (U.S.). Committee on Economic and Environmental Impacts of Increasing Biofuels Production., et al., 2011). Increased demand for energy, the cost of oil imports, increased extraction costs of less accessible fossil fuel reservoirs, energy security, and concern over CO_2 emissions led the Department of Energy and the USDA to explore the feasibility of using biofuels to supply up to 30% of the U.S. transportation fuels requirement by 2030 (Perlack, et al., 2005). To reach this goal the United States Energy Independence and Security Act (EISA) of 2007 mandated production of up to 15B gals of biofuels from starch-based grain such as corn, and 21B gals from cellulosic biofuels and other non-grain sources (Rahall, 2007). In 2009, approximately 12B gals of ethanol were produced in the U.S. primarily by fermentation of corn grain starch (United States Energy Information Agency, 2009). Because the EISA limited production of ethanol from grain to ~15B gals to minimize competition between the use of corn grain for food, feed and biofuels, the next substantial increase in biofuels production is likely to be based on conversion of cellulosic biomass to biofuels and production of biofuels from algae.

Usefulness of C4 grasses for biomass generation

The USDA's 'Billion Ton Study' estimated that the U.S. has the potential to produce ~1.3 billion dry tons of biomass for biopower and biofuels generation without compromising food, feed and fiber supplies (Perlack, et al., 2005). Crop and forest residues were identified as sources of biomass as well as biomass derived from a new generation of dedicated energy crops such as the perennial C4 grasses switchgrass (*Panicum virgatum)* and Miscanthus (*Miscanthus x giganteus)* (Dohleman and Long, 2009). The C4 grasses were targeted for bioenergy production because of their high photosynthetic efficiency and capacity for biomass accumulation (Rooney, et al., 2007; Heaton, et al., 2008). Energy crops with high biomass yield are an important aspect of reducing the total acreage required for biofuels production minimizing potential competition for land utilization.

Perennial C4 grasses have long growing seasons and the ability to store nutrients in rhizomes at the end of the growing season for regrowth the following season (Beale and Long, 1997;Carpita and McCann, 2008). In a recent study, Miscanthus was 59% more productive than grain maize (*Zea mays*) in a large-scale trial in the U.S Corn Belt due to the longer duration of Miscanthus growth and higher photosynthetic rates in cooler portions of the season (Dohleman and Long, 2009). Miscanthus had a peak dry biomass of 60 dT ha^{-1} in small-scale plots under optimal conditions (Heaton, et al., 2008) and in larger scale plots, Miscanthus yielded 30.3 dT ha^{-1} in 2007, compared to maize which had a peak

aboveground dry biomass of 19.2 dT ha^{-1} in the same year (Dohleman and Long, 2009). Harvestable Miscanthus yield ranged from 10-40 dT ha^{-1} in different locations in Europe (Lewandowski and Schmidt, 2006) and a meta-analysis of Miscanthus found an average yield of 24.9 dT ha^{-1} (Heaton, et al., 2004). In general, switchgrass produces a lower yield of shoot biomass than Miscanthus, with a dry matter yield of 9.9-23.0 dT ha^{-1} in research trials depending on year and location (Sanderson, et al., 1996b; McLaughlin and Kszos, 2005) and from 5.2-11.1 dT ha^{-1} in established field trials of 3-9 ha size across diverse environments (Schmer, et al., 2008).

Ideal attributes of Sorghum bicolor

Most of the perennial C4 grasses targeted for development as dedicated energy crops are polyploids with large complex genomes that can often be challenging for genetic analysis and plant breeding (Carpita and McCann, 2008). For example, lowland switchgrass, a tetraploid, and upland switchgrass, typically an octaploid, are primarily outcrossing species that are phenotypically and geographically distant (Bhandari, et al., 2011; Bouton, 2007). In contrast, the C4 grass sorghum (*Sorghum bicolor* L. Moench) is a diploid inbreeding species with a relatively small (800 Mbp) (Price, et al., 2005) sequenced genome (Paterson, et al., 2009) and a tractable hybrid breeding system (Rooney, 2004). Sorghum has been grown for over 100 years in the U.S. using well-established sustainable annual grain and forage production systems (Rooney, 2004). Worldwide, sorghum is grown on ~65 MHa, primarily in drought prone regions,

and provides an important staple food source for ~500 million people in more than 30 countries (International Crops Research Institute for the Semi-Arid Tropics, 2011).

Sorghum diverged from rice (*Oryza sativa*) ~50 MYA and this grass species is widely dispersed throughout Africa, India and Australia (Price, et al., 2005; Dillon, et al., 2007). The sorghum germplasm collection derived primarily from Africa is extensive (GRIN, n=41,000), diverse, and enriched in genes for drought tolerance, adaptations to nutrient limitation, and other abiotic and biotic constraints. Over the past 10 years, the sorghum research community has created an extensive set of genetic and genomic resources including integrated genetic/physical and cytogenetic maps aligned to the genome sequence (Kim, et al., 2005), TILLING populations (Xin, et al., 2008), transgenic capability (Mall, et al., 2011), association panels (Casa, et al., 2008), transcriptome information (Buchanan, et al., 2005; Dugas, et al., 2011) and numerous populations for QTL mapping and gene discovery. With all of these tools available, sorghum has the potential to be an excellent reference genomic platform and genetic model for understanding and optimizing the design of dedicated C4 grass energy crops.

Origins of United States biomass crops

Following disruption of oil supplies in the 1970s, a selection of first generation high biomass energy sorghum genotypes was developed specifically for bioenergy production (Hons, et al., 1986). However, large-scale development has occurred only recently (Rooney, et al., 2007) following the

genetic characterization of a regulatory system that modulates photoperiod sensitivity and flowering time in sorghum (Rooney and Aydin, 1999; Murphy, et al., 2011). Using this system, highly photoperiod sensitive, late flowering energy sorghum hybrids have been developed that exhibit long duration of vegetative growth and high biomass accumulation (Rooney, et al., 2007). These energy sorghum hybrids are similar to Miscanthus and sugarcane (*Saccharum officinarum*) in stature and growth duration, providing both a genetic model and complementary annual biomass feedstock for biofuels production.

Goal of this study

The goal of the current study was to characterize the growth and development of a first generation energy sorghum hybrid under field conditions to obtain a baseline of data on this plant's phenology and its potential for biomass accumulation. Overall, this study demonstrates that energy sorghum will be a good genetic model for C4 energy grasses and that the species has exceptional potential as a lignocellulosic crop for biofuels production.

Results

Grain and energy sorghum growth and development

In both years of this study, differences in the size and developmental stage of grain and energy sorghum genotypes were evident (Fig. 8). The grain sorghum genotypes used in this study are photoperiod insensitive and flowered in June, reaching grain maturity in early August. In contrast, the energy sorghum hybrid TX08001 is highly photoperiod sensitive due to the action of six

maturity genes that repress flowering in long days (Quinby, 1974; Rooney and Aydin, 1999; Murphy et al., 2011). As a result, when grown in College Station, TX, the energy sorghum hybrid initiates flowering in mid-September and reaches anthesis in mid to late October.

Figure 8: Photograph of the grain sorghum hybrid 84G62 (left two plots) and the energy sorghum hybrid TX08001 (right two plots) grown with limited irrigation in College Station, Texas on July 23 2008, approximately 90 days after emergence. The grain sorghum has reached the grain filling stage whereas the energy sorghum hybrid is vegetative and will remain so until mid-September.

Grain sorghum genotypes accumulated 50-100 g DW plant^{-1} of shoot dry biomass by 90 DAE (Fig. 9). In contrast, energy sorghum hybrids accumulated 125-175 g DW plant^{-1} of shoot dry biomass over that same period of growth (Fig. 9). Grain sorghum stopped accumulating biomass approximately 120-140 DAE, once plants reached grain maturity. In contrast, energy sorghum hybrids continued to grow rapidly after grain sorghum matured and by the end of October, the total shoot biomass reached ~300-420 g DW plant^{-1}. The highest biomass accumulation occurred in energy sorghum hybrids that were irrigated throughout the growing season (TX08001, 2009).

Based on the measurement of 9 individual plants per time point, on an area basis the energy sorghum hybrids accumulated ~41 dT Ha^{-1} and ~49.5 dT Ha^{-1} under limited irrigation in 2009 and 2008 respectively, and ~59 dT Ha^{-1} under irrigated conditions in 2009. Similar peak shoot biomass yields have been reported for miscanthus based on individual plant measurements (Heaton et al., 2008). The machine harvestable yield of TX08001 from larger plots (5 to 1000 m^2) grown without irrigation ranged from 17.5 dT Ha^{-1} to 25.5 dT Ha^{-1} in 2008 and from 15 dT Ha^{-1} to 21.5 dT Ha^{-1} in 2009 when fields were harvested in early September.

Leaf appearance rates (LAR) were similar for the grain and bioenergy hybrids. By June, both genotypes had produced or initiated ~17-20 leaves (Fig. 10A). The duration of vegetative growth was approximately twice as long for the energy sorghum hybrid, resulting in plants that produced a total of ~45

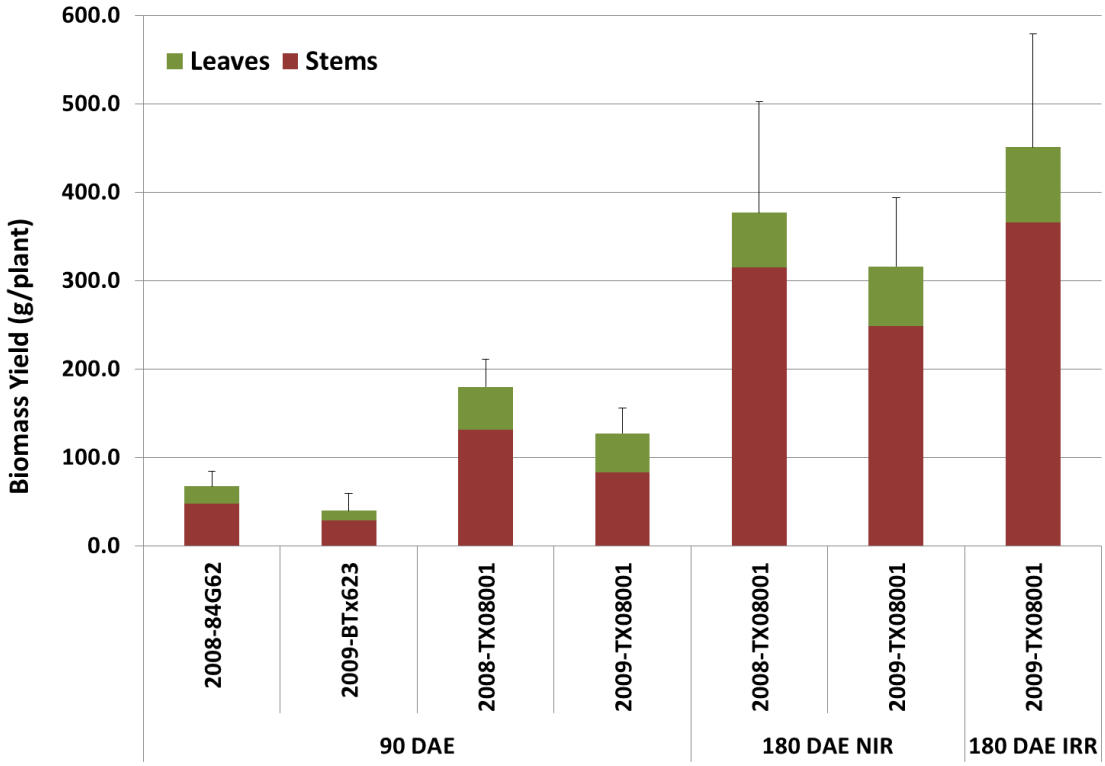

Figure 9: Biomass yield (g DW plant^{-1}) of two early flowering grain sorghum genotypes (84G62, BTx623) and the late flowering energy sorghum hybrid TX08001 in 2008 and 2009. Biomass yield was obtained at 90 DAE or 180 DAE from plants grown with limited irrigation until early July (L-IRR) or with irrigation throughout the season (IRR).Green bars represent leaf biomass while maroon bars represent stem biomass. Nine plants were analyzed at each harvest; error bars represent one standard deviation.

leaves. In contrast the grain sorghums produced ~17-20 leaves that were fully expanded by late June when plants reached anthesis (Fig. 10A).

A large difference in duration of stem growth was also observed (Fig. 10B). The number of internodes and length of the main culm of grain and energy hybrid increased in parallel until June when floral induction occurred in grain sorghum. In grain sorghum, stem elongation (excluding the peduncle)

ceased just prior to anthesis when upper internodes were fully elongated. Grain sorghum main stems reached an average maximum length in early July of around 90 cm with relatively short internodes consistent with the presence of recessive alleles at three dwarfing loci (Hart et al., 2001). Energy sorghum hybrid internode number and stem length increased steadily until September and October, respectively. Increased length of energy sorghum stems was due to higher internode number (due to delayed flowering) and longer internode length resulting in a final average total stem length of approximately 4 meters. The increased internode length of the energy sorghum relative to grain type sorghum indicates that the hybrid is probably recessive for only two dwarfing genes.

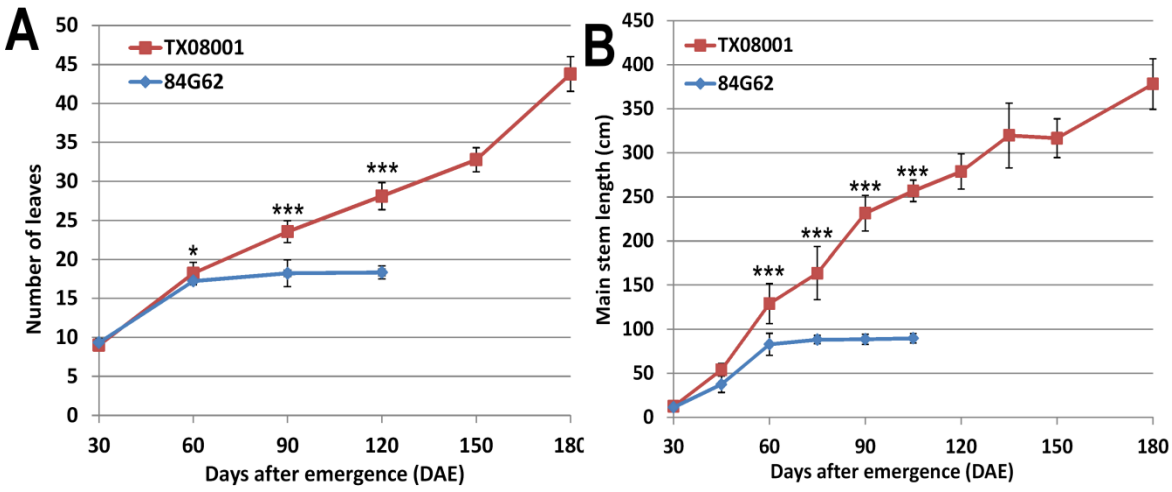

Figure 10: Time course of leaf production and stem growth of the energy sorghum hybrid TX08001 (maroon lines) and the grain sorghum 84G62 (blue lines). (A) Number of leaves produced per plant and (B) main stem length (cm) at different times after plant emergence (DAE). For each harvest, n=9 plants and error bars represent one standard deviation. (***, $p < 0.01$; *, $p < 0.1$)

Biomass accumulation and partitioning

Total shoot biomass of grain and energy hybrids increased in parallel until 40-60 DAE (Fig. 11A). After 60 DAE, the rate of total biomass accumulation in the grain sorghum hybrid decreased as plants reached anthesis and entered the grain-filling phase. Because of the relatively late planting date, midge damage minimized grain production and this contributed to the decreased rate of biomass accumulation in grain sorghum during this phase of development. In contrast, the rate of biomass accumulation in the energy sorghum hybrid continued at a high rate until the harvest on August 20 2008. From August 20 until mid-September, no further shoot biomass accumulation occurred, most likely due to a lack of water (Fig. 11B and the table on page 196). When non-irrigated energy sorghum hybrids stopped accumulating biomass in August, most if not all of the subsoil plant-available moisture was depleted. The lower leaves of these plants senesced but the plants recovered and produced new leaves after rainfall in September. On September 14 2008, severe winds and rain from Hurricane Ike caused significant root lodging. The lodged plants reoriented the upper portion of their stems/leaves, allowing shoot growth to resume and continue until the final harvest (Fig.11A).

Leaf biomass accumulated rapidly in both grain and energy sorghum hybrids during the first 30-50 days of growth and then more slowly when stem growth accelerated (Fig. 11B and C). The duration of biomass accumulation in leaves and stems was shorter in grain hybrids ($p < 0.01$), ending approximately

at anthesis in late June (Fig. 11B and C). In energy sorghum, from 60 DAE onwards, leaf biomass accumulated at a slow rate compared to stem biomass (Fig. 11B and B). In energy hybrids, most of the shoot biomass accumulated in stems after July 1, resulting in in an increasing ratio of stem/leaf biomass during the course of development (Fig. 12). By the end of the season, approximately 83% of the above-ground biomass of energy sorghum was present in stems.

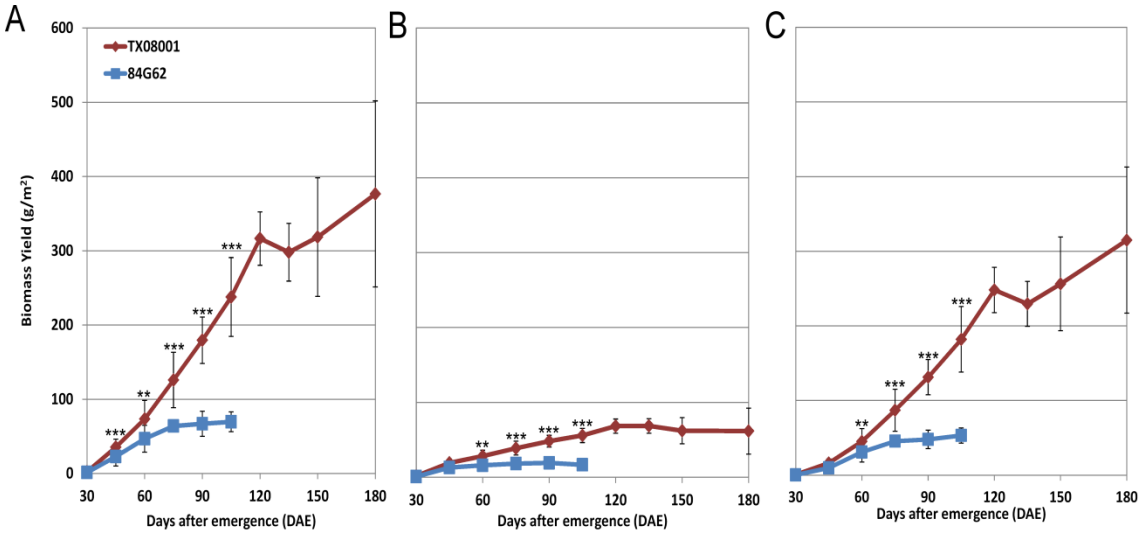

Figure 11: Time course of total, leaf and stem biomass accumulation of the grain sorghum 84G62 (blue lines) and energy sorghum TX08001 (maroon lines). (A) Accumulation of total shoot biomass (g DW plant^{-1}), (B) leaf biomass (g DW plant^{-1}) and (C) stem biomass (g DW plant^{-1}) during plant development. For each harvest, n=9 plants and error bars represent one standard deviation. (***, $p < 0.01$; **, $p < 0.05$)

GLA, LAI, and RUE

From emergence to 60 DAE, green leaf area (GLA) per plant increased rapidly in both grain and energy sorghum hybrids (Fig. 13). By July 1, energy

sorghum hybrids had a greater total GLA than the grain sorghum hybrids ($p < 0.01$) and this difference increased until 120 DAE when grain sorghum reached maturity. For energy sorghum hybrids, the maximum total GLA per plant, 6476 cm^2, occurred in in early August, while for grain sorghum the maximum plant GLA of 2735 cm^2 occurred in early July (Fig. 13). The GLA of energy sorghum hybrids declined from September to October due to senescence of the lower leaves (Fig. 13).

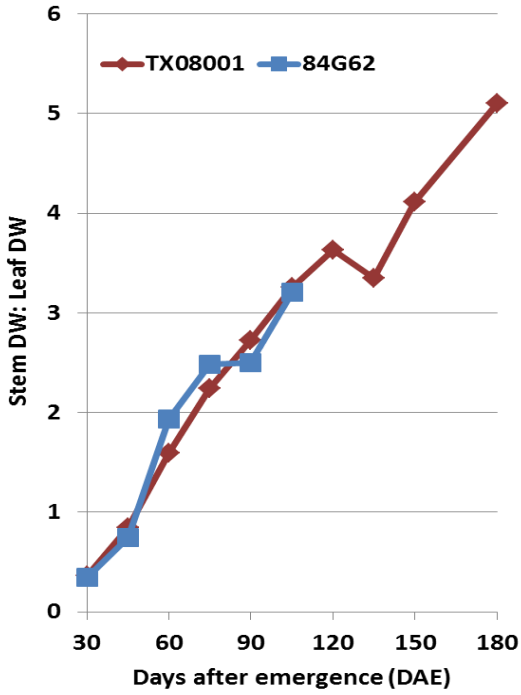

Figure 12: Stem to leaf DW ratio for the energy sorghum TX08001 (maroon line) and grain sorghum 84G62 (blue line). For each harvest, n=9 plants.

The larger canopy of the energy hybrids results in greater total light interception (S_i) compared to the grain hybrids (Table 3). In July (90 DAE) LAI was 3.4 for grain sorghum, and 5.6 for the bioenergy sorghum hybrid. Plants with LAI > 4 will intercept > 90% of the incident PAR (Bégué, 1993). Field ceptometer measurements showed that in early August when LAI was maximal, the grain sorghum BTx623 intercepted ~87% of the incident PAR whereas irrigated TX08001, which had the highest LAI, intercepted nearly all of the incident light (>98%).

Table 3: LAI (GLA $m^2 plant^{-1}$), TLI (MJ), and RUE (g m^{-2}) of TX08001 and 84G62 for each month of the growing season in 2008.

	LAI (GLA $plant^{-1}$)		S_i (MJ m^{-2})		RUE (g MJ^{-1})	
Harvest Date	84G62	TX08001	84G62	TX08001	84G62	TX08001
15 DAE	3E-2	3E-2	3.7	4.0	0.4	0.6
30 DAE	0.2	0.2	37.1	32.3	0.4	0.9
60 DAE	1.8	2.4	278.0	322.4	1.0	1.4
90 DAE	3.4	5.6	450.7	517.2	1.2	2.3
120 DAE	3.0	7.7	524.3	657.2	0.2	2.2
150 DAE	-	8.4	-	576.5	-	1.4
180 DAE	-	6.1	-	839.5	-	1.2

Biomass accumulation in energy sorghum increased rapidly starting in July when radiation intercepted by the canopy reached a maximum (Fig.14). Irrigated energy sorghum continued rapid biomass accumulation, at a rate of approximately 53 g d^{-1}, until September when rates of biomass accumulation declined. Non-irrigated energy hybrids accumulated biomass at the same rate

as irrigated hybrids until late August in 2008, and mid-August in 2009, indicating that lack of water was the major factor contributing to early cessation of biomass accumulation prior to mid-September.

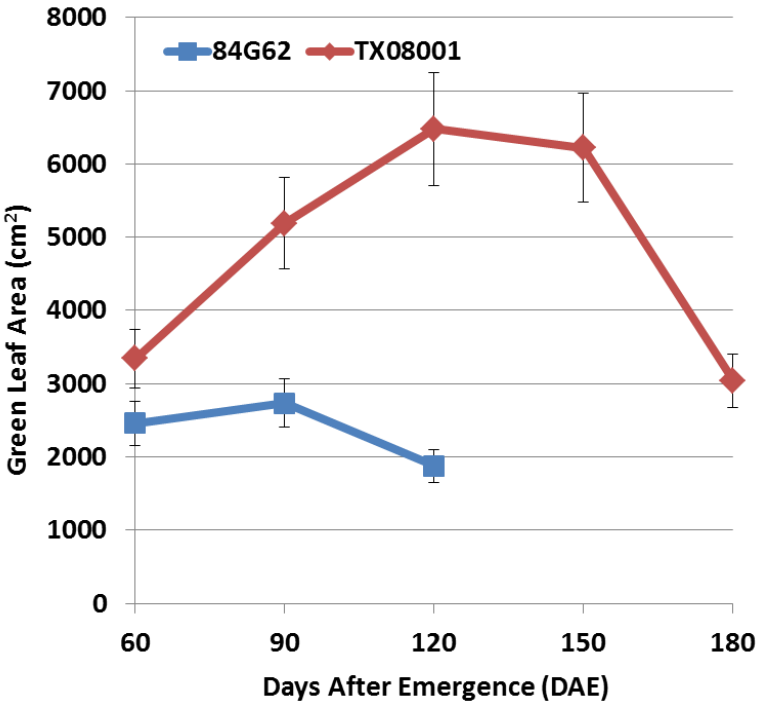

Figure 13: Total green leaf area (cm^2) for the energy sorghum TX08001 (maroon line) and grain sorghum 84G62 (blue line). For each harvest, n=9 plants.

Miscanthus and other C4 grasses with long duration of growth have some of the highest values of radiation use efficiency (ε_c). The ε_c of grain sorghum hybrids reached a maximum of 1.2 g MJ^{-1} in July, a value similar to previous reports for grain sorghum (1.24 g MJ^{-1}; Sinclair and Muchow, 1999). The ε_c of the energy sorghum hybrid increased to a maximum ~2.2-2.3 g MJ^{-1} in July and

August under irrigated conditions once plant canopies were fully established and stem growth and biomass accumulation was maximal.

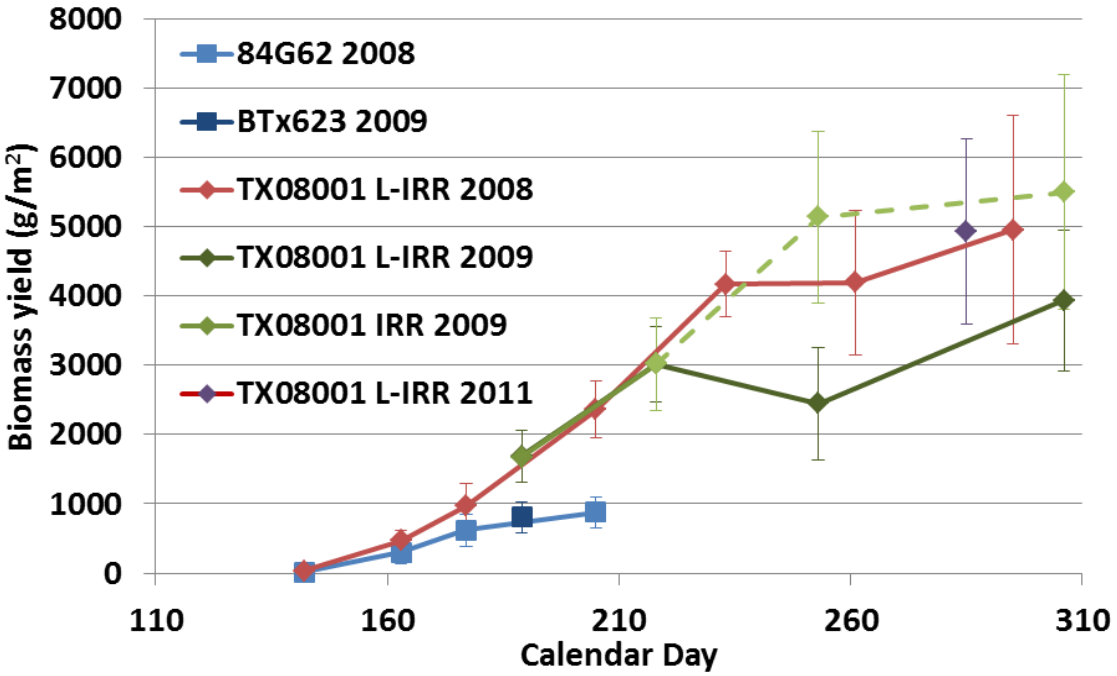

Figure 14: Time course of above ground biomass accumulation (g m^{-2}) of energy and grain sorghum in different years and irrigation treatments. (A) Biomass accumulation of 84G62, BTx623, and TX08001 in 2008 to 2011. Dashed line for TX08001 IRR 2009 represents uncertainty due to plant lodging that occurred in September. For all harvests, n=9, and error bars represent one standard deviation.

Discussion

The growth, development and genetic potential of a first generation energy sorghum hybrid, TX08001, were characterized and compared to grain sorghum. The energy sorghum hybrid had much longer growth duration and accumulated more than twice the biomass of grain sorghum ($p < 0.01$). The

basis of this difference in biomass yield was investigated to learn more about the processes affecting biomass accumulation and to identify ways to further improve energy sorghum. Grain and energy sorghum plants grew similarly for the first 30 DAE. By 30 DAE, both genotypes had produced and expanded approximately 10 leaves but showed limited stem elongation.

Floral initiation and duration of vegetative growth

Floral initiation in grain sorghum occurred at ~30 DAE; these plants reached anthesis by 70 DAE, and by 120 DAE grain was mature. In contrast, floral induction of the energy sorghum hybrid was delayed until mid-September when day lengths decreased below 12 h 20 m (Rooney and Aydin, 1999). Energy sorghum hybrids began to reach anthesis in mid-October followed by a grain development phase that was terminated by low temperature in November. The increased duration of the vegetative phase from 30 days in grain sorghum to 150 days in energy sorghum had a large impact on plant morphology and biomass accumulation. The longer duration vegetative growth phase of the energy sorghum hybrid allowed the production of ~45 leaves over the season compared to ~17-20 leaves by grain sorghum. Increased leaf number and leaf size resulted in higher green leaf area and greater LAI in energy sorghum (LAI = 5.6-7.6 in July/August) compared to grain sorghum (LAI = 3.4 in July). In July the canopy of grain sorghum intercepted ~87% of the incident light whereas energy sorghum hybrid canopies intercepted ~98%.

Plant size and biomass accumulation rates

The energy sorghum hybrid produced stems that were longer and of greater biomass than stems of grain sorghum. The longer vegetative growth phase of TX08001 resulted in production of ~25 more internodes than grain sorghum. In addition, the grain sorghum genotype used in this study contains recessive alleles at three dwarfing loci which reduce internode length and plant stature in order to reduce lodging and increase the harvest index of grain (Rooney, 2004). In contrast, the energy sorghum hybrid had longer internodes than grain sorghum indicating that TX08001 is probably recessive for only two dwarfing loci. The increased stem length of energy sorghum hybrids provided a strong sink for biomass. Over the season, stem:leaf biomass ratios increased steadily once stem elongation started, and by the end of the season, stem biomass was 83% of the energy hybrid's total biomass.

RUE and light interception efficiency

The ε_c of grain and energy sorghum hybrids grown with irrigation increased during development, reaching maximum values in July. The irrigated energy sorghum hybrid plants had an ε_c of 1.6-1.7 g MJ^{-1} averaged over the season. The ε_c of TX08001 reached a maximum of 2.3 g MJ^{-1} in July when the energy sorghum's canopy was fully developed, stem growth rate was maximal, and the highest biomass accumulation rates were recorded. Similar high ε_c values have been reported for sweet sorghum (1.9-2.5 g MJ^{-1}) (Mastrorilli et al., 1995), Miscanthus (2.3 g MJ^{-1}) (Beale and Long, 1995) and the C4 grass

Echinochioa polystachya (2.3 g MJ^{-1}) (Piedade et al, 1991). Maximum ε_c values for the grain sorghum were lower (1.2 g MJ^{-1}) but similar to values previously reported for grain sorghum (1.4-1.8 g MJ^{-1}) (Kinry et al., 1998; Muchow and Sinclair, 1994). Biomass accumulation and estimated ε_c of grain sorghum in July were probably affected by midge damage due to later than normal planting.

There was some uncertainty in light interception and biomass accumulation values for TX08001 in September and October under fully irrigated conditions because rain storms in mid-September caused plants in these plots to root lodge (Fig. 14, indicated by the dashed line, large standard deviation). Even so, the general agreement between measured biomass accumulation and predicted potential for biomass accumulation over the season indicates that the energy sorghum hybrid grown under fully irrigated conditions in well-managed field plots was functioning near its genetic potential at least until early September.

Response of TX08001 to limited irrigation

TX08001 grown with limited irrigation only until early July stopped accumulating biomass in early August (2009) or mid-August (2008) depending on when water in the soil profile became limiting (Fig. 14). During the more severe drought conditions in 2009, the shoot biomass of the energy sorghum declined in August and early September in parallel with lower leaf senescence. In all years, plants recovered and accumulated additional biomass starting in September when rainfall occurred. Large plots (5-1,000 m^2) of TX08001 grown

under completely dryland conditions had the lowest biomass yield and were most adversely affected by lack of water. Plant biomass yield ranged from 15-25 dT Ha^{-1} under these conditions, less than 50% of the theoretical genetic potential of TX08001 under optimal growing conditions.

The data on biomass accumulation of plants based on machine harvesting of large, non-irrigated field plots cannot be directly compared to data on individual plants obtained from fully irrigated plots that were thinned to a standard density and where portions of plots with poor stands were not assayed. However, it is reasonable to conclude that lack of water can be the primary factor limiting energy sorghum biomass yield when plants are grown under non-irrigated conditions in College Station, Texas.

Late flowering contributes to biomass accumulation

TX08001 is a first generation energy sorghum hybrid constructed specifically to have increased photoperiod sensitivity to lengthen the duration of its vegetative growth in order to increase biomass accumulation. In this sense the late-flowering energy sorghum shows a pattern of development similar to other high biomass C4 grass crops such as Miscanthus, sugarcane, and switchgrass. The yield of TX08001 varied from 15-59 dT Ha^{-1} depending on irrigation and water supply, plot management, and method of yield measurement (individual plant or machine plot harvest). A large variation in biomass yield (22 dT Ha^{-1} to 61 dT Ha^{-1}) was also reported for Miscanthus in multi-location tests and based on small and large plot assays (Heaton et al., 2004; Heaton et al.,

2008; Doleman and Long, 2009). While energy sorghum biomass yield was significantly higher than grain sorghum, much higher yields have been reported for the tropical C4 grass *Penisetum typhoides* (80 dT Ha^{-1}) (Begg, 1965) and the C4 Amazon floodplain grass *Echinochloa polystachya* (99 dT Ha^{-1}) (Piedade et al., 1991). Therefore, an additional goal of this study was to identify potential ways to modify energy sorghum to further increase biomass yield and yield potential.

Yield potential of bioenergy hybrid Sorghum bicolor

To investigate this question in a systematic manner, the well-established theoretical framework for analysis of genetic yield potential of crops (Monteith, 1977), further modified by Zhu et al. (2010) was utilized. This framework can be expressed by the equation:

$$Y = S_t \varepsilon_i \varepsilon_c \varepsilon_p$$

where S_t (GJ m^{-2}) is the total incident photosynthetically active solar radiation during a crop's growing season, ε_i is the crop canopy light interception efficiency, ε_c (g MJ^{-1}) is the conversion efficiency representing total canopy photosynthesis minus respiration, and ε_p is the partitioning efficiency or harvest index.

Duration of growth affects biomass accumulation

The development of energy sorghum hybrids with longer vegetative growth duration increased S_t as well as yield potential for bioenergy hybrid sorghum genotypes as compared to grain sorghum by extending the growing season of the bioenergy hybrid into early November compared to early August

for grain sorghum. Increased biomass yield potential of Miscanthus compared to maize was also due in part to increased S_t caused by earlier vegetative growth of Miscanthus in the spring (~4 weeks) as well as longer duration of growth into the fall (Dohleman and Long, 2009). Further increases in S_t and biomass yield might be achieved if energy sorghum could be planted earlier in the spring in regions where low temperatures would not inhibit sorghum growth. Sorghum is more sensitive to low temperature than Miscanthus and maize (Zhu et al., 2010), and this will limit early planting in more northern regions of the U.S.

Light interception efficiency and biomass generation

The efficiency of light interception, ε_i, of grain and energy sorghum is high once the canopy is fully established. The energy sorghum hybrid established leaf area more rapidly, to a greater extent, and maintained high LAI for longer duration. This contributed to increased biomass yield on a seasonal basis, as was found in comparisons of Miscanthus and maize (Dohleman and Long, 2009). Genetic improvement in the rate of canopy establishment in the spring has the potential to further increase biomass yield of energy sorghum. Grain sorghum intercepted ~87% of the incident light with an LAI of 3.4 in July whereas TX08001 intercepted 98% of the incident light with an LAI of 5.6-8.3 from July to September. It is possible that the very high LAI of TX08001 is not optimal, and hybrids with lower LAI but still high ε_i might increase yield by improving the efficiency of energy sorghum's leaf production. Lower LAI may

also reduce nitrogen requirements without penalizing water use efficiency in bioenergy hybrid sorghum.

C4 grasses that utilize NADP-malic enzyme such as maize, sorghum, Miscanthus and sugarcane have high ε_c and some of the highest rates of photosynthesis. Grain sorghum genotypes have high rates of CO_2 fixation at high light intensity under field conditions (44-55 µmol CO_2 m^{-2} s^{-1}) (Balota et al., 2008). The high midday rates of sorghum photosynthesis are similar to maize (~57.7 µmol CO_2 m^{-1} s^{-1}) but higher than Miscanthus (38 µmol CO_2 m^{-2} s^{-1}) (Dohleman and Long, 2009). The canopy of energy sorghum is much deeper than that of grain sorghum and this may alter ε_c by changing the portion of the canopy and photosynthetic apparatus operating under low efficiency at or near light saturation (Ort et al., 2011). Coupling direct measurement of CO_2 and water exchange rates, biomass accretion, and light penetration in the tall canopies of energy sorghum are needed to have a better understanding of whether the ε_c and water use efficiency of this crop can be further optimized by changing canopy architecture.

Harvest indices of biomass for biofuels

The harvest index, ε_p, of an energy crop is the harvestable portion of the shoot biomass per total biomass rather than the ratio of grain per shoot biomass typically used for grain crops. The yield equation gives harvestable yield (Y) in terms of MJ m^{-2}, and for the purposes of this discussion it is assumed that the energy content of energy sorghum biomass is approximately 18 MJ g^{-1}, although

this number will vary depending on biomass composition. The composition of sorghum biomass varies considerably but is typical of C4 grasses with large stems such as sugarcane or Miscanthus (data not shown). The relative amount of root biomass of Miscanthus (up to 38% of total biomass (Clifton-Brown and Lewandoski, 2000) and switchgrass (~50% of total biomass) (Frank et al., 2004) is significantly higher than sorghum (~20%, see below), as would be expected for perennial grasses that need to regrow each spring. Miscanthus shoot biomass accumulation peaks early in the fall and then decreases 25-40% as plants translocate carbon and nitrogen to roots for the next season (Jorgensen, 2011; Heaton et al., 2008). In contrast, grain sorghum and energy sorghum hybrids only require sufficient root biomass for water and nutrient extraction and to avoid lodging during annual growth. Preliminary data on shoot:root biomass ratios of TX08001 were obtained in 2011 by growing TX08001 under typical field conditions into October then harvesting nine entire plants including roots for analysis of shoot and root biomass. The shoot biomass of these plants was similar to TX08001 analyzed in 2008-2009 and the root biomass of these plants was ~20% of the total biomass accumulated (data not shown).

The high shoot to root biomass ratio and elevated shoot biomass ε_p for the energy sorghum hybrids were significantly higher than ε_p estimated for perennial C4 grasses such as Miscanthus and switchgrass. Root lodging often occurred in energy sorghum plots in the fall due to a combination of tall plant architecture, relatively small root systems, high plant biomass, and increased

rainfall that loosened soils in September. Selection for lodging resistance may increase root biomass and reduce ε_p, but may also improve radiation interception in the autumn.

C4 grass species and biomass generation

Sorghum, Miscanthus, sugarcane and switchgrass are all promising C4 grasses for biomass and bioenergy production (Rooney et al., 2007). When these species are grown in conditions that result in long vegetative growth phases and sufficient time to establish high LAI, they rapidly accumulate biomass due to high RUE and stems that are strong sinks for biomass accumulation. However, among these C4 grasses there are significant differences in seasonal shoot biomass accumulation potential due to: (1) location (latitude, radiation, evaporative demand, rainfall, and soil quality), and length of growing season, (2) differences in photosynthetic activity and ε_c, (3) differential adaptation to cool temperatures that affects photosynthesis and growth throughout the growing season, (i.e. Miscanthus is better adapted than maize and sorghum for early season growth (Jorgensen, 2011)), (4) variation in flowering time (i.e., switchgrass flowers earlier than Miscanthus or energy sorghum (McLaughlin and Kszos, 2005), (5) differences in shoot:root biomass accumulation and ε_p, and (6) differences in water availability, water use efficiency, and adaptation to drought. Miscanthus is better adapted to cooler regions in Europe and the central/upper mid-west, while sorghum is better adapted to warmer climates in Texas and the Gulf Coast region that are

subjected to hot, dry periods of variable length during the growing season. Switchgrass genotypes exist that are adapted to either region.

Response to limited irrigation

Energy sorghum grown under dryland or limited irrigation was very resilient, even after long periods of water deficit. The two years of this study had different weather patterns, but in both years, plants grown under dryland conditions or with limited irrigation experienced weeks of significant water deficit. The year 2009 was hot and dry. Under these conditions, the energy sorghum hybrid stopped growing in August, followed by an acceleration of normal lower leaf senescence and a period of quiescence that lasted for weeks in both years. When rain occurred in September, the plants initiated new leaf growth and accumulated additional biomass although at a lower rate consistent with reduced canopy and decreasing temperature and radiation.

Energy sorghum's response to severe water deficit is similar to the previously documented quiescence adaptation of selected sorghum genotypes (Mutava et al., 2011). This suggests that there may be an opportunity to improve the drought tolerance and water use efficiency of energy sorghum further because breeders have not previously selected sorghum genotypes/hybrids with long growth duration that need to tolerate long periods of water limitation. Moreover, sorghum's drought tolerance and wide adaptation will allow energy sorghum to be grown for biofuel production in regions that are of marginal use for grain and food production. Taken together, it can be concluded that energy

sorghum is a very useful genetic model for the development of dedicated C4 energy grasses. Moreover, TX08001, the energy sorghum hybrid used in this study, has been commercialized as ES5200 by Ceres Inc., making this first generation energy sorghum available for large scale biomass production.

Materials and methods

Experimental design, plant genotypes, and data analyses

The experimental design for collection of morphometric and biomass data from sorghum plants grown under field conditions was based on protocols developed for rice, wheat and other crops used previously to provide data for modeling crop growth and development (Thornton, et al., 1991). In 2008 and 2009, field studies were conducted at the Texas A&M University Field Station near College Station, Texas (30°37'40"N, 96°20'3"W, 100 m above sea level). At this location, soils are a Belk (very fine, mixed, active thermic Chromic Hapludert) (United States Department of Agriculture, 2011). The 2008 plots were planted on April 23 and the 2009 plots on April 14. In both years, fertilizer (100 kg N ha^{-1}) was applied prior to planting for both grain and energy sorghum. Row spacing was 76 cm and furrows were overplanted and thinned to a population density of 132,000 plants ha^{-1}. The inner rows were sampled within the field plots to mitigate potential edge effects. In 2008, the bioenergy sorghum hybrid TX08001 and the grain sorghum hybrid Pioneer 84G62 were planted in a plot three rows wide x 50 m. Both genotypes were irrigated twice between planting and early July and not irrigated thereafter (limited irrigation, L-IRR). In

2009, energy sorghum hybrids were planted in a block that was eight rows wide x 50 m long. TX08001 was planted and irrigated as needed throughout the season (IRR), and in plots that received irrigation as needed until July 7, and not irrigated thereafter (L-IRR). To mitigate border effects, all samples were harvested and measurements made on inner rows.

In both years, plants in the inner rows of the plots were divided into groups of three plants for analysis at different points during the season. For each data point, three adjacent plants from three random locations in the inner rows were harvested and individual plant characteristics were measured to obtain average trait values. The calculation of total dry weight (DW) per hectare was based on the number of theoretical plants in a hectare estimated from measured plant and row spacing within the plots. In 2009 and 2010 larger plots (5 m^2, 1,000 m^2) were planted and plants were grown under dryland conditions (no irrigation). The large plots were machine harvested in early September.

Analysis of physical and compositional traits

In 2008, plants were collected every two weeks starting two weeks after planting, and ending on October 21. At each harvest date, and for each plant, stem length was measured and the total plant, main stem, tiller stem, main stem leaves, tiller leaves and panicles (when present) were weighed for fresh weight (FW) and then bagged individually and dried for DW analysis following drying in an oven for three days at 70^{o}C. Green leaf area (GLA) was determined by passing each green leaf of the main stem through a planimetric leaf area meter

(LiCOR LI-3100C, Lincoln Nebraska, USA) prior to drying. Every second sampling included additional measurements, such as individual main stem internode FW and DW and individual main stem leaf green leaf area (GLA). In 2009, biomass samples were collected once a month starting in early July, with the final sampling on November 2.

Light interception measurements

Light interception was measured and estimated from the leaf area index (LAI). Light interception was measured using a linear PAR (400-700 nm) ceptometer (AccuPAR LP-80, Decagon Devices Inc., Pullman, Washington, USA). Light interception by the canopy was assessed at three randomly selected locations within the plots of each genotype in early August 2009. The measurements were taken on a clear, sunny day around midday. Light interception was recorded at multiple positions relative to the growing plants: above the canopy, within rows, and across rows at ground level. Measurements from several locations within the plots were used to obtain estimates of average light interception by each genotype.

Radiation use efficiency analysis

Radiation use efficiency (ε_c, g DW MJ^{-1}) was calculated as the ratio of biomass accretion (ΔB, g DW m^{-2}) to intercepted solar radiation (ΔS_i, MJ m^{-2}) by the crop for a given time interval:

$$\varepsilon_c = \frac{\Delta B}{\Delta S_i}$$

Intercepted solar radiation was estimated using the leaf area index (LAI), which was calculated as the ratio of green leaf area per plant to total ground area per plant. The solar radiation intercepted by the canopy on a daily basis (S_i) was calculated using the following equation:

$$S_i = S_t(1 - e^{-0.5*LAI}),$$

where S_t is the total incident solar radiation (MJ m^{-2} d^{-1}) retrieved from NASA via (http://earth-www.larc.nasa.gov/cgi-bin/cgiwrap/solar/agro.cgi?email=agroclim@larc.nasa.gov). The coefficient 0.5 is a daily extinction coefficient for solar radiation that is adequate for canopies with an approximately spherical leaf angle distribution. The term in parenthesis is the fractional radiation interception (ε_i).

Calculations of RUE were used to compare bioenergy hybrid and grain varieties of *S. bicolor* in terms of how much biomass each genotype was able to generate based on the light it actually intercepted.

Statistical calculations

All reported measurements are average values based on individual measurements of nine plants. Where applicable, standard deviations were calculated based on these measurements. Pairwise comparisons between 84G62 and TX08001 were carried out for multiple traits at each harvest. These comparisons were done using a Student's T-test with independent samples, calculated with Microsoft Excel. Statistically significant differences between

genotypes are noted in figures and tables where applicable (*, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$).

CHAPTER III

HIGH NUE OF BIOENERGY HYBRID *SORGHUM BICOLOR* IS VITAL TO ITS USEFULNESS AS A BIOFUEL CROP

Background and introduction

Nitrogen (N) is an important component of amino acids for proteins, chlorophyll, and a multitude of N-containing secondary metabolites. As a result, N supply has a significant impact on biomass and grain yields of plants (Perry, et al., 2010). Abundant N in the growing environment minimizes limitations on the rate of photosynthesis for a plant and, in turn, maximizes the rate of accumulation of biomass (Lawlor, 2002; Frink, et al., 1999). Therefore, N budget is an important consideration for any crop that is used to accumulate biomass for use in generation of biofuel (van Oosterom, et al., 2001). As it has been established in chapter one, generation of large quantities of high quality biomass is a critical component of increasing biofuel production (National Research Council (U.S.). Committee on Economic and Environmental Impacts of Increasing Biofuels Production., et al., 2011). This chapter will address the concerns pertaining to nitrogen fertilization of plants grown for biomass for generation of biofuels.

Nitrogen uptake by Sorghum bicolor

N is taken into a plant through the roots as nitrate or ammonium (Foyer, C.H., Hanma, Z., 2011). As a plant grows it continues to extract available N from the surrounding soil, growing its roots as needed to access areas of soil

with high concentrations of N (Takei, et al., 2002). Over time, it is possible that an individual plant will exhaust the locally available supply of N, which can impair photosynthesis and biomass generation (Caravetta, et al., 1990; Eilrich, et al., 1964). In order to maximize yield, most crops are fertilized with a high-N fertilizer prior to planting seeds (Byrt, et al., 2011). In some cases, an additional application of N is made during the growth cycle of the crop to mitigate potential biomass losses due to limited N supply (International Crops Research Institute for the Semi-Arid Tropics, 2011). This is not necessary in all cases, however, as there is significant variability in nitrogen demand between species and even between genotypes within a species (Brown, 1978).

Variation in nitrogen demand

Given that the role of N is most pronounced in the photosynthetic process, it follows that N will be a more important component of leaf tissue than stem or reproductive tissue in a plant (Taub and Lerdau, 2000). As such, in a given crop, genotypes which grow with a lower leaf-to-culm ratio may have higher NUE than genotypes which grow with a high leaf-to-culm ratio. For example, in the case of bioenergy hybrid sorghum (*Sorghum bicolor*), which was discussed in chapter two of this dissertation, the leaf area indices (LAI) of those plants are more than double what would be necessary to capture nearly all of the incident radiation (Bégué, 1993). The implication of this result is that higher NUE may be achieved by alteration of the leaf architecture, provided that such changes would not adversely affect other aspects of biomass generation (Sage

and Pearcy, 1987a; Taub and Lerdau, 2000). Longer growing plants also tend to require proportionally higher N availability, as N is continually extracted from the soil throughout the growing cycle to support growth and development of new tissues (Thomas, et al., 2002). In addition, the N demand of a crop will vary with the way in which the crop is grown, whether annually or perennially.

In perennially grown crops, vegetative plant tissue is typically harvested periodically. This harvesting method removes a portion of the total photosynthetic tissues of a plant, and therefore N, from the field, meaning that new tissue growth for the same plant will require additional uptake of N from the same soil in order to continue generating biomass (Dohleman, et al., 2009; Ranjith and Meinzer, 1997). As a result, perennial crops may suffer decreased rates of biomass accumulation following initial harvests.

Annually grown and harvested crops are wholly harvested and replanted from one season to the next, allowing for re-fertilization of the soil between plantings. While generation of biomass using this annual strategy may be more costly in terms of fertilization, increases in yield may be large enough to overcome this cost increase (Lewandowski and Schmidt, 2006).

When considering a crop's applicability for biomass generation, its N demand becomes highly relevant. Exogenous application of N is costly, both environmentally and economically (Frink, et al., 1999; Cui, et al., 2009). Environmental concerns include pollution and energy consumption associated with fertilizer production as well as contamination of groundwater by high-N

runoff (Hons, et al., 1986). Economically, N fertilization can account for as much as 15% of the total cost of biomass production (Tamang, et al., 2011). Given these concerns, optimal generation of biomass for biofuel hinges on maximizing biomass output from minimal N input.

Nitrogen use efficiency

For any crop, the relative accumulation of biomass can be quantified in terms of N use with the measure referred to as nitrogen use efficiency (NUE). Specifically, NUE is a measure of the total biomass generated by a plant per unit of N used by that plant (Hirel, et al., 2007). While there are multiple methods used to calculate NUE, which vary based on the intended application of the calculation, NUE is calculated in this case as the ratio of total biomass to total nitrogen mass for dry plant tissue (Schmitt and Edwards, 1981).

Many crops have been considered as potential energy crops for biomass generation. Among these, C4 grasses including Miscanthus (*Miscanthus x giganteus*), Sugarcane (*Saccharum officinarum*), and sorghum have been proposed as ideal energy crops due to their extremely high rates of biomass accumulation (Byrt, et al., 2011; Jorgensen, 2011). These crops have varied N demands and exhibit varied NUEs. The growth strategy (annual or perennial) has an effect on the NUE of each of these crops as previously described, but there are also genetic factors affecting NUE independent of growth strategy (Vermerris, 2011). Based on NUE, *S. bicolor* is an ideal energy crop. Specifically, bioenergy hybrid varieties of *S. bicolor* exhibit extremely high NUE

over a very lengthy period of vegetative growth, resulting in generation of large quantities of biomass from minimal N input (Rooney, et al., 2007).

N demand of long-growing bioenergy hybrid Sorghum bicolor

Concerns have been raised that *S. bicolor* biomass accumulation may suffer diminished yield over the course of a lengthy duration of vegetative growth due to limiting N availability (Vermerris, 2011). As a *S. bicolor* plant grows, new leaves are initiated and developed at the top of the plant as older leaves, closer to the soil and therefore shaded by upper leaves, senesce and fall away (Vanderlip, 1993). Each new leaf requires N for construction, as do growing portions of culm. This means that continued vegetative growth requires a continual supply of N in the soil for uptake (Richard-Molard, et al., 2008). It has previously been established that a deficit of N in the soil immediately surrounding a plant leads to a diminished capacity for biomass accumulation. There is a theory that the N in each leaf is trapped above the soil when the leaf senesces and falls away, meaning that the growing plant must extract additional N from the soil for each newly generated leaf (Perry, et al., 2010). This continued extraction of N would lead to an eventual exhaustion of available N in the soil. This concern needs to be addressed before bioenergy hybrid *S. bicolor* can be considered as a legitimate energy crop for biomass generation.

To gain further understanding of the NUE of bioenergy hybrid *S. bicolor*, nitrogen content was measured in stem, leaf, and leaf sheath tissue at a series of points throughout the growth cycle. NUE was considered in terms of

individual tissues as well as on a whole plant basis. Soil N was assessed prior to planting and following the final harvest for the field where these plants were grown, demonstrating the rate at which bioenergy hybrid *S. bicolor* plants extract

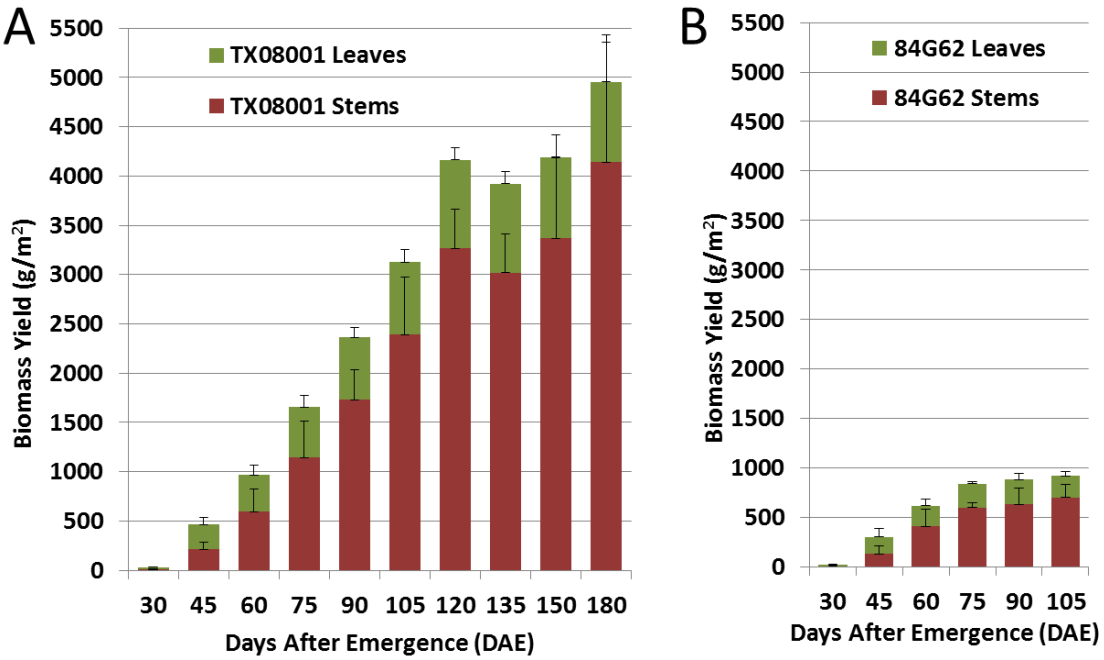

Figure 15: Stacked bar graphs of biomass yield. Green bars represent leaf biomass, maroon bars represent stem tissue. Y-axis is in g/m^2. These graphs are based on data from 2008 study. (A) biomass yield of TX08001 bioenergy hybrid *S. bicolor*. (B) biomass yield of 84G62 grain type *S. bicolor*. For each harvest, n=9, error bars show one standard deviation.

N from the soil compared to other crops. There are many crops that are considered ideal energy crops for use in generation of biomass for biofuels, many of which have been mentioned here, including miscanthus, sugarcane, and maize (Byrt, et al., 2011; Jorgensen, 2011; Vermerris, 2011). These data

demonstrate that bioenergy hybrid *S. bicolor* varieties are an ideal energy crop for two main reasons. First, bioenergy hybrid varieties of *S. bicolor* exhibit high NUE during vegetative growth. Second that high level of NUE is maintained over the course of an extremely long duration of vegetative growth.

Results

Biomass accumulation and partitioning

Biomass yield of TX08001 and 84G62 was measured in fifteen day increments throughout the growth cycle. 84G62 accumulates leaf biomass until 60 days after emergence (DAE) (Fig. 15a, green bars). At this point there is a shift from accumulation to maintenance of leaf biomass. Culm biomass, however, accumulates consistently until 90 DAE, which is coincident with the approximate timing of anthesis for this genotype (Fig. 15a, maroon bars and blue arrow). At maturity, culm tissue accounts for approximately 76% of the biomass in this genotype.

TX08001 generates over five times as much biomass as 84G62 during its growth cycle. The growth of vegetative tissue follows a similar profile in both genotypes. TX08001 plants accumulate leaf biomass until 120 DAE, after which leaf biomass is maintained rather than accumulated (Fig. 15b, green bars). Culm biomass accumulation in TX08001 is rapid and, with the exception of a brief lag from 120 to 150 DAE, continuous (Fig. 15b, maroon bars). This temporary alteration of culm biomass generation rate can be attributed to limiting water conditions during a particularly dry growing season.

Though the two genotypes differ substantially in eventual biomass yield, the culm-to-leaf ratios observed at the end of the growth cycles are somewhat similar. The composition of vegetative tissues for 84G62 at the end of the growth cycle is approximately 76% culm, 24% leaf. The harvest index of TX08001 is similar, as these plants yield approximately 84% culm and 16% leaf tissue.

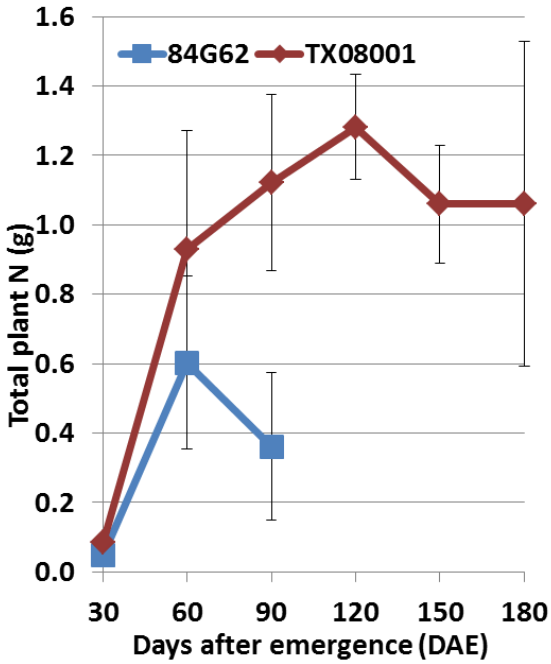

Figure 16: Total shoot nitrogen content in grams. Maroon line represents nitrogen content of shoots of TX08001 bioenergy hybrid *S. bicolor* plants. Blue line represents nitrogen content of shoots of 84G62 grain type *S. bicolor* plants. For each harvest, n=9, error bars show one standard deviation.

N accumulation and NUE

Total nitrogen content of accumulated biomass was measured in 84G62 and TX08001 at successive points throughout their growth cycles. Both genotypes take up nitrogen from the soil. The most rapid rate of nitrogen uptake in both genotypes occurred during the period from 30 to 60 DAE. During this time, 84G62 took up nitrogen at a rate of approximately 18 mg d^{-1}. TX08001 took up approximately 28 mg d^{-1} during the same portion of the growth cycle. Both genotypes take up nitrogen at an initially rapid rate, but neither the rate nor the whole shoot nitrogen content is maintained throughout the growth cycle.

In the case of 84G62, whole shoot nitrogen content decreases between 60 and 90 DAE (Fig. 16, blue line). This timeframe is consistent with floral induction, suggesting that nitrogen from the vegetative shoot tissues is being shifted into reproductive tissues once flowering is initiated. TX08001 takes up nitrogen in a different way (Fig. 16, maroon line). The shift from rapid uptake to moderate uptake corresponds to the previously established timing of leaf canopy closure in this genotype. At the point of canopy closure, nearly all incident light is intercepted by the existing leaf canopy and any new leaf area generated will likely shade established leaves. After 120 DAE, plant nitrogen content appears to decrease. This timing corresponds to the decreased rate of biomass generation discussed above.

Total biomass yield and total plant nitrogen content were used to generate nitrogen use efficiency (NUE) measurements for 84G62 and TX08001

at successive points throughout the growth cycle. Until 90 DAE, NUE for 84G62 and TX08001 are nearly identical (Fig. 17). After 90 DAE though, 84G62 reaches maturity and the NUE of this genotype no longer increases. TX08001 continues to grow vegetatively for much longer than 84G62 and is still growing at the point of the final harvest in this case (180 DAE). As a result, the NUE for this genotype continues to increase throughout the duration of its very long vegetative growth.

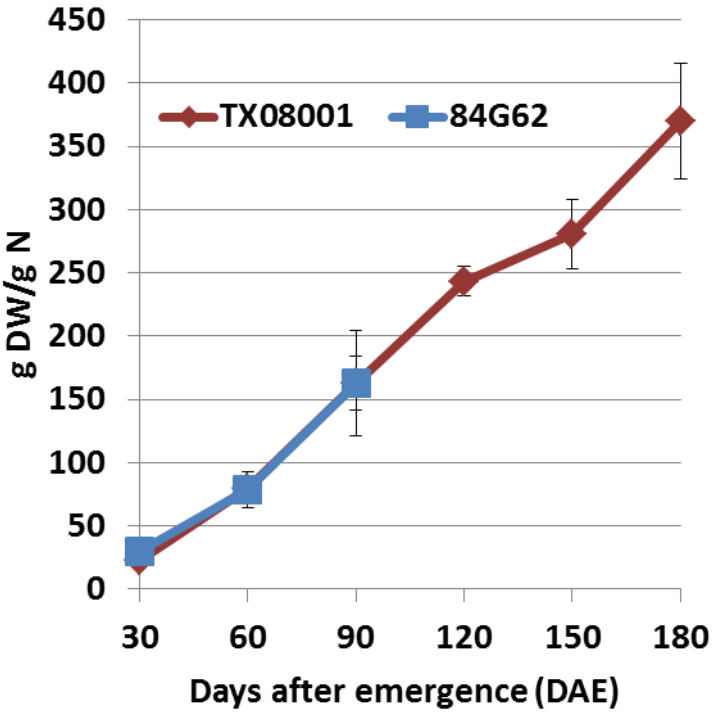

Figure 17: Nitrogen use efficiency of *S. bicolor* genotypes. NUE is calculated as total grams of DW per gram of nitrogen in shoot tissue. Maroon line represents NUE of TX08001 bioenergy hybrid *S. bicolor* plants. Blue line represents NUE of 84G62 grain type *S. bicolor* plants. For each harvest, n=9, error bars show one standard deviation.

Nitrogen content of leaves was determined in addition to whole plant nitrogen content. Like total plant nitrogen content, leaf nitrogen content increases rapidly at first in TX08001 leaves (Fig. 18A). The maximum rate of leaf nitrogen uptake in TX08001 was 18 mg d^{-1} and was only observed from 30 to 60 DAE. From 60 DAE onward, the nitrogen content of the leaves was either maintained or decreased; there was no additional accumulation of nitrogen.

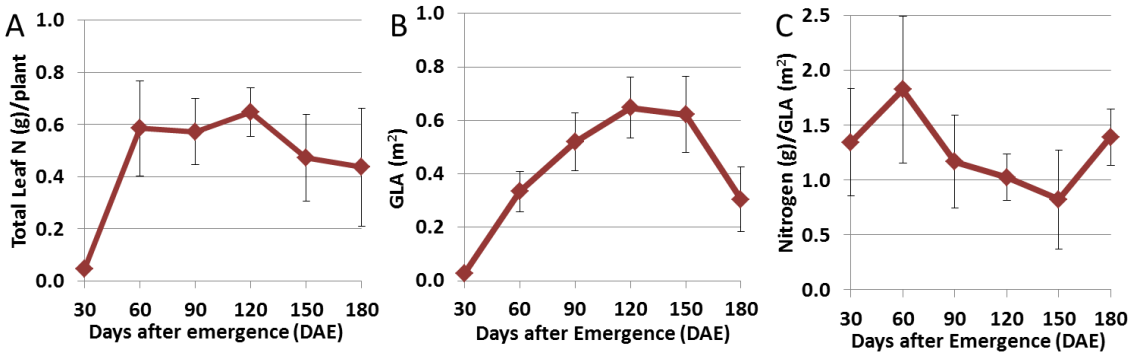

Figure 18: Leaf traits of TX08001 bioenergy hybrid *S. bicolor* plants. (A) Total nitrogen (grams) in leaf tissue of TX08001 plants across growing season. (B) Green leaf area (m^2) on a per-plant basis for TX08001 plants across growing season. (C) Nitrogen per GLA (g/m^2) for leaf tissue of TX08001 plants across growing season. For each harvest, n=9, error bars show one standard deviation.

Green leaf area (GLA) was also measured over the course of the growth cycle of TX08001 plants. Total GLA increases most rapidly early in the growth cycle and decreases in rate later in the growth cycle. This pattern is consistent with the concept of construction and eventual closure of the leaf canopy (Fig. 18B). Following putative canopy closure, there is some additional GLA accumulation seen, but the total plant GLA reaches an eventual plateau of 0.648

m^2 $plant^{-1}$ at 120 DAE. The decrease in total biomass accumulation rate seen late in the growth cycle of TX08001, earlier attributed to limited water conditions, is also observed in this case as a decrease in GLA. Late-season decreases in GLA can be attributed to environmental factors like wind and water deficit as well as to biological factors like plant lodging and senescence.

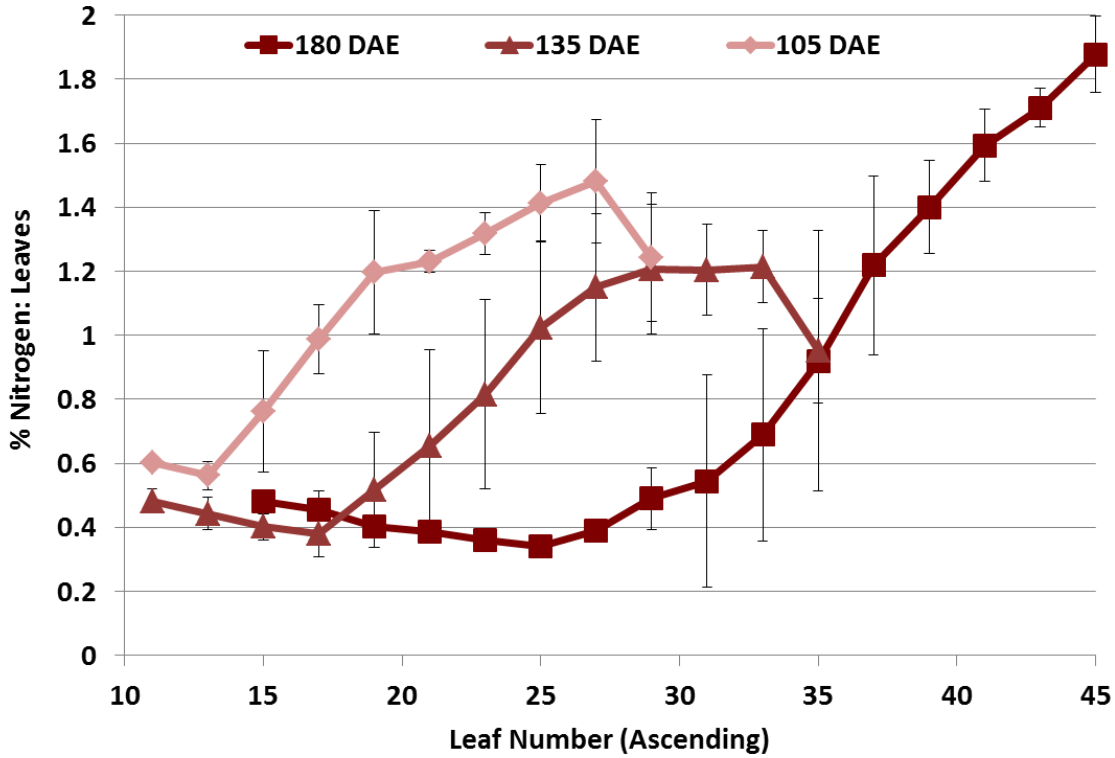

Figure 19: Change in nitrogen percentage in leaves over growing cycle. Leaves are numbered in ascending order, with leaf 1 at the ground. Every other leaf was assessed for this figure. Light pink line with diamond markers shows nitrogen percentage of individual leaves at 105 DAE. Medium pink like with triangle markers shows nitrogen percentage of individual leaves at 135 DAE. Maroon line with square markers shows nitrogen percentage of individual leaves at 180 DAE. For each harvest, n=9, error bars show one standard deviation.

Total N per unit leaf area decreased over the course of the growth cycle in TX08001 (Fig. 18C). This is consistent with the results obtained for both leaf N content and GLA. As GLA per plant increases and total leaf N per plant remains approximately the same, the total N per unit leaf area must decrease.

N recycling in leaf tissue

There is a pronounced profile of nitrogen percentage for each individual leaf up the stem in TX08001 plants. Following canopy closure, the top 16 leaves of a TX08001 plant have the highest nitrogen percentages, with the maximum being measured at the top of the plant and a lower percentage being measured at each successive leaf down the stem. Leaves below these top 16 are shaded by the upper leaves and have little incident radiation. The lower leaves do little photosynthesis and as such have lower nitrogen demands. The minimum nitrogen percentage is observed in the lowest leaves, with the smallest nitrogen percentage measured being 0.34% in leaf 25 at 180 DAE (Fig. 19). The nitrogen percentage of any individual leaf is a dynamic attribute which changes considerably over the lifetime of the leaf. Leaf 25, for example, has one of the highest nitrogen percentages of all of the leaves at 105 DAE (1.41%). Thirty days later, at 135 DAE, this same leaf has lost some nitrogen (1.02%). By 180 DAE, the leaf has lost nearly all of its nitrogen (0.34%).

As new leaves are made at the top of the plant, older leaves are shaded by the new leaves and nitrogen is recycled from the old leaves up to the new, growing leaves which have full exposure to sunlight (Thomas, et al., 2002). The

physical appearances of the leaves also demonstrate a nitrogen recycling phenomenon. The top 16 leaves are green and at least partially exposed to sunlight. Leaves below this top set are typically thin, brown, and shaded from exposure to sunlight (Fig. 20).

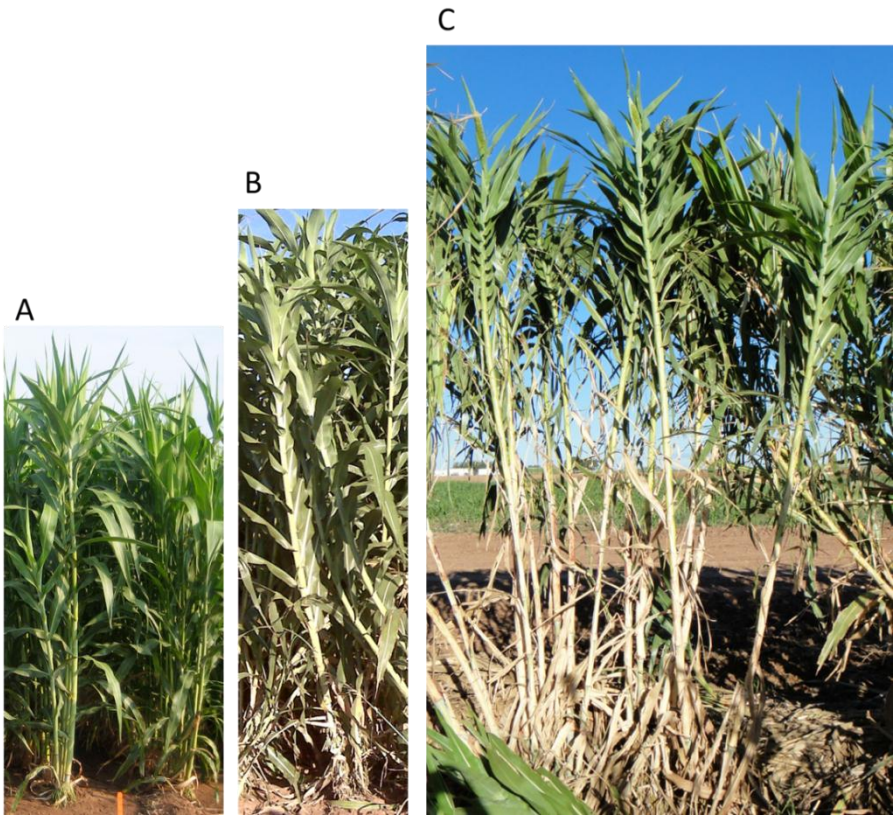

Figure 20: TX08001 bioenergy hybrid *S. bicolor* growing in the field in College Station, Texas. (A) photo taken on 23 July, 2008. Plants are ~ 90 DAE. (B) photo taken on 31 August, 2008. Plants are ~120 DAE. (C) photo taken on 13 October, 2011. Plants are ~180 DAE.

Nitrogen is not the only component that is differentially distributed among the leaves of a TX08001 plant. Assessment of composition of leaf tissue using NIR (near-infrared) spectroscopy reveals that many components, both structural and soluble, are present at different levels in each leaf. Cellulose, starch, and lignin are present in their highest levels in the lowest leaves of the plant, with lower levels in upper leaves (Fig. 21). In contrast, ash and protein are at their highest in the upper leaves of the plant with lower levels in the lowest leaves. Not all components vary, however. Glucan, arabinan, and sucrose are evenly distributed throughout the leaves.

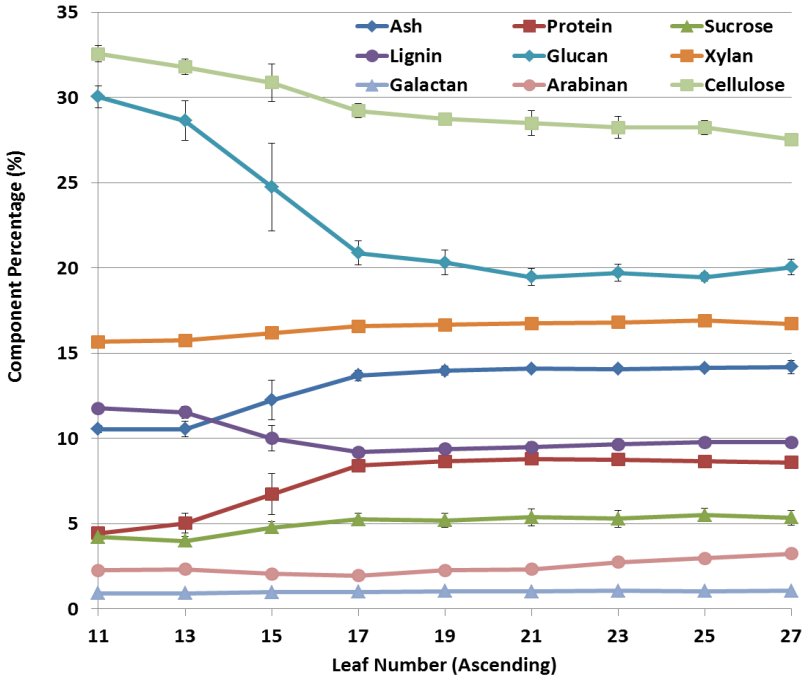

Figure 21: Composition of leaf tissue of TX08001 bioenergy hybrid *S. bicolor* plants at 120 DAE, as determined by NIR. Leaves are numbered in ascending order with leaf 1 at the ground level. Every other leaf is assessed. Each data point is the average of two NIR scans for each of nine tissue samples (N=18). Error bars show one standard deviation.

N recycling in stem tissue

The nitrogen present in leaf tissue accounts for less than half of the total nitrogen in the plants. The remainder of this nitrogen is located in stem tissue. As was the case for the leaves, total nitrogen in the stems increased most rapidly early in the growth cycle. This rate slowed at 120 DAE, when the stem nitrogen level reached a plateau of 0.656 g/plant (Fig. 22a). When reported as a percentage, however, the stem nitrogen content looks very different. Stem nitrogen percentage decreases continually throughout the growth cycle, with the minimum percentage being measured at 180 DAE (0.19%) (Fig. 22b). This value is consistent with reported nitrogen percentages for dead tissue (G. Hammer, personal communication).

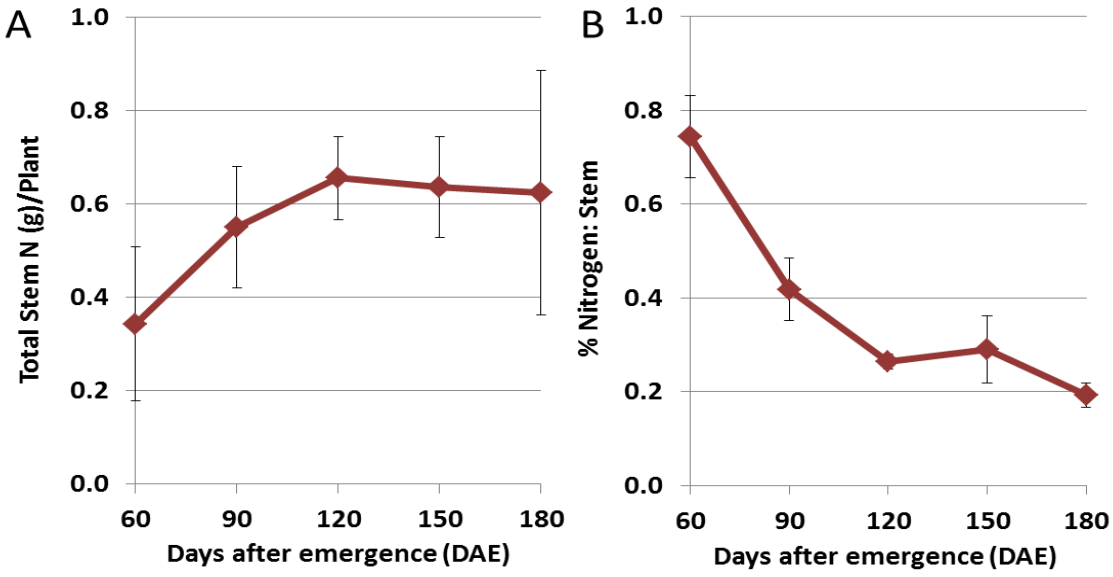

Figure 22: Stem nitrogen content of TX08001 bioenergy hybrid *S. bicolor* plants. (A) Total stem nitrogen per plant (g) over the course of growth cycle. (B) Total stem nitrogen percentage over the course of growth cycle. For each harvest, n=9, error bars show one standard deviation.

Nitrogen percentage in stem tissue can also be measured for individual stem sections to create a profile of nitrogen distribution across the stem. Such measurements show that the nitrogen percentages of individual stem sections resemble the nitrogen percentages of individual leaves when graphed (Fig. 23).

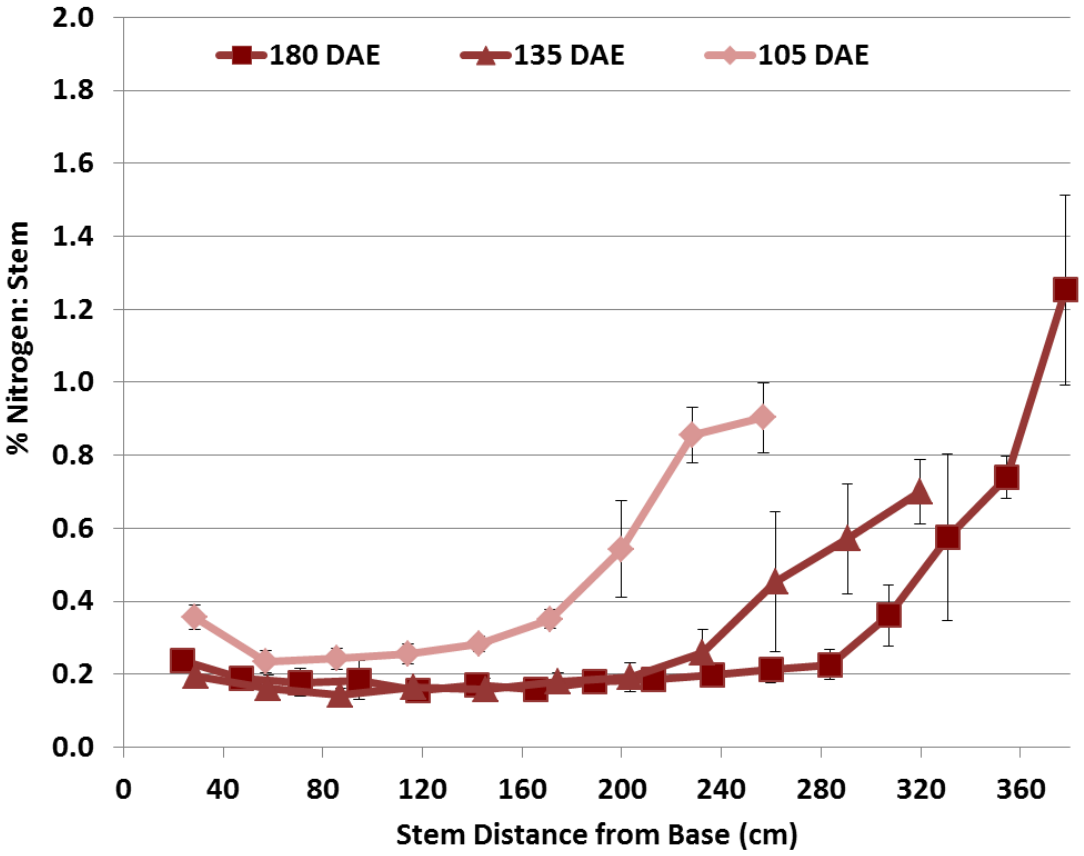

Figure 23: Change in nitrogen percentage in leaves over growing cycle. Stem sections are shown in ascending order, labeled by distance from the base of the plant. Light pink line with diamond markers shows nitrogen percentage of individual stem sections at 105 DAE. Medium pink like with triangle markers shows nitrogen percentage of individual stem sections at 135 DAE. Maroon line with square markers shows nitrogen percentage of individual stem sections at 180 DAE. For each harvest, n=9, error bars show one standard deviation.

For example, the stem section corresponding to 240 cm above the plant base has 0.88% nitrogen at 105 DAE. By 135 DAE, this level drops to 0.35%, and the percentage has fallen to 0.2% by 180 DAE. As was the case for the leaf tissue, this phenomenon suggests that the plant is recycling nitrogen from its lowest tissues up to the growing top of the plant to continue vegetative growth without additional nitrogen uptake from the soil.

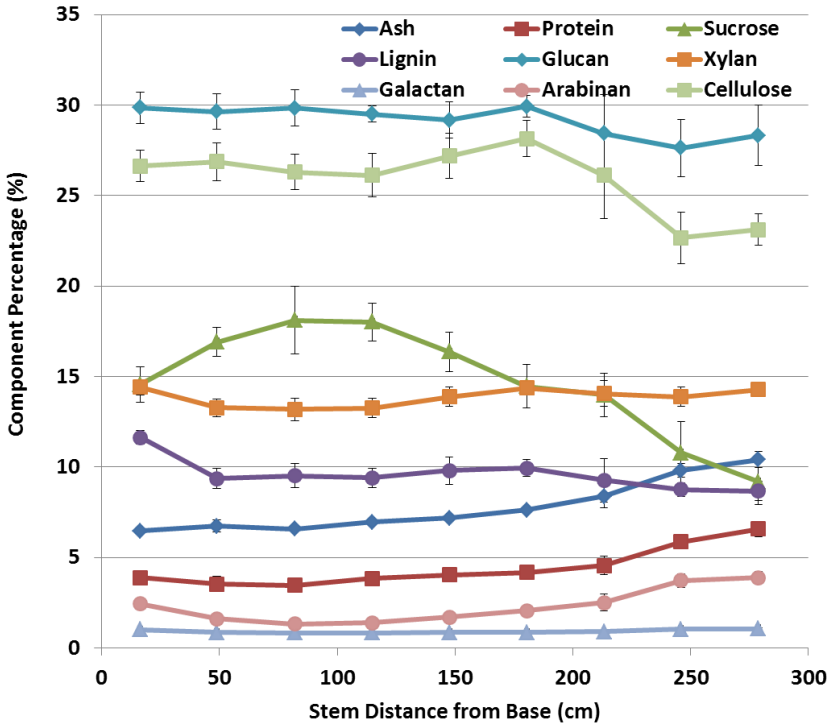

Figure 24: Composition of stem tissue of TX08001 bioenergy hybrid *S. bicolor* plants at 120 DAE, as determined by NIR. Stem sections are shown in ascending order from base of plant. Each data point is the average of two NIR scans for each of nine tissue samples (N=18). Error bars show one standard deviation.

Analysis of stem tissue using NIR spectroscopy is also useful for assessment of whole stem composition (Fig. 24). Sucrose, cellulose, and starch are at their highest levels near the base of the plant, with lowest levels recorded near the top of the plant. In contrast, ash, protein, and arabinan are lowest at the base of the plant and increase in concentration near the top of the plant.

Discussion

NUE is an important consideration for generation of biomass for biofuels (Ragauskas, et al., 2006; Simmons, et al., 2008). While maximum biomass output is of great importance, the cost of generating that biomass is of equal importance. Total nitrogen uptake by a plant, as well as partitioning of that nitrogen within a plant, are critical factors to take into account when considering the fitness of a crop for generating biomass for biofuels (Fazio and Monti, 2011). As these data demonstrate, bioenergy hybrid genotypes of *S. bicolor* can grow vegetatively for extremely long durations, and require no additional fertilization to support that growth beyond typical fertilization levels for other, short-duration *S. bicolor* genotypes.

NUE increases throughout growth

NUE is not a static attribute of a plant; rather it is a variable phenomenon. The level of NUE measured in a juvenile does not appear to correspond to the NUE of that same plant at maturity. Figure 17 demonstrates that, in the case of TX08001, NUE actually increases in a nearly linear mode throughout the growth cycle. The same mode of increase of NUE is observed for 84G62 until the point

of anthesis, but this genotype has a much shorter duration of vegetative growth than TX08001. As a consequence, TX08001 has considerably more time to accumulate biomass and increase its NUE.

Early in the growth cycle, both genotypes have nearly identical levels of NUE. This may be explained by the fact that this is the time when both genotypes are constructing and filling in their canopies. This is a time of high-intensity leaf growth and stem development. Growth of leaves, including synthesis of chlorophyll for photosynthesis, requires very large amounts of N (Smith and Frederiksen, 2000). Stem development also creates a sink for N, as elongating and dividing cells require protein for construction of cell walls (Murray, et al., 2008b; Murray, et al., 2008a). This is a portion of the growth cycle when the plant is at its least efficient in terms of nitrogen use.

As the plants get older and the canopy is filled in, each new leaf shades old leaves, and shaded leaves senesce and fall away. Therefore, once the canopy has been filled in, a plant will maintain an approximately constant number of green leaves on its stem, and the level of net leaf growth will be significantly decreased (Gerik, et al., 2003; Gerik and Neely, 1987). This helps to explain further the rapid uptake of nitrogen by *S. plants* of both genotypes early in the growth cycle as well as the plateau in total plant nitrogen content observed after 90 DAE. However, this pattern of leaf growth does not explain how the TX08001 plants are able to continue vegetative growth for extremely long durations on fields where the N fertilization level is the same as that

required for growth of 84G62, a grain type with a significantly shorter growth cycle than TX08001.

Increased NUE through N remobilization

Soil analyses performed before and after growth of these plants indicate that TX08001 did not use significantly more N from the soil than 84G62 (data not shown). There was still residual N present in the soil immediately adjacent to the TX08001 plants at the end of the growth cycle. This result suggests that TX08001 plants were not limited by N availability, but rather the observed level of N uptake was sufficient to sustain growth. This result invalidates the hypothesis that long-growing, high biomass producing genotypes of *S. bicolor* like TX08001 will require additional N fertilization compared to typical fertilization levels in order to achieve high yields of biomass (Byrt, et al., 2011; Lewandowski and Schmidt, 2006). In addition, the soil analysis results preclude the putative need for re-fertilization mid-growth cycle for sustained vegetative growth by TX08001 plants.

The question of how these plants were able to support such immense vegetative growth using so little N is intriguing. The answer is likely complex and includes the process of remobilization of N within live leaf tissues prior to senescence. In figure 18a, the N level in the leaves of TX08001 is observed to stabilize at approximately 0.6% after 60 days of growth. Then, figure 19 illustrates the trend of nitrogen content on a per-leaf basis and provides a concrete clue to the process being used to maximize NUE in these plants. As

an individual leaf ages, its N percentage decreases from an initially high level, in a linear fashion, to a level that is consistent with levels of N in dead plant tissues. The change in N percentage over time indicates that each leaf is losing its N while still live tissue. These leaves are not merely dead, brown leaves. Rather, green leaves that are still fully attached to the plant are being removed of their N content prior to senescence. The plant is recycling its own nutrients from leaves that are no longer necessary for photosynthesis, due to shading, and using those nutrients to build new leaves at the growing top of the plant. While the hypothesis of N remobilization within the plant to sustain new growth has been proposed in the case of filling grain in *S. bicolor* plants at maturity, this is evidence that the same strategy is being used by vegetative TX08001 plants to support continued growth rather than maturity.

The specific distribution of N to each individual leaf across the TX08001 plant also provides evidence as to the nature of the N recycling that may be occurring in the leaves. Beginning with the highest level at the very top of the plant, N percentage then decreases incrementally with each leaf down the stem. Light availability in the canopy is also at the highest level at the very top of the plant and decreases incrementally with each leaf down the stem due to shading by the leaf or leaves above it, until such depth that no more light can penetrate the canopy (Bégué, 1993). The concurrence of light availability and relative N percentage indicates that each leaf is able to individually modulate its photosynthetic capacity, and thus N demand, based on the light available to it.

Additional research will be necessary to confirm this hypothesis and such experiments should include artificial shading of individual leaves and measures of carbon dioxide assimilation by individual leaves in shaded versus sun-exposed conditions.

Table 4: NUE levels for candidate biomass crops. Ranges reflect within-species variation as reported in the literature.

Species	NUE (g DW g^{-1} N^{-1})	Citation
Saccharum Oficinarum	499	Ranjith and Meinzer, 1997
Energy Sorghum bicolor	370	This dissertation
Miscanthus x giganteus	125 – 333	Jorgensen, 2011
Zea mays	100 – 166	Cui et al., 2009
Grain Sorghum bicolor	163	This dissertation
Panicum virgatum	36 - 119	Bowman, 1991
Triticum aestivum	106	Schmitt and Edwards, 1981

NUE and biomass generation of grass species

There is considerable variation in NUE observed both within *S. bicolor* as a species, as well as between *S. bicolor* and other grass species. Many such species have been previously considered in terms of fitness for use in generation of biomass for biofuel, including sugarcane (*Saccharum officinarum*), miscanthus, switchgrass (*Panicum virgatum*) and maize (*Zea mays*) (Ranjith and Meinzer, 1997; Frank, et al., 2004; Makino, et al., 2003). While absolute biomass generation capacity is of importance, NUE for each of these candidate biofuel crops is highly important as well.

Table 4 shows reported NUE values for a range of species that have been suggested as potential crops for use in generation of biomass for biofuels (Bowman, 1991; Ranjith and Meinzer, 1997; Frank, et al., 2004; Makino, et al., 2003). TX08001 sorghum ranks very highly among these grass species. In fact, sugarcane is the only crop listed that exceeds TX08001 in terms of NUE. No other grass crop in this group has been reported to achieve even half of the level of NUE reported for TX08001.

While sugarcane exhibits the highest NUE of these crops, this value is somewhat misleading. Sugarcane is grown in tropical conditions with extremely long growth seasons (Ranjith and Meinzer, 1997). Given that the NUE for TX08001 is shown to increase steadily over the course of the growth cycle, it is reasonable to propose that the NUE would continue to increase additionally if the plants were grown in a place where the growth conditions were maintained at a favorable level for a longer period than is achievable in the fields in College Station, TX. Extrapolation of the line representing NUE in Figure 17 shows that TX08001 could hypothetically reach an NUE level comparable to sugarcane in 243 days of growth, or 80 additional days of growth over what was achieved in College Station, TX. Additional research to confirm this hypothesis would be highly valuable. Future efforts should include planting TX08001 in a tropical setting, or planting in College Station, TX at an earlier time of year to maximize potential growing days. Only then will it be possible to discover whether

TX08001 is able to grow vegetatively for long enough to meet or exceed the NUE of sugarcane.

There are many factors to consider when evaluating the fitness of a crop for use in generation of biomass for biofuels. As was demonstrated in the introduction, NUE is a very important attribute of any crop because of its role in the cost of production of biomass as well as the environmental cost of excessive fertilization. If ethanol from lignocellulosic sources of biomass is to become a reality of the energy landscape, it will be necessary to examine NUE further. Determination of the gene(s) that control NUE in sorghum will be instrumental to increasing NUE and therefore cost effectiveness of generation of biomass for biofuels from this crop.

Materials and methods

Genotypes used in the study

This study is a comparative study of the growth of grain type and bioenergy hybrid type *Sorghum bicolor*. The grain type, 84G62, is a hybrid line sold by Pioneer and well-known for its high yield of grain even under dryland growing conditions (White, 2006). The bioenergy hybrid type, TX08001, is a tri-hybrid produced by crossing ATx2752 x BTx623 x R.07007.

Field planting conditions

Prior to planting, the field was treated with nitrogen at a rate of 100 kg/ha. No additional nitrogen was applied after planting. Following planting, these plots

were irrigated twice prior to the 15th of July, and rain-fed for the remainder of their growth.

Plot design

Plants used in this study were grown in fields at the Texas A&M University Field Station outside College Station, TX. The soil in this field consists of Belk Clay (United States Department of Agriculture, 2011). Planting occurred on the 23rd of April, 2008. Sufficient seed was sown to yield an excess of plants, which were thinned to 10 cm spacing following seedling emergence. Each row of plants was 50 m in length, with consecutive rows planted 76 cm apart. The 50 m rows were subdivided into sections of approximately 5 meters, which were referred to as ranges. For each genotype planted in this study, the field plot consisted of three rows.

Sampling scheme

Samples were collected from the field every 15 days. All plants were sampled from the center row of each three-row plot, such that edge effects should be negligible. For each sample, nine plants were collected for each genotype. These plants were chosen by first selecting three random ranges within the row, then selecting three consecutive plants at each location. Random ranges were pre-selected using a random number generator. Each plant was cut at soil level and returned to the laboratory for measurement.

Plant measurements and processing

All plants were first weighed to determine fresh weight. Then, stem height from the ground to the collar of the top fully expanded leaf was recorded. Leaf blades were removed, and green leaf area was measured using a planimetric leaf area meter (Licor LI-3100C, Lincoln, Nebraska, USA). Following area measurement, leaf fresh weight was recorded, and leaves were dried in a drying oven for three days at 60° C with blowers constantly circulating the air, at which point leaf dry weights were recorded. After leaf blades were removed from the stem, stem fresh weight was measured and the stem was then separated into sections.

For each stem section, length and fresh weight were recorded. Then, the stem sections were dried using the same conditions as those used to dry the leaves and stem dry weights were recorded. In some instances, stem sections were too large to dry in three days. Such stem sections were left in the drying oven for up to five days until all residual moisture had been removed.

For any plant which had produced tiller stems, those stems were collected along with the main stem for that plant. Each tiller stem was processed and measured in the same way as the main stem such that all tiller and main stem data could be combined to yield total plant measurements for fresh weight, dry weight, and green leaf area parameters.

Composition analysis

Dried plant tissue was ground before being analyzed for composition. Grinding of dried leaf tissue was done using a Cyclone Sample Mill (Udy Corporation, Fort Collins, Colorado, USA). Grinding of stem tissue was done in a two part process, first using a Total Blender Fourside (Blendtec, Orem, Utah, USA), and then using a Krups F203 Fast-Touch Coffee Grinder (Krups, Shelton, Connecticut, USA). For all samples, grinding was carried out until the ground sample could pass through a 2 mm mesh.

Nitrogen content analysis was carried out using a Leco FP-528 Nitrogen/Protein Determinator (Leco Corporation, St. Joseph, Missouri, USA). As this is a combustion-based method requiring destruction of the tissue being measured, small samples (0.15 g) were used for this measurement.

Total composition was analyzed by near infrared spectroscopy (NIR). These measurements were carried out using a FOSS Rapid Content Analyzer (FOSS, Eden Prairie, Minnesota, USA). This is a non-destructive analysis method, and each sample was analyzed twice to minimize error due to variation in sample particle size and to account for environmental factors which could influence the readings. Each reported measurement is the average of two measurements for all nine plants in a given sampling.

CHAPTER IV

OPTIMIZING YIELD AND COMPOSITION:

A QUANTITATIVE GENETICS APPROACH

Background and introduction

When considering a crop as a potential source of biomass for biofuels, there are many attributes that may contribute to that crop's fitness for use. The potential biomass yield of a crop and the composition of that biomass are both of great importance (Rooney, et al., 2007). This chapter will address the genetic basis of variation for a large number of traits relating to biomass generation, plant maturity, and biomass composition (Semagn, et al., 2010). Through quantitative trait locus (QTL) mapping and subsequent analysis, genomic loci have been identified that contribute to measured variation in many physical traits.

QTL mapping: process and benefits

QTL mapping is a process that utilizes phenotype and genotype data from the progeny of a genetic mating to facilitate identification of the genetic loci that control variation in physical trait expressed in the population (Collard, et al., 2005; Hackett, 2002). When a QTL map is generated, it answers the question of whether there is a statistically significant correlation between genotype and phenotype variation in the population analyzed at every site across the genome (Kearsey, 1998). The strength of the correlation between the two is reported in terms of log of odds (LOD) (Rice, et al., 2001). This multi-step process begins

with identification of parental lines with significant phenotypic variation. For these experiments, the two parents selected were SC170 and Maldandi 35-1 (M35-1). SC170 is a short-statured, thick-stemmed, caudatum variety of *S. bicolor* which comes from river deltas in Ethiopia where water supplies are typically plentiful (Hart, et al., 2001). M35-1 is a very tall, thin-stemmed genotype which is grown for grain in India (Starks and Doggett, 1970). These two varieties, which vary in terms of stem length and diameter, leaf area, and many other traits, are ideal for use in construction of a population for QTL mapping.

Once two parents have been selected, a cross is made, F1 status is verified, and the offspring are advanced using self-pollination for multiple generations (Xu, et al., 2000). In this case, plants from the F5 generation were used. This self-pollination over repeated generations decreases the level of heterozygosity in the offspring in a stepwise fashion each generation (Smith and Frederiksen, 2000). Given that the phenotypic effects of both heterozygous dominant and homozygous dominant alleles are identical except in cases of dominance variance, decreasing the level of heterozygosity can increase the power to identify QTL (Dodds, et al., 2004).

The next step is to grow the population and record measurements for all phenotypes of interest (Guan, et al., 2011). For this experiment, those phenotypes included measures of plant size and weight, plant maturity, and stem biomass composition. Once phenotypes have been measured, tissue

needs to be sampled for DNA extraction as well. This isolated DNA must then be genotyped across the entire genome in a specific way. The actual sequence of the DNA does not matter; rather it is important to identify the parental source of the DNA at each location across the genome (Mace and Jordan, 2011). This allows for identification of the parental allele that is contributing to phenotypic variation in each case.

When the process is completed, the result is a QTL map, which identifies discrete loci within the genome that are contributing to measured variation in a trait (Agrama, et al., 1999). For each trait mapped, there will be a unique QTL map generated. These maps, and the individual QTL locations, can be compared between traits to reveal overlapping, or co-located, QTL (Mace and Jordan, 2011). While QTL mapping of each individual trait can be informative, using the co-location of QTL from multiple traits can increase the power of this method. When the same QTL is identified for multiple individual traits, this locus is often a putative macro regulator that is identified for one or more macro traits as well (Ming, et al., 2002). The plants in this population vary on a multitude of levels, but by using multi-trait QTL mapping, it is possible to decipher complicated phenotypic effects in terms of their component genetic effects.

For most macro traits, like biomass yield and stem length, QTL that have been identified will also be seen in the QTL maps of component traits that relate to the macro trait in question. These putative macro regulators can be of use once identified, even if the actual gene underlying the QTL has not been

identified. The parental allele contributing to increased phenotype values for each trait can be of significant importance as well, as this piece of information can be of use in creating a hypothesis as to the function of the gene. For each QTL, one parent is identified whose allele contributes to an increased phenotypic value. The other parental allele necessarily contributes to a decrease in phenotypic value for the same trait at the same locus.

Putative macro regulators of Sorghum bicolor

Some putative macro regulators have previously been identified in *S. bicolor*. With respect to the results contained in this chapter, there are a number of genomic loci that influence stem length that are of greatest interest. These are called the dwarfing loci, and there are four of them (Dw1 through Dw4) (Quinby and Karper, 1954). Dw2 is linked to the maturity locus Ma1 on chromosome six (Klein, et al., 2008). Dw4 has not been mapped to any specific genetic location at this time.

Dw3 is located on chromosome seven and the gene underlying this macro regulator has been identified (Multani, et al., 2003). The gene at this locus that modifies stem length encodes an ABC type-B auxin transporter. An important phytohormone, auxin is produced by plants in the shoot apical meristem and then transported throughout the plant as a signal to initiate or continue growth of plant tissue. Genotypes containing Dw3 express a functional auxin transporter and have internodes that are longer than the internodes of plants that are dw3. The presence of a duplication of 882 bp in the fifth exon of

the gene encoded by Dw3 leads to a non-functional gene, a recessive allele of this gene (George-Jaeggli, et al., 2011). In the recessive case, plants exhibit altered polar transport of auxin and the result is a shortened stem, often with a thicker diameter. The parents of the population used in this study, SC170 and M35-1, differ at this locus. M35-1, which is a tall variety of *S. bicolor* with slender stems, carries a dominant copy of the Dw3 gene. SC170, on the other hand, has short and thick stems and carries the recessive *dw3* allele at this locus.

S. bicolor dwarfing locus 1 (Dw1) is somewhat less thoroughly-understood (Quinby and Karper, 1954). As was the case for the Dw3 locus, the parents of this population differ in their Dw1 genotypes. SC170 harbors a recessive copy of this gene, evident by the short stems of SC170 plants. M35-1, with its tall stems, carries the dominant form of the Dw1 gene. Though the specific gene represented by the Dw1 locus has yet to be identified, the dominant and recessive alleles can be assigned based on the convention used for naming the other dwarfing loci in this system.

Results

The following sections will describe the locations and contributions to phenotypic variation of all of the QTL for each trait measured in this study. For each trait, a QTL map has been generated for the entire genome. In addition, a table has been created for each trait that includes genetic position, physical position, additive variation, and percent variance explained for each QTL

identified. Any of these maps or tables not found within the text are listed in the appendices.

Macro traits

Macro traits were measured to identify genomic regions that may be contributing to control of multiple specific traits. Such traits are typically complex and based on many component traits (Brown, et al., 2008). Often, loci found to be responsible for a portion of the variation of a macro trait are also found to be responsible for much of the variation measured in the various component traits that contribute to variation in these macro traits (Semagn, et al., 2010).

Table 5: Phenotype values and QTL identified for macro traits in SC170xM35-1 F5 generation.

Trait Name		Range		# QTL	% Variance Explained
	Units	Min	Max		
Days to Anthesis	days	69	85	5	58%
# Nodes	.	9	15	3	45%
LAR	days leaf^{-1}	5	8	3	31%
Total DW	g	43	200	4	45%
Stem Length	cm	48	218	3	93%
GLA	cm^2	2125	5272	3	40%

Macro traits include total biomass yield, days to anthesis, total stem length, total green leaf area (GLA), and other traits listed in table 5 (Table 5). For each of these traits, the QTL identified can explain a portion of the total variation observed in the population. In addition, some macro traits are

somewhat nested within other macro traits. For example, both number of nodes and leaf appearance rate (LAR) are nested within the number of days to anthesis. As such, some QTL will overlap between these traits even though each is independently considered a macro trait.

Total dry biomass yield

Four QTL were identified for total dry weight (DW) of each plant (Fig. 25). For three of these QTL, presence of the allele from M35-1 increased DW, while for the fourth QTL, the presence of the allele from SC170 increased DW. The first QTL identified, on chromosome one, has a peak at 54.6 Mbp and contributes 10.82 g DW to any plant that has the M35-1 allele at this locus. Two QTL were identified on chromosome nine.

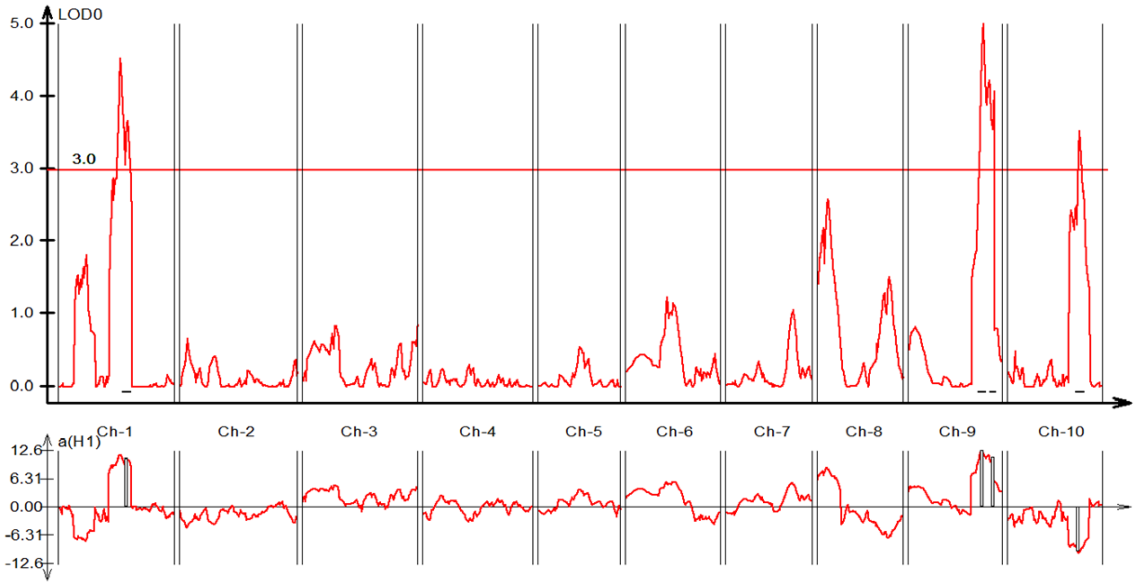

Figure 25: QTL map of total DW in SC170xM35-1 F5 generation.

The first one peaks at 54 Mbp, the second at 57.8 Mbp. The M35-1 allele at these two QTL contributes 12.42 g DW and 11.04 g DW, respectively. The second of these two loci coincides with the putative master regulator locus DW1. The fourth locus identified is on chromosome 10. For total DW, this is the only QTL where the presence of the allele from SC170 increases total DW of a plant. This QTL peaks at 56 Mbp and the allele from SC170 contributes 9.95 g DW. Taken together, these four QTL explain 44.75% of the variation in total DW observed in this set of F5 plants.

Maturation-related macro traits

Days to anthesis

Days to anthesis are measured by counting the number of days between emergence of the seedlings from the soil following planting and the appearance of pollen shed on the panicle following exsertion. Though both SC170 and M35-1 reach anthesis in 77 days in the field in College Station, Texas, their maturity genotypes are not identical. The existence of allelic variation in multiple genes contributing to regulation of flowering time makes it possible to generate a QTL map for days to anthesis in this population (Lin, et al., 1995).

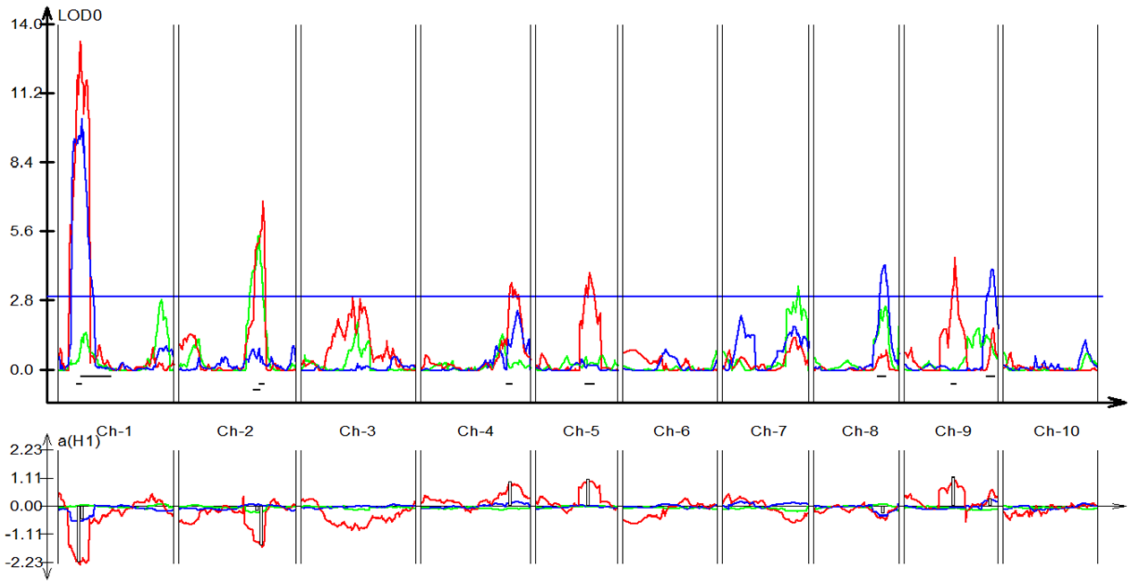

Figure 26: QTL map of maturity-related traits in SC170xM35-1 F5 generation. Traits include: days to anthesis (red), number of nodes (blue), leaf appearance rate (green).

There are five QTL that contribute to variation in days to anthesis in this population (Fig. 26). The five QTL explain 57.85% of the total variation in days to anthesis in this population. On chromosome one, a QTL was identified that can increase the number of days to anthesis by 2.22 days when the allele from SC170 is present. This QTL peaks at 8.2 Mbp, and the bounds of the locus includes Phytochrome A (PhyA), which can modify flowering time in *Oryza sativa* (Osugi, et al., 2011). The QTL identified on chromosome two peaks at 68.3 Mbp and increases the number of days to anthesis by 1.55 days when the allele from SC170 is present.

The three remaining loci that modulate days to anthesis are on chromosomes four, five, and nine; in all cases the allele from M35-1 increases

days to anthesis by approximately one day. The QTL on chromosome four peaks at 62.8 Mbp, while the QTL on chromosome five peaks at 54.9 Mbp and the QTL on chromosome nine peaks at 9.7 Mbp. As is often the case, the QTL on chromosome nine is very large, spanning from 7.5 Mbp through 47.8 Mbp. This indicates that the QTL identified maps in the pericentromeric region of this chromosome. There is often little crossing over that occurs in and near centromeres, so the genetic map distance (in cM) is very small in spite of the large number of bases in the same region. While such QTL are not ideal for positional cloning, the specific peak identified can often be very informative in spite of the size of the QTL (Mace and Jordan, 2011).

Number of nodes

This is the total number of stem nodes produced by a mature plant, counted from the first node at ground level to the node at the base of the peduncle (Vanderlip, 1993). Since each main stem leaf emerges from its respective main stem node, measurements of number of nodes are equivalent to measurements of the number of leaves on a plant. Given that the plants used in these studies had grown to anthesis, many of their lower leaves had senesced and were no longer attached to the plant. As such, counting the number of nodes is a more accurate measurement of leaf number for mature plants.

Three QTL were identified that regulate the total number of nodes (Fig. 26). In total, these QTL explain 45.15% of the total variation measured. The two QTL located on chromosomes one and eight each contributed to increased

numbers of nodes when the allele from SC170 is present, and have peaks at 10.9 Mbp and 52.6 Mbp, respectively. There may be overlap between the QTL identified on chromosome one and the QTL identified for days to anthesis that is also on chromosome one. The QTL located on chromosome nine peaks at 58.2 Mbp and contributes to an increased number of nodes when the allele from M35-1 is present.

Leaf appearance rate

LAR is measured by taking the total number of leaves (nodes) produced by a mature plant, divided by the number of days that it took to produce those leaves. The result is a number that represents the number of days necessary for the production and emergence of each new leaf (van Oosterom, et al., 2011).

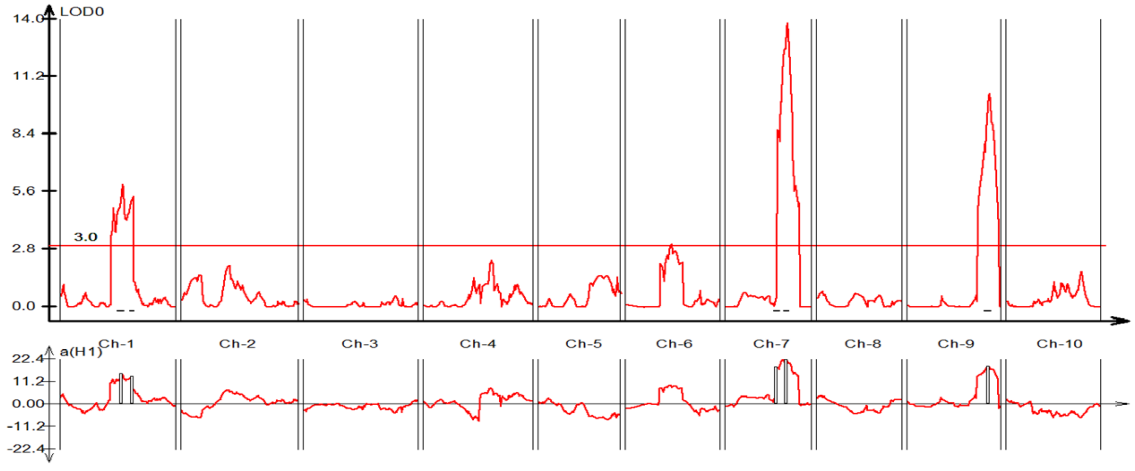

Figure 27: QTL map of stem length in SC170xM35-1 F5 generation.

Three QTL were identified to be controlling LAR in this background and can explain 30.57% of the total variation measured (Fig. 26). The strongest effect comes from the QTL on chromosome two, where the allele from SC170 can increase LAR by 0.21 days per leaf. The peak of this QTL is at 64.7 Mbp. The other two loci identified both increase the number of nodes when the allele from M35-1 is present. The first, on chromosome one, peaks at 68.2 Mbp while the second peaks at 52.8 Mbp on chromosome eight. The QTL on chromosome eight for LAR shows considerable overlap with the QTL for number of nodes on chromosome eight.

Table 6: Range of phenotype variation and QTL identified for selected stem biomass traits in SC170 x M35-1 F5 population.

Trait Name		Range		Number QTL	% Variance Explained
	Units	Min	Max		
Stem FW	g	53	525	3	49%
Stem DW	g	12	119	5	56%

Stem length

Stem length, which is measured from the first internode at the base of the plant to the base of the peduncle, not including the peduncle, varied greatly in this background (Table 6). The QTL identified all contribute to an increased stem length when the allele from M35-1 is present (Fig. 27). As expected, M35-1 stems are significantly longer than SC170 stems and the three QTL identified explained 93% of the total variation in this trait.

Two of these QTL, located on chromosomes seven and nine, correspond to known dwarfing loci in *S. bicolor* (Quinby and Karper, 1954). The QTL on chromosome seven peaks at 58.5 Mbp, coincident with Dw3, and alone accounts for 49% of the variation in stem length in this population. On chromosome nine, a QTL was identified which explains another 20.7% of the variation observed for stem length. This QTL peaks at 56.3 Mbp and maps coincident with the Dw1 locus in *S. bicolor*.

One additional stem length QTL is located on chromosome one and explains 23.3% of the variation observed for stem length. This QTL peaks at 51.5 Mbp and does not overlap with any known dwarfing loci in *S. bicolor*. For the remainder of this dissertation, this locus is referred to as the Dwx locus.

Total green leaf area

Total GLA was measured using a planimetric scanning apparatus and care was taken to ensure that only green leaves were assessed to contribute to this measurement. Following the conclusions made in chapter III, it is important to measure only green leaf area rather than total area as it has been established that senesced leaves may remain attached to the stem for some time after the end of their ability to photosynthesize.

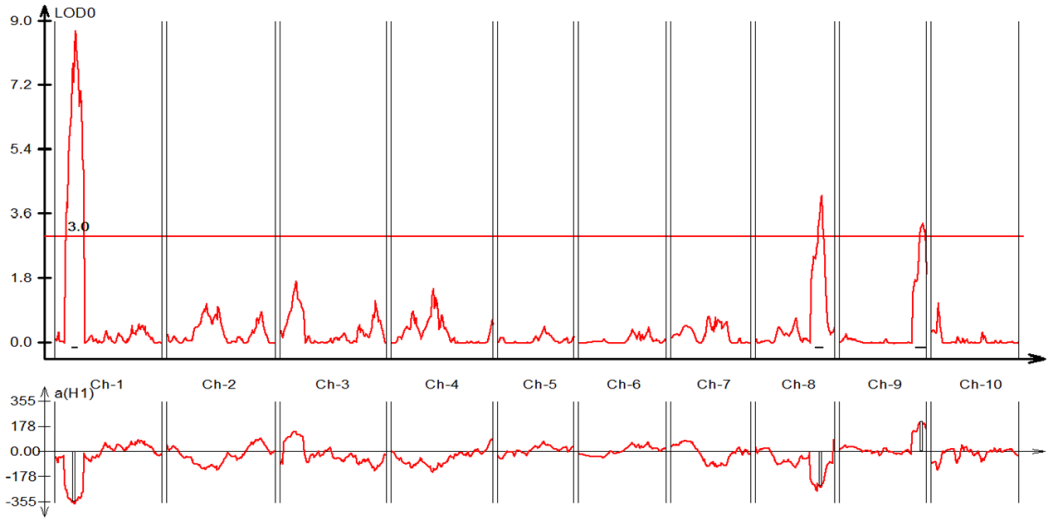

Figure 28: QTL map of total green leaf area (GLA) in SC170xM35-1 F5 generation.

While only 39.5% of the variation in GLA was attributable to the three QTL identified, these QTL are interesting because they are coincident with previously identified QTL for other macro traits in this population (Fig. 28). The QTL identified on chromosome one, for example, explains 21.7% of the variation in GLA observed. The allele from SC170 is responsible for increased GLA at this locus. With a peak at 8.2 Mbp, this QTL overlaps with the observed QTL for days to anthesis, putatively identified as PhyA. The presence of the SC170 allele of PhyA can increase both days to anthesis and total GLA. This finding is consistent with the fact that each successive leaf produced by an *S. bicolor* plant during the vegetative phase is larger than the previous leaf (Thomas, et al., 2002). Therefore, plants that flower later will be expected to have larger upper leaves and a greater total GLA.

The QTL for GLA identified on chromosome eight also contributes to increased GLA when the allele from SC170 is present. This QTL peaks at 52.8 Mbp, aligning very closely to the QTL on chromosome eight for both number of nodes and LAR. This allele from SC170 contributes to increased GLA, most likely by decreasing LAR and thereby increasing the number of nodes and leaves produced by the plant given a constant number of days to floral induction. The increased rate of leaf production would also be expected to result in greater seedling vigor, earlier establishment of a complete leaf canopy, and increased biomass accumulation in any *S. bicolor* genotype with a long duration vegetative phase (van Oosterom, et al., 2011).

The third QTL identified for GLA is located on chromosome nine, with a peak at 58.5 Mbp. The allele at this locus that contributes to greater GLA comes from M35-1. While this QTL only accounts for 7.6% of the variation in GLA, it is coincident with the previously identified QTL for number of nodes and total DW. The allele from M35-1 leads to increased GLA, a greater number of nodes, and an increase in total DW.

Stem traits

Each complete stem was weighed so that QTL maps could be generated for both whole stem FW and whole stem DW. There was considerable variation observed in both stem FW and stem DW (Table 6). For the purposes of this study, once macro stem traits were measured, the stem was then subdivided into internodes for further measurement. This allowed for quantification of

genetic variation in individual internode size in addition to total stem size. Both the length and diameter of each internode was measured, allowing for resolution between genomic regions controlling only length, only diameter, or both for each internode of the stem. There was considerable variation in internode length and diameter among the lines in this population, which will be discussed in a later section.

While there were a small number of QTL identified that had specific effects on either length or diameter of one individual internode, most of the QTL identified affected phenotypic variation of internodes 4-10 of the stem. While the peduncle is visually similar to the rest of the stem, this is actually a separate organ and is not discussed in this dissertation.

Stem biomass traits

Stem biomass was assessed in terms of fresh weight (FW) and dry weight DW (Table 2). There were three QTL identified for stem FW, all three of which were also identified for stem DW (Fig. 29). There was one QTL on each of chromosomes one, seven, and nine. All of these QTL increase the FW and DW of the stem when the allele from M35-1 is present. The QTL with the largest effect is, which is located on chromosome nine, has a peak at 55.8 Mbp. This locus explains over 20% of the variation in both stem FW and DW. This QTL is coincident with the QTL on chromosome nine that corresponds to Dw1 in *S. bicolor*.

The other two QTL that modulate both stem FW and DW have peaks at 54.6 Mbp and 58.4 Mbp on chromosomes one and seven, respectively. The QTL on chromosome seven corresponds to Dw3 in *S. bicolor* while the QTL on chromosome one corresponds to the Dwx locus identified in this study.

In addition to the three QTL that affect both FW and DW in the stems, there were two additional QTL that contribute to variation in stem DW that were not found to contribute to variation in stem FW. The first of these, located at 26.3 Mbp on chromosome one, explained 7.5% of the variation in stem DW and

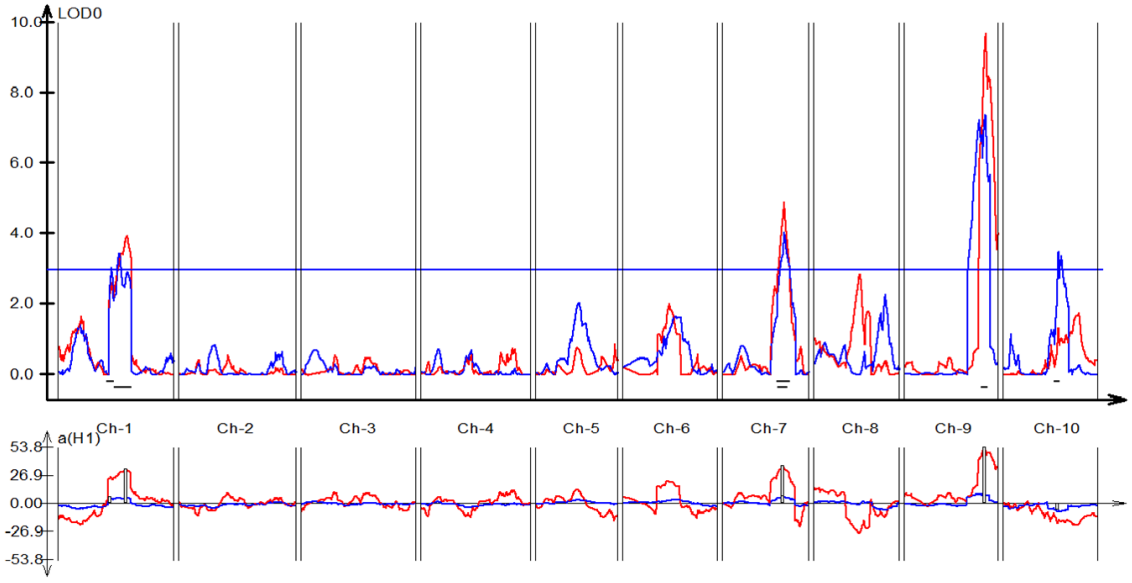

Figure 29: QTL maps of stem FW (red) and stem DW (blue) for SC170xM35-1 F5 generation.

the allele that contributes to increased stem DW at this locus is from M35-1. In contrast, the QTL on chromosome ten contributes to increased stem DW when the allele from SC170 is present and can account for 8.6% of the variation. The

peak of this QTL is at 51.1 Mbp. Neither of these two QTL modulates variation in any of the other macro traits listed previously.

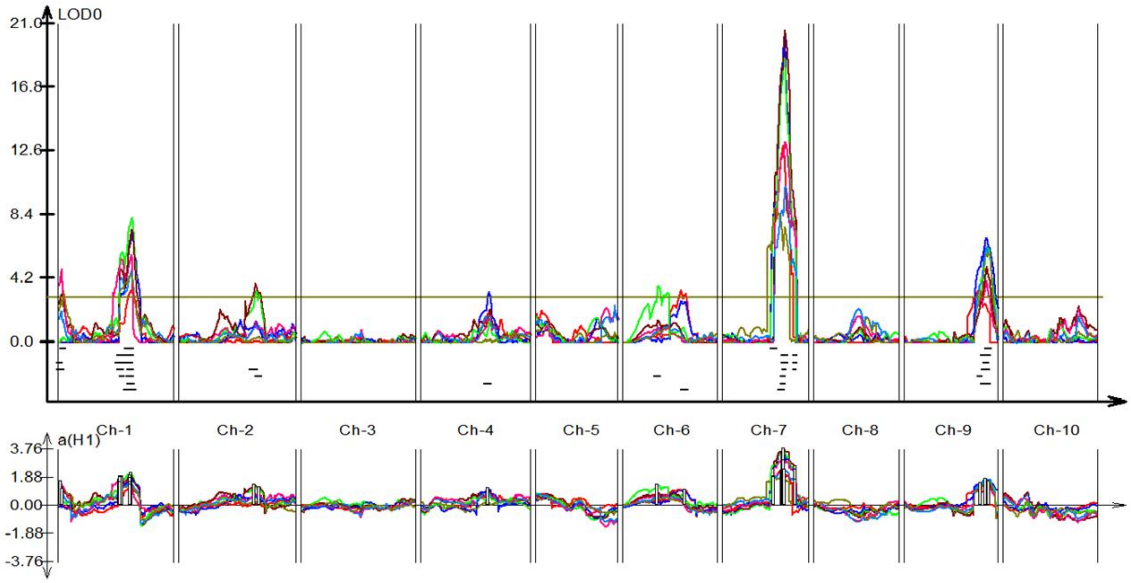

Figure 30: QTL maps of individual internode lengths in SC170xM35-1 F5 generation. Traits include: Internode 4 length (red), Internode 5 length (blue), Internode 6 length (light green), Internode 7 length (maroon), Internode 8 length (fuschia), Internode 9 length (light blue), and Internode 10 length (gold).

Individual internode lengths

Each internode of the main stem was measured individually in the SC170*M35-1 F5 population. Internodes one to three, the most juvenile internodes, were highly variable in size and did not yield reliable QTL with high LOD scores; for internodes above number ten, the length data was also highly variable and not useful for QTL mapping (data not shown). However, length measurements for internodes four through ten yielded high quality QTL maps for each internode (Fig. 30).

Table 7: Range of phenotype variation and QTL identified for individual internode length traits in SC170xM35-1. All units are cm.

Trait Name	Range		Number QTL	% Variance Explained
	Min	Max		
Int 4 Length	2	19	3	47%
Int. 5 Length	2	24	4	75%
Int. 6 Length	3	27	3	69%
Int. 7 Length	2	24	4	99%
Int. 8 Length	2	24	4	56%
Int. 9 Length	3	21	3	49%
Int. 10 Length	3	20	4	55%

Three main QTL modulated the length of most all of the internodes between four and ten. These three loci were located on chromosomes one, seven, and nine, and correspond to putative Dwx, Dw3, and Dw1, respectively (Quinby and Karper, 1954; Brown, et al., 2008). For each of these QTL, the allele for increased internode length comes from M35-1. Dw3, which is located on chromosome seven, has the largest effect on internode length and explains 20 to 60% of the variation observed in internode length, depending on internode number (Table 7). Dwx is located on chromosome one and accounts for 7 – 25% of the total internode length variation. Dw1, which is located on chromosome nine, accounts for 7 – 15% of the total variation in internode length.

For many of the internodes measured in this study, additional QTL modulate the length of a specific individual internode (Table 7). For example, in

addition to Dw3 and Dwx, the length of internode four is controlled by a QTL on chromosome six, which peaks at 48.9 Mbp and explains 6.5% of the total variation in length. In the case of internode seven, over 99% of the variance in length is explained by Dwx, Dw3, Dw1, and a QTL on chromosome two that peaks at 62.5 Mbp.

Table 8: Range of phenotype variation and QTL identified for individual internode diameter traits in SC170xM35-1. All units are mm.

Trait Name	Range		Number QTL	% Variance Explained
	Min	Max		
Int. 4 Diam.	13	23	2	28%
Int. 5 Diam.	12	23	3	39%
Int. 6 Diam.	12	22	3	43%
Int. 7 Diam.	11	21	3	48%
Int. 8 Diam.	11	20	3	39%
Int. 9 Diam.	10	19	1	14%
Int. 10 Diam.	10	18	1	11%

Individual internode diameters

The diameter of each individual internode was measured for use in QTL mapping. Internodes four through 10 yielded diameter data that was useful for QTL mapping (Table 8). A QTL for stem diameter, shared by internodes 4 – 8, was identified on chromosome seven, with a peak at 58.4 Mbp (Fig. 31). This QTL corresponds to Dw3; the SC170 allele of Dw3, which decreases internode

length, contributes to increased internode diameter. Between 15 and 20% of the total variation in internode diameter is explained by this QTL.

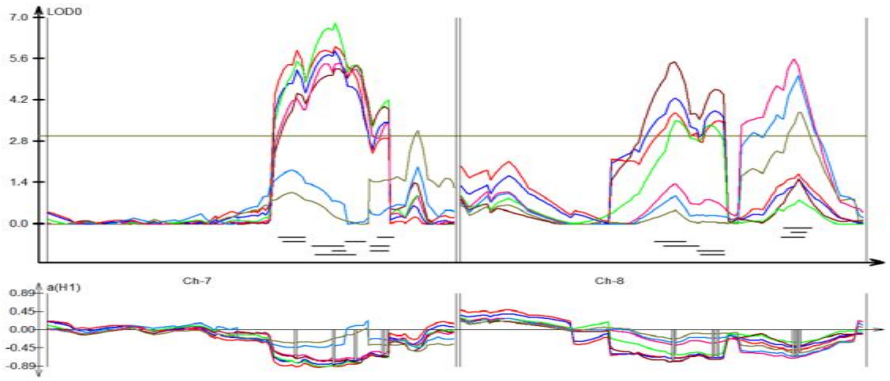

Figure 31: QTL maps for internode diameter in SC170xM35-1 F5 generation. Only chromosomes 7 and 8 are shown. Traits include: Internode 4 length (red), Internode 5 length (blue), Internode 6 length (light green), Internode 7 length (maroon), Internode 8 length (fuschia), Internode 9 length (light blue), and Internode 10 length (gold).

For the lower internodes (four and five), a QTL on chromosome eight with a peak at 46 Mbp explains ten percent of the variation observed. For internodes six and seven, a QTL on chromosome eight with a peak at 6.5 Mbp explains 10 – 15% of the remaining variation. Internodes 8-10 have a QTL on chromosome eight with a peak at 53.1 Mbp that explains 10-15% of the total variation. For internodes five through eight, a second QTL on chromosome seven, with a peak at 61.8 Mbp explains approximately 10% of the total variation. This set of regulatory loci demonstrates the complex nature of the regulatory system that modulates stem diameter of *S. bicolor*.

Leaf traits

The leaves of SC170*M35-1 F5 plants were characterized in multiple ways. First, total GLA was measured as previously reported. In addition, individual leaves were measured; length, width, and area were recorded for each leaf. Chlorophyll per unit leaf area estimates were obtained in the form of SPAD readings. Then the whole leaf biomass was measured when fresh (FW) and after drying (DW). Table 9 shows the range in variation observed in the leaf mass and area traits measured in this population.

Table 9: Phenotype variation and QTL identified for macro leaf traits and individual leaf area traits in SC170xM35-1 F5 generation.

Trait Name		Range		Number QTL	% Variance Explained
	Units	Min	Max		
Leaf FW	g	57	155	2	28%
Leaf DW	g	13	43	4	70%
Flag leaf area	cm^2	66	343	1	17%
Leaf 2 area	cm^2	153	555	1	20%
Leaf 3 area	cm^2	238	655	1	30%
Leaf 4 area	cm^2	317	715	2	37%
Leaf 5 area	cm^2	295	702	4	49%
Leaf 6 area	cm^2	237	646	4	50%
Leaf 7 area	cm^2	170	666	3	51%

Leaf biomass traits

Leaf biomass FW and DW were measured and the data was used in QTL analysis. All of the leaves that were still physically attached to the stem of a plant at the time of harvest were assayed. As was the case for stem FW and DW, there was some genetic control found to be shared between these two traits, as well as some control specific to each trait (Fig. 32). Only one QTL was found to be in common between these two traits. The peak of this QTL is located at 53.2 Mbp on chromosome eight, with the allele from SC170 contributing to both increased leaf FW and leaf DW. This QTL only explains 11 and seven percent of the variation in leaf FW and leaf DW, respectively.

There was only one additional QTL identified to be controlling leaf FW. Located on chromosome one with a peak at 2.3 Mbp, the allele from SC170 at this locus contributes to increased leaf FW and explains 16.6% of the total variance observed for leaf FW. In contrast, leaf DW exhibits genetic control by three loci in addition to the one locus that leaf FW and DW have in common. Two of these, both located on chromosome one, contribute to increased leaf DW when the allele from SC170 is present. These two loci have peaks of 6.2 Mbp and 11.6 Mbp, respectively, and explain a total of 50.4% of the variation in leaf DW when taken together. The remaining QTL for leaf DW is on chromosome nine, with a peak at 58.5 Mbp. At this locus, the allele from M35-1 contributes to increased leaf DW, and can explain 12.1% of the total variance in the leaf DW measurements. This QTL is positioned very near the Dw1 locus

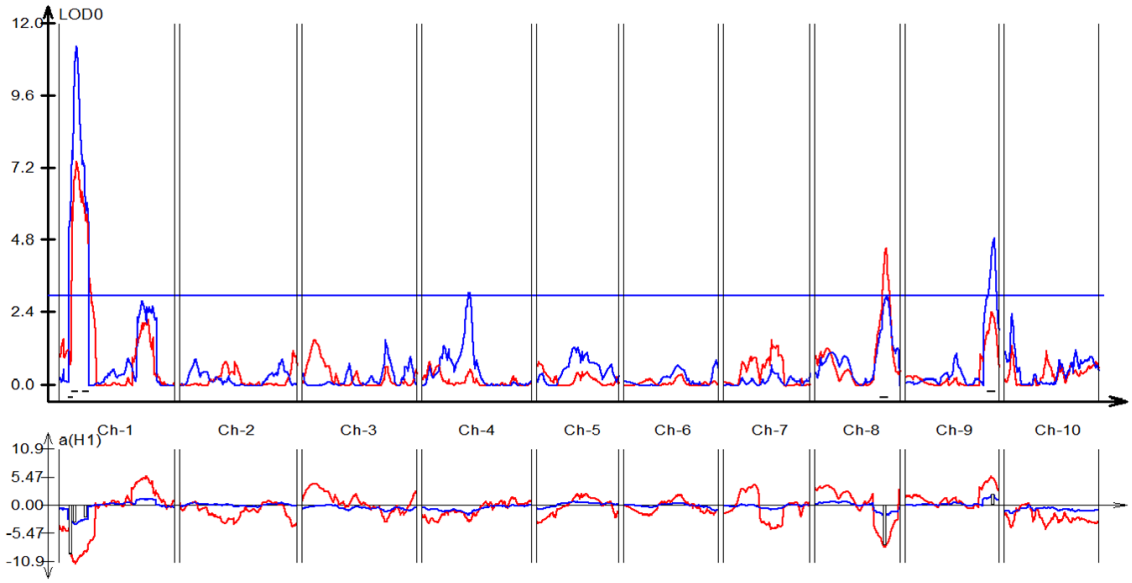

Figure 32: QTL maps of leaf FW (red) and leaf DW (blue) for SC170xM35-1 F5 generation.

Leaf areas

Individual leaf areas were measured for each of the top leaves of each plant, beginning with the flag leaf. Successive leaves below the flag leaf are numbered in descending order beginning with two and going through seven (flag – 6). There is a complex but apparent pattern of changing control of leaf area when moving down through the canopy (Fig. 33).

The pattern begins with the flag leaf, where the area of this leaf is apparently controlled by a QTL on chromosome ten that has a peak at 57.1 Mbp. This QTL explains 16.5% of the total variation observed, and the presence of the SC170 allele at this locus contributes to a larger flag leaf. The same QTL is responsible for the variation seen in the areas of leaf two and leaf three. In

both cases the allele from SC170 contributes to a larger leaf. In the case of leaf two, this QTL explains 19.6% of the variation. For leaf three, the variation explained by the QTL on chromosome 10 increases to 29.6% of the total.

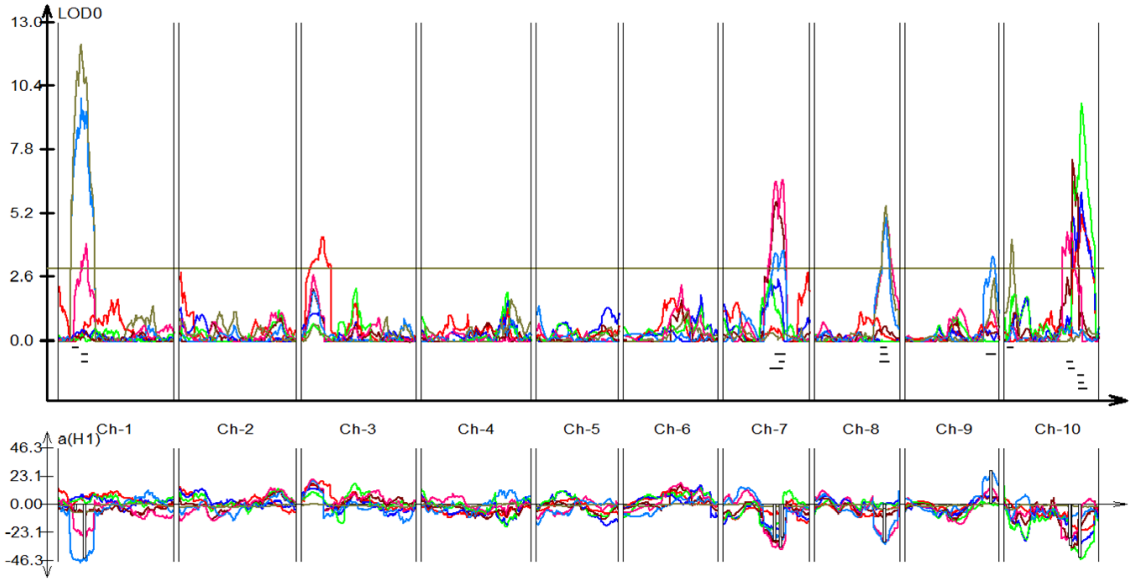

Figure 33: QTL maps of individual leaf areas in SC170xM35-1 F5 generation. Traits included are: flag leaf (red), leaf 2 (blue), leaf 3 (green), leaf 4 (maroon), leaf 5 (pink), leaf 6 (light blue), leaf 7 (gold).

The area of leaf four is controlled by two QTL. The same QTL responsible for variation in the areas of the top three leaves is also identified for leaf four, with the SC170 allele contributing to a larger leaf size, and with this QTL explaining 20.7% of the variation. A second QTL modulates the area of leaf four area. It is located on chromosome seven, with a peak at 57.2 Mbp, and explains 15.7% of the variation observed. This QTL is consistent with the position of the Dw3 locus in *S. bicolor*.

There are four QTL identified that modulate the area of leaf five. Two of these, located on chromosomes seven and 10, are the same QTL identified for the areas of the previously described leaves (Fig. 33). Together they explain 28.0% of the variation in leaf five area. The remaining two QTL, the first on chromosome one, the second on chromosome eight, explain an additional 20.8% of the variation. For both of these QTL, the SC170 allele confers larger leaf five area. On chromosome one, the QTL peak is located at 11.6 Mbp. The QTL on chromosome eight has a peak at 52.8 Mbp, consistent with the position of the QTL found to modulate both leaf FW and leaf DW. There may be some element of genetic control shared by the QTL that regulates leaf five area, leaf FW, and leaf DW.

Four QTL were also identified to be controlling leaf six area as well. The three QTL on chromosomes one, seven, and eight, are the same as those identified as controlling leaf five area (Fig. 33). Together, these three QTL explain 41.9% of the variance measured in leaf six area. However, the QTL on chromosome ten, which had been present in the QTL map of every leaf area from the flag leaf down, did not modify leaf six area. Instead, a QTL on chromosome nine was identified which has a peak at 59.8 Mbp and explains 8.0% of the total variance. This QTL is unique because it is the only individual leaf area QTL identified so far where the contribution of the allele from M35-1 leads to larger leaf area.

The control of leaf seven area mirrors that of leaf six area with one exception. The QTL on chromosome seven, hypothesized to co-locate with Dw3, is not identified as having an effect on leaf seven area. The three QTL on chromosomes one, eight, and nine explain a total of 51% of the variance observed in leaf seven area in this population.

Table 10: Phenotypic variation and QTL identified for individual leaf length traits in SC170xM35-1 F5 generation.

Trait Name		Range		Number QTL	% Variance Explained
	Units	Min	Max		
Flag leaf L.	cm	33	117	1	13%
Leaf 2 L.	cm	51	94	2	30%
Leaf 3 L.	cm	57	101	4	41%
Leaf 4 L.	cm	61	101	2	24%
Leaf 5 L.	cm	58	100	4	43%
Leaf 6 L.	cm	55	107	3	49%
Leaf 7 L.	cm	45	104	3	51%

Leaf lengths

In addition to individual leaf areas, individual leaf lengths were used to create QTL maps in this background. There is considerable variation in leaf lengths for each of the leaves considered (Table 10). As might have been expected, there are some loci that appear to control leaf length on a global scale, evident by their presence on the QTL maps for multiple individual leaves, as well as loci that are clearly unique to individual leaf lengths (Fig. 34).

The length of the flag leaf in this population is controlled by a single locus which is located on chromosome three with a peak at 56 Mbp (Fig. 34). This QTL explains 13.1% of the variance observed and it is the allele from SC170 that contributes to a longer flag leaf.

Leaves two through five are controlled by a selection of shared loci. There are two main QTL identified that are shared by all four of these leaf length traits. The first is on chromosome six, with a peak at 50.2 Mbp. This QTL explains 8-15% of the total variation observed and the allele from M35-1 contributes to longer leaf length. In contrast, the second QTL in common is located on chromosome 10 with a peak at 55 Mbp and it is the SC170 allele which contributes to longer leaf lengths in this case. This second QTL explains 7-15% of the total variance in leaf length for leaves two through five. The position of this QTL indicates that it is the same locus that was previously identified as controlling a portion of the variation in total DW, with the allele from SC170 contributing both to larger total DW and increased leaf length. There are a small number of additional QTL that are unique to individual leaf lengths within this group which can be seen in figure 34.

The length of leaf six is controlled by three loci. For all three loci, the presence of the allele from SC170 leads to increased leaf length (Fig. 34). Taken together, these loci explain 48.5% of the variance measured in leaf six length. The first is located at 11.6 Mbp on chromosome one. The second locus is on chromosome seven, with its peak at 58.5 Mbp. This position corresponds

to Dw3, which has been discussed previously. The third locus, located on chromosome eight, has a peak at 52.8 Mbp and co-locates with previously identified QTL for leaf FW, leaf DW, and the individual areas of lower leaves.

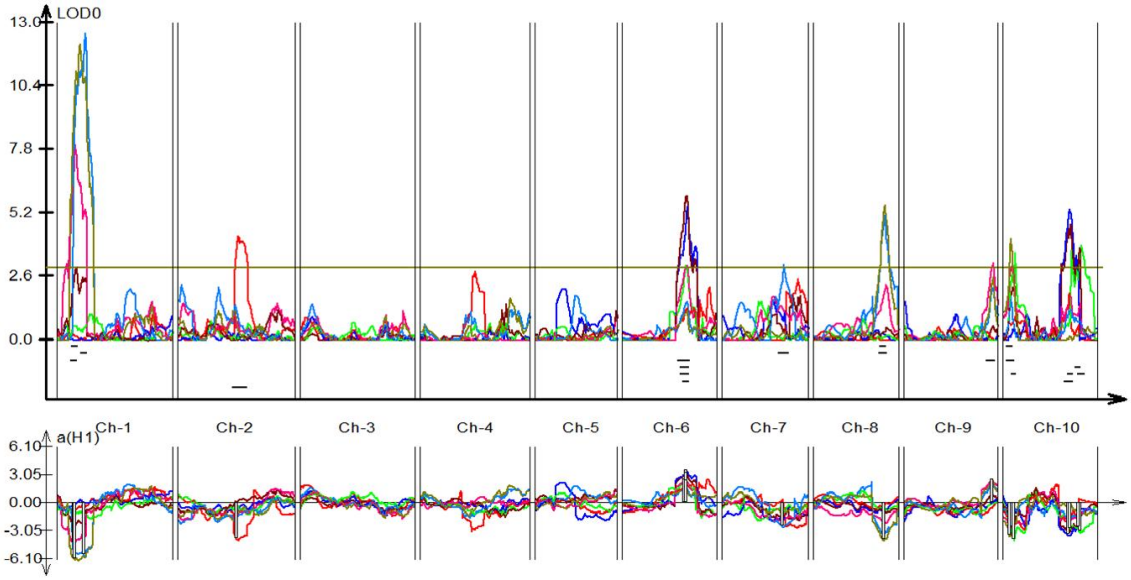

Figure 34: QTL maps of individual leaf lengths for top 7 leaves in SC170xM35-1 F5 generation. Traits included are: flag leaf (red), leaf 2 (blue), leaf 3 (green), leaf 4 (maroon), leaf 5 (pink), leaf 6 (light blue), leaf 7 (gold).

Leaf seven length is also controlled by the locus on chromosome eight, which explains 12.2% of the variance in leaf seven length. There are two other loci contributing to the regulation of the length of leaf seven (Fig. 34). These are located on chromosomes one and 10 and explain 29.4% and 9.1% of the total variance, respectively. For the QTL on chromosome one, the peak is located at 6.4 Mbp. The QTL on chromosome 10 has its peat at 1.6 Mbp.

Leaf widths

The QTL identified for individual leaf widths do not follow a simple progression like the QTL for leaf areas did. The regulation of leaf width appears to be quite complex and QTL identified vary greatly between the different leaves of *S. bicolor* plants of this background (Fig. 35 and Table 11).

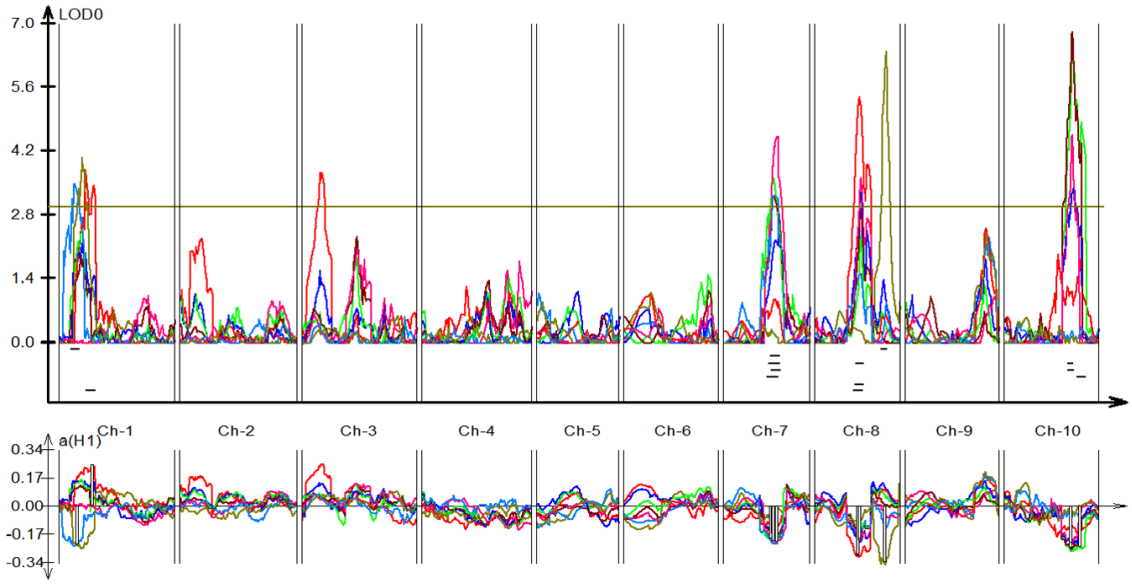

Figure 35: QTL maps of individual leaf widths for top 7 leaves in SC170xM35-1 F5 generation. Traits included are: flag leaf (red), leaf 2 (blue), leaf 3 (green), leaf 4 (maroon), leaf 5 (pink), leaf 6 (light blue), leaf 7 (gold).

The top leaf, or flag leaf, exhibits genetic control of width by two loci (Fig. 35). The first, located on chromosome one with a peak at 14.8 Mbp, accounts for 10.1% of the variance observed. For this QTL, the presence of an allele from M35-1 leads to an increase in flag leaf width. The second QTL identified is on chromosome eight, with a peak at 6.4 Mbp. This second locus accounts for

15.3% of the variance; the allele from SC170 contributes to greater leaf width. For the width of leaf two, only this QTL on chromosome eight was identified as playing a role in control of the phenotype. The presence of the SC170 allele at this locus explains 10.3% of the variance observed for increased leaf two width.

Table 11: Phenotypic variation and QTL identified for individual leaf width traits in SC170xM35-1 F5 generation.

Trait Name		Range		Number QTL	% Variance Explained
	Units	Min	Max		
Flag leaf W.	cm	1.3	4.8	2	25%
Leaf 2 W.	cm	2.9	5.9	1	10%
Leaf 3 W.	cm	4.1	7.1	2	25%
Leaf 4 W.	cm	4.5	7.1	2	28%
Leaf 5 W.	cm	4.5	7.7	3	36%
Leaf 6 W.	cm	3.9	7.1	1	10%
Leaf 7 W.	cm	3.4	7.5	2	28%

The regulation of leaf width is very different for leaves three and four. The QTL map generated for each of these leaf width traits identified two QTL (Fig. 35). The first, located at 56.2 Mbp on chromosome seven, explains approximately 9% of the variance observed and the presence of the SC170 allele increases leaf width in both traits. This QTL position is consistent with the previous identification of this locus as Dw3. The second QTL controlling width of leaves three and four is located on chromosome 10. The peak of this QTL is at 57.1 Mbp and the allele from SC170 leads to increased leaf width at this locus.

Regulation of leaf five width is unique in that it includes both the QTL on chromosomes seven and 10 that were most recently described, as well as the QTL on chromosome eight that was identified in the QTL maps of leaf width for the uppermost leaves (Fig. 35). Taken together, these three QTL can account for 36.5% of the total variance in leaf five width. Leaf six width regulation is quite simple and involves only the locus on chromosome seven that is currently referred to as Dw3. The regulation of the width of leaf seven is the same as the regulation described for the width of the flag leaf.

Table 12: Phenotypic variation and QTL identified for individual leaf SPAD levels in SC170xM35-1 F5 population.

Trait Name	Range		Number QTL	% Variance Explained
	Min	Max		
Leaf 2 SPAD	44	58	1	13%
Leaf 3 SPAD	44	64	1	19%
Leaf 4 SPAD	48	65	1	14%
Leaf 5 SPAD	46	65	1	9%
Leaf 6 SPAD	48	65	3	34%
Leaf 7 SPAD	43	64	1	17%

Leaf chlorophyll (SPAD) estimates

In addition to the physical size and weight of the leaves of a plant, it is also important to measure the photosynthetic capacity of those leaves. SPAD assays of chlorophyll content per unit leaf area were used as an indirect assay

of photosynthetic electron transport capacity (Ranjith and Meinzer, 1997; Thomas, et al., 2002).

Using the same numbering scheme described for individual leaf areas, the top seven leaves of each plant in the SC170*M35-1 F5 population were assessed for variation SPAD readings (Table 12). While there were no significant QTL identified for the SPAD level of the flag leaf in this case, there were strong QTL identified for control of SPAD level of other, lower leaves (Fig. 36).

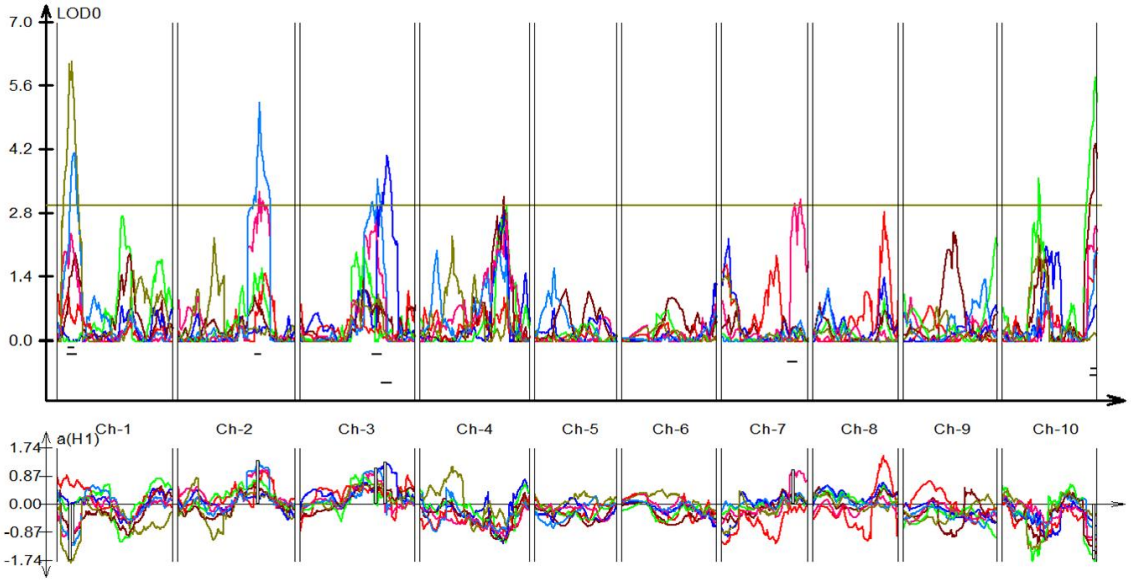

Figure 36: QTL maps of individual leaf SPAD levels for top 7 leaves in SC170xM35-1 F5 generation. Traits included are: leaf 2 (blue), leaf 3 (green), leaf 4 (maroon), leaf 5 (pink), leaf 6 (light blue), leaf 7 (gold). No QTL were identified for flag leaf SPAD level, so this trait is omitted here.

The SPAD level QTL map generated for leaf two showed a single QTL located on chromosome three, which explained 12.7% of the variance observed. This QTL had a peak at 67.5 Mbp and the allele from M35-1 contributed to increased SPAD level for this locus.

The SPAD levels of leaves three and four were both controlled by the same single QTL (Fig. 36). Located on chromosome ten with a peak at 60.3 Mbp, this locus explains 18.8% and 13.7% of the total variation in SPAD level for the two leaves, respectively. In both cases, the presence of the allele from SC170 contributes to an increased SPAD level.

Leaf five SPAD level is controlled by a solitary QTL located at 61.8 Mbp on chromosome seven that does not co-locate with Dw3 (Fig. 36). Leaf six SPAD level is also uniquely controlled by three QTL. A QTL on chromosome one, with its peak at 6.2 Mbp, explains 10.9% of the variance observed and the presence of an allele from SC170 leads to increased SPAD level. The other two QTL, located on chromosomes two and three, explain a total of 22.8% of the variance observed. In both cases, the allele from M35-1 contributes to increased SPAD levels. In the case of leaf seven SPAD, there is only one QTL identified. It is located on chromosome one, with a peak at 2.9 Mbp, and explains 17.2% of the variance observed. At this locus, the allele from SC170 contributes to increased SPAD levels in leaf seven.

Composition traits

Once dried, stem biomass was ground and scanned using near-infrared spectroscopy (NIR) to assess the composition of the material. These scans are then analyzed to yield relative composition of the biomass based on nine categories of components (Table 13) (Wolfrum et al., in prep). Each of these components can then be treated as a trait for QTL mapping (Murray, et al., 2008b; Murray, et al., 2008a). For some of the components, the total quantity in grams, in addition to the relative quantity, in percentage, was used for QTL mapping. In such cases, both results are reported.

Table 13: Phenotypic variation and QTL identified for stem biomass composition traits in SC170xM35-1 F5 population.

Trait Name	Range		Number QTL	% Variance Explained
	Min	Max		
% Cellulose	18.4%	28.4%	2	29%
% Lignin	80.0%	13.0%	3	59%
% Xylan	99.6%	15.5%	3	44%
% Galactan	0.7%	0.9%	4	59%
% Arabinan	0.4%	2.7%	4	49%
% Sucrose	13.1%	28.6%	4	50%
% Protein	0.3%	3.1%	3	31%
% Ash	2.5%	5.2%	2	19%
% Extractives	33.1%	50.9%	3	58%

Major cell wall components

Cellulose

QTL mapping was carried out on multiple cell wall component molecules, including one of the main components of plant cell walls, cellulose (Powell, et al., 1991). When mapped as a relative percentage of the total composition, two QTL were identified as regulating cellulose level (Fig. 37). Both were located on chromosome one, each explains 14.5% of the total variance, and in both cases, the allele from M35-1 contributes to increased cellulose percentage. Their peaks are located at 7.9 Mbp and 11.6 Mbp, respectively.

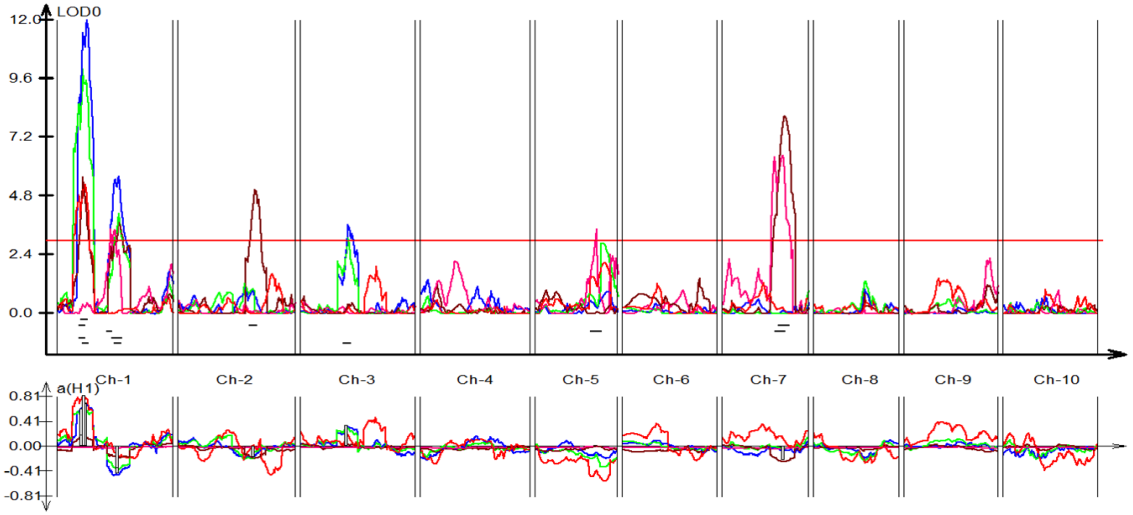

Figure 37: QTL maps of individual cell wall component percentages of stem biomass in SC170xM35-1 F5 population. Components included are: cellulose (red), lignin (blue), xylan (green), galactan (maroon), and arabinan (pink).

Mapping of QTL for the total number of grams of cellulose present in the stem tissue yielded completely different QTL compared to those described in the previous paragraph (Fig. 38). Two loci were identified for this trait, and in both cases the allele from M35-1 contributed to a greater number of grams of cellulose in the stem. The first of these, on chromosome seven, has a peak at 58.5 Mbp and explains 16.8% of the total variance. This locus co-locates with Dw3. The second QTL identified in this case was on chromosome nine, with a peak at 56.0 Mbp. This locus explains 23.2% of the total variance measured and is located in the same position as Dw1.

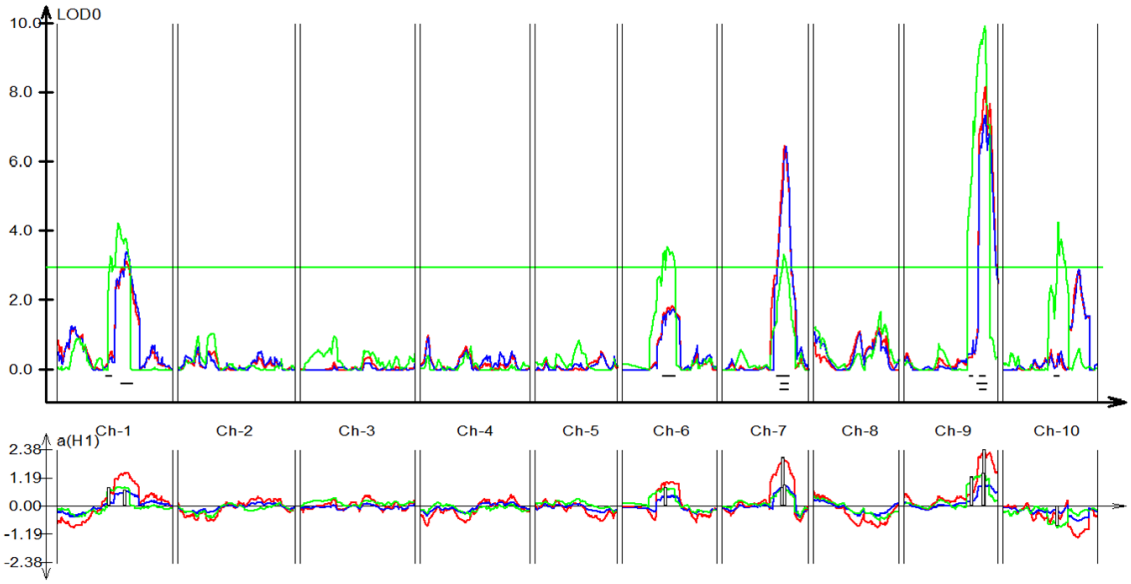

Figure 38: QTL maps of individual cell wall components (in grams) of stem biomass in SC170xM35-1 F5 population. Components included are: cellulose (red), lignin (blue), and xylan (green).

Xylan

Another major cell wall component is xylan, which is also referred to as hemicellulose (Sanderson, et al., 1996a). Mapping of the relative percentage of xylan in stem tissue identified three QTL (Fig. 37). The first of these, located at 10.9 Mbp on chromosome one, explains 26.7% of the variance observed. In this case, the allele from M35-1 contributes to an increased percentage of xylan. The second QTL identified is also located on chromosome one, with a peak at 51.3 Mbp; this location is consistent with the position of Dwx. The allele from SC170 contributes to increased xylan percentage in this case. The final locus identified for this trait, with a smaller phenotypic effect (7.3%), was located on chromosome three, with a peak at 12.7 Mbp. The presence of the allele from M35-1 leads to increased xylan percentage at this locus.

Mapping of QTL for total grams of xylan present in stem tissue reveals much more complex regulation. A total of six loci were identified for this trait (Fig. 38). Only one of these loci, located on chromosome 10, contributes to an increase in the number of grams of xylan in the presence of the allele from SC170. With a peak at 51.1 Mbp, and explaining 9.9% of the variance, this locus aligns with a previously identified QTL for stem DW. All five of the remaining QTL lead to a greater number of grams of xylan in the stem in the presence of the allele from M35-1. There is a locus on chromosome one which also co-locates with a QTL for stem DW, peaking at 26.2 Mbp. A locus on chromosome six is unique to this trait, with a peak at 38.5 Mbp. Chromosome

nine has two loci for grams of xylan. The first explains 20.4% of the variance observed and peaks at 51.9 Mbp. The second, explaining 26.8% of the variance, is consistent with the position of Dw1. Finally, a locus was identified on chromosome seven which is in the location of Dw3.

Lignin

When mapped as a percentage of total composition, lignin content yielded three QTL (Fig. 37). All three of these are consistent in physical location and phenotypic effect with the loci identified for xylan percentage. The three loci explain 59% of the total variance observed in lignin percentage.

When mapped in terms of total grams of lignin present in the stem, the loci identified were completely different (Fig. 38). Three loci were again identified, but these three loci corresponded to putative Dwx, Dw3, and Dw1. For all three loci, the presence of the allele from M35-1 contributes to an increase in the number of grams of lignin in the stem.

Minor cell wall components

Minor cell wall components were assessed in terms of relative percent composition. These components, galactan and arabinan, are the building blocks of pectin, another important part of plant cell walls (Zhao, et al., 2009; Sanderson, et al., 1996a).

Galactan

Three loci were identified as playing a role in regulation of galactan percentage in stem tissue (Table 13). Each of these three QTL are co-located

with QTL identified for other traits (Fig. 37). First, on chromosome one, with a peak at 26.3 Mbp, is a QTL which was also identified for stem DW. Second, on chromosome five there is a QTL which was previously noted to have an effect on days to flowering. Third and finally, a locus was identified on chromosome seven which is consistent with the position of Dw3. The loci located on chromosomes one and seven can contribute to increased galactan percentage in the presence of an allele from SC170 while the locus on chromosome five can contribute to increased galactan percentage in the presence of an allele from M35-1. Taken together, these three loci explain 58.7% of the total variance observed for galactan percentage in stem tissue.

Arabinan

When QTL mapping of arabinan percentage in stem biomass was carried out, four loci were identified to have a role in regulating this trait (Table 13). As was the case for galactan, each of these loci is in a position that is consistent with a locus that has previously been identified for another trait in this study (Fig. 37). The first of these loci, located on chromosome one, is coincident with a QTL for number of nodes and can contribute to increased arabinan percentage in the presence of the allele from M35-1. The remaining three loci for this trait all contribute to increased arabinan percentage in the presence of an allele from SC170. The remaining three are located on chromosomes one, two, and seven and correspond to QTL for lignin and xylan percentages, internode seven length,

and the Dw3 locus, respectively. In total, these loci account for 48.8% of the variation observed for stem biomass arabinan percentage.

Other components

Protein

The percentage of stem biomass that is made up of protein is an important trait to map. Unlike leaf tissue, where photosynthetic capacity will determine much of the protein content, there is little photosynthesis occurring in stems of *S. bicolor* plants and as such the protein content can give insight into the overall structure and biochemical activity of the stem (Makino, et al., 2003). When QTL mapping was carried out for protein as a percentage of dry weight, three loci were identified (Fig. 39). All three loci identified can lead to increased stem protein percentage in the presence of the allele from SC170, and the cumulative effects of these three loci can account for up to 31.3% of the total variance observed (Table 13).

The first locus identified co-locates with the previously identified locus for stem DW that is located on chromosome one. The second, also located on chromosome one, co-locates with a QTL shown to be controlling the relative percentages of lignin and xylan. The third and final QTL identified for this trait, on chromosome seven, is in the position of Dw3.

Sucrose

When mapped as a relative percentage of total stem biomass, sucrose levels yielded four loci (Fig. 39 and Table 13). The first of these, located at 10.9

Mbp on chromosome one, is the only locus for this trait where the allele from SC170 can lead to increased trait value. For all three other QTL identified, the presence of the allele from M35-1 is what contributed to increased sucrose percentage. The remaining three loci occur on chromosomes one, two, and seven and correspond to QTL for lignin and xylan percentage, internode seven length, and Dw3, respectively. In total, these four loci account for 49.9% of the total variance observed in sucrose percentage.

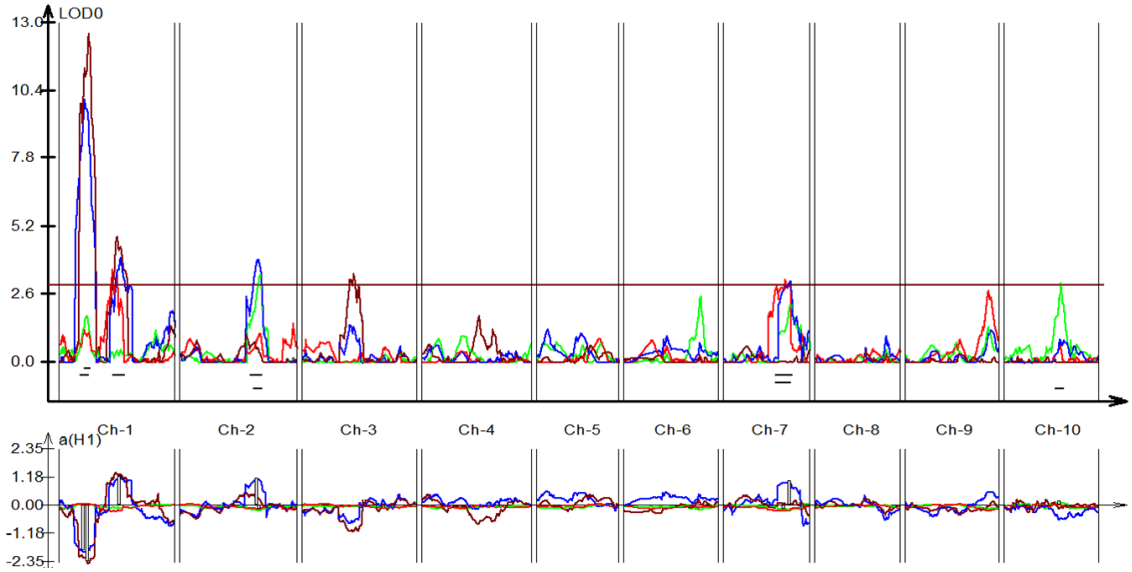

Figure 39: QTL maps of individual soluble component percentages of stem biomass in SC170xM35-1 F5 population. Components included are: protein (red), sucrose (blue), ash (green), and extractives (maroon).

Ash

Ash, another category of component assessed by NIR, corresponds to mineral components and other non-hydrocarbon materials from the plant stem (Hartley, et al., 2011; Monti, et al., 2008). When QTL mapping was carried out,

two loci were identified (Table 13). On chromosome two, with a peak at 64.8 Mbp, a QTL was located that corresponds to the position of a QTL for days to anthesis (Fig. 39). At this locus, the allele from SC170 can increase ash percentage, and this locus accounts for 9.7% of the total variance observed. The second locus identified was located on chromosome 10, with a peak at 52.0 Mbp. This locus explains 8.9% of the total variance and the presence of the allele from M35-1 leads to an increase in ash percentage in this background.

Extractives

The extractives component category includes components of biomass that are soluble in water or ethanol (Vassilev, et al., 2012). When the relative percentage of extractives in the total stem biomass was used for QTL mapping, three loci resulted (Table 13). Two were located on chromosome one, with peaks at 13.1 Mbp and 49.9 Mbp (Fig. 39). The first of these accounts for 38.6% of the variance observed while the second accounts for 11.9% of the variance. The allele that contributes to increased percentage of extractives comes from SC170 and M35-1, respectively. The third QTL identified is located on chromosome three, with a peak at 50.8 Mbp. This locus accounts for 7.8% of the total variance observed and the allele from SC170 leads to increased extractive percentage in this population.

Discussion

QTL that modulate nearly 100 traits in the F5 generation of SC170*M35-1 *S. bicolor* plants were identified, making it possible to examine the patterns of

genetic regulation that modify macro traits and their component individual traits. QTL mapping of any single trait can be a useful way to investigate the genomic region(s) contributing to control of that trait; comparison of QTL maps for many traits at once yields even more useful information (Mace and Jordan, 2011). If QTL for more than one trait map to the same genomic region, it is possible that allelic variation in a gene within that QTL modulates the expression of several traits (Ming, et al., 2002). Two of the genomic regions that affect many traits correspond to Dw3 and Dw1, located on chromosomes seven and nine, respectively (Quinby and Karper, 1954). In this study, the most significant genomic regions for overlapping QTL are on chromosomes one, seven, eight, nine, and 10. On each of these chromosomes, there is at least one locus that modulates expression of a macro trait or traits, and additional QTL that affect specific component traits. QTL that modulate many traits help formulate informed hypotheses about the biochemical function of the gene or genes involved, as well as connections between physiological and morphometric traits.

Putative macro regulators

Stem length loci

Dw3

The Dw3 locus in *S. bicolor*, which has been previously shown to correspond to an ABCB auxin transporter, appears to be modulating several

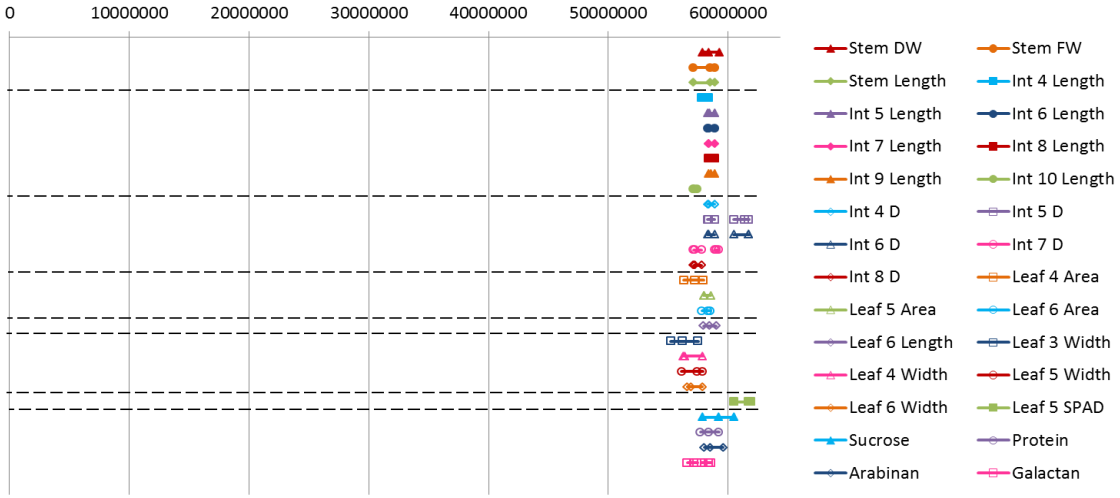

Figure 40: Combined QTL maps for chromosome 7 of SC170xM35-1 F5 generation. All QTL identified in this study on chromosome 7 are shown for comparison. Closed markers indicate increased phenotype values caused by the allele from M35-1 for that trait. Open markers indicate increased phenotype values caused by the allele from SC170 for that trait. The x-axis represents the entire length of chromosome 7. Dashed lines separate groups of traits. In descending order, trait groups are: macro traits, internode lengths, internode diameters, leaf areas, leaf lengths, leaf widths, leaf SPAD levels, biomass component percentages.

traits in this population (Fig. 40) (Multani, et al., 2003; George-Jaeggli, et al., 2011; Campbell, et al., 1975). The Dw3 locus is coincident with QTL for stem length, FW, and DW. Each internode between number four and number nine is modified by a QTL coincident with Dw3. The dominant Dw3 allele from M35-1 increases the length of internodes 4 – 9 as well as total stem FW and DW. In contrast, the Dw3 allele from M35-1 decreases the diameters of internodes four through eight, individual leaf areas, lengths, widths, and several stem biomass composition traits.

Given the existing knowledge of the identity of the gene corresponding to Dw3, the phenotypes observed can be explained. Plants with the dw3 recessive allele are shorter and have thicker stems because of the decreased flow of auxin to stem internodes during plant growth relative to plants with the Dw3 allele (Multani, et al., 2003). Since auxin is not being transported out of the apex in a normal way in plants that are dw3, auxin levels may increase at the apex and possibly cause an increase in stem diameter. For this reason, when it comes to the action of the Dw3 locus, having an allele that increases stem length could also cause plants to have narrower stems and smaller leaves compared to plants that have the recessive dw3 allele. The actions of the alleles at this locus demonstrate a trade-off between stem length and diameter. The implication of this finding is that, if Dw3 is used in breeding as a way of lengthening stems for increased biomass yield, some of those potential gains will be mitigated by decreases in stem diameter.

Dw1

Located on chromosome nine, the Dw1 locus is also contributing to variation in multiple traits in this population (Fig. 41) (Quinby and Karper, 1954). This locus appears to contribute to variation in both stem and leaf FW and DW, as well as total DW. For every QTL identified on chromosome nine, including those loci that are coincident with Dw1, it is the allele from M35-1 which contributes to increased trait values. This includes the aforementioned macro traits as well as individual internode lengths and individual leaf areas and

lengths. This set of QTL and the allelic contributions are different from Dw3. In this case, Dw1 contributes to increased biomass yield and stem length without having an apparent effect on stem diameter.

Dw1 may be having a larger effect on final biomass yield as compared to Dw3 because of the absence of an effect on stem diameter. While Dw3 appeared to regulate the balance between stem length increase and stem diameter increase, there is no such trade off in the action of Dw1. Rather, this locus contributes to increased stem length when the allele from M35-1 is present and no effect is seen on stem diameter.

This finding has great importance for future breeding of bioenergy hybrid *S. bicolor* genotypes. Often, Dw3 has been a target of breeding efforts due to the depth of knowledge that exists about the locus and the prevalence of Dw3 alleles in *S. bicolor* breeding germplasm. The findings presented here indicate that using Dw1 in breeding efforts will have a greater effect on biomass yield because there is no trade-off between a longer stem and a thicker stem. Whereas DW3 appears to contribute to the eventual lengthening of each cell within an internode, it is possible that the gene represented by the Dw1 locus

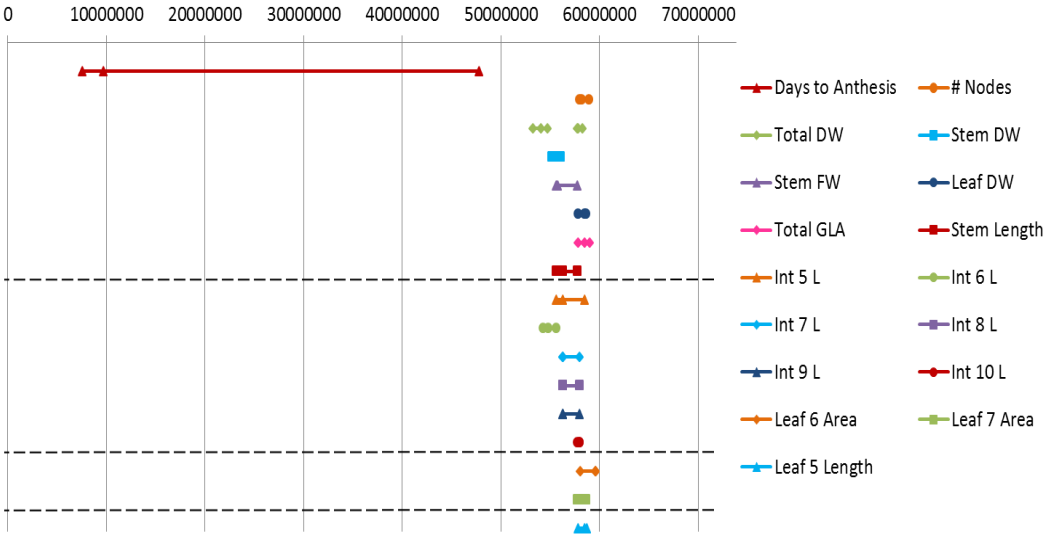

Figure 41: Combined QTL maps for chromosome 9 of SC170xM35-1 F5 generation. All QTL identified in this study on chromosome 9 are shown for comparison. Closed markers indicate increased phenotype values caused by the allele from M35-1 for that trait. Open markers indicate increased phenotype values caused by the allele from SC170 for that trait. The x-axis represents the entire length of chromosome 9. Dashed lines separate groups of traits. In descending order, trait groups are: macro traits, internode lengths, leaf areas, leaf lengths.

has some effect on determining the number of divisions that apical cells go through between the formation of each node. Future studies should include microscopic analysis of longitudinal stem sections to test this hypothesis.

Dwx

A new locus, Dwx, that regulates stem length, was identified in this population. Located on chromosome one, the allele from M35-1 contributes to increased stem length, stem FW and DW, and total DW in addition to increased individual internode lengths (Fig. 42). In contrast, the SC170 allele at this locus

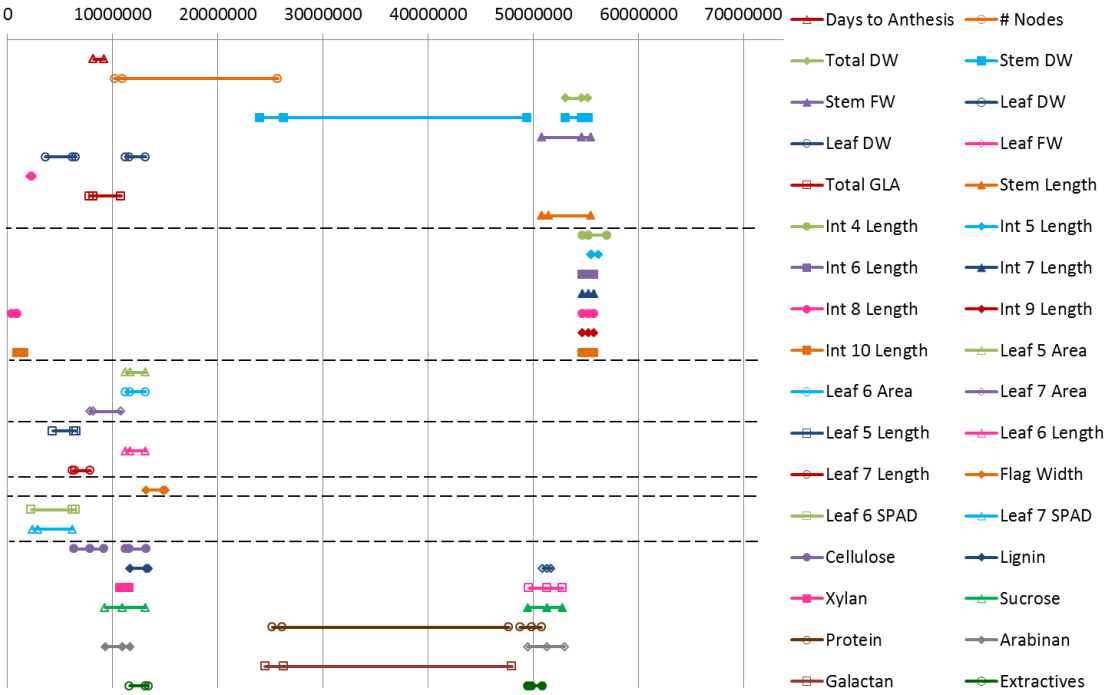

Figure 42: Combined QTL maps for chromosome 1 of SC170xM35-1 F5 generation. All QTL identified in this study on chromosome 1 are shown for comparison. Dwx is located from 51 to 58 Mbp on this chromosome. Closed markers indicate increased phenotype values caused by the allele from M35-1 for that trait. Open markers indicate increased phenotype values caused by the allele from SC170 for that trait. X-axis represents entire length of chromosome 1. Dashed lines separate groups of traits. In descending order, trait groups are: macro traits, internode lengths, leaf areas, leaf lengths, leaf widths, leaf SPAD levels, biomass component percentages.

contributes to increased lignin, xylan, and arabinan content. The action of this locus contributes to increased biomass yield and stem length with a linked decrease in secondary cell wall components. This finding has led to the hypothesis that this locus contains a gene that is responsible for secondary cell wall deposition in the stem. According to this idea, the allele that contributes to

longer internodes and lower secondary cell wall component percentages comes from M35-1 and would likely encode a gene that slows or represses the formation of secondary cell wall components. This would explain why plants with the allele from M35-1 are taller and have increased sucrose content but decreased cell wall component content. Likewise, plants with the allele from SC170 are shorter and exhibit less sucrose content but a greater content of secondary cell wall components.

Within the region bounded by the DWx locus, there are multiple transcription factors, including a NOT1 transcription factor, as well as a large group of beta-expansin precursor genes. These are prime candidate genes for this locus based on the QTL results obtained. Future research on this locus would include further fine mapping, sequencing, and expression studies on genes within the locus that have the potential to have effects on cell wall deposition, cell wall precursor synthesis, and/or expression of genes relating to cell walls and stem lengthening.

Biomass yield loci

Multiple loci were identified that modulate total biomass yield, including Dw1, Dwx, and a QTL on chromosome ten which does not appear to be related to stem length (Fig. 43). While the likely contributions of Dw1 and Dwx to eventual biomass yield have already been discussed, the locus on chromosome ten requires further analysis. This QTL does not influence the length or diameter of the stems, and the only stem biomass component affected is ash. Multiple

leaf parameters are modulated by this locus. Individual leaf areas and lengths are affected as well as a small number of individual leaf widths. For all of these QTL, the allele from SC170 is contributing to increased trait values. This means that the allele from SC170 may contribute to increased total DW through increased leaf size and photosynthetic capacity.

There are many genes within this locus on chromosome ten, any one of which may be contributing to eventual biomass yield through leaf size. Future experiments should begin with fine mapping of this locus. Fine mapping will reduce the size of this already small QTL even further, which will decrease the number of possible candidate genes as well. Following fine mapping, it may be possible to move straight to sequencing of likely candidate genes and even expression studies.

Maturity, node number, and LAR loci

One of the most interesting loci for days to anthesis that was identified in this experiment is located on chromosome one. A putative flowering-time control gene, phytochromeA (PhyA), is known to exist within this QTL (Fig. 42). The product of this gene is a light-sensing protein that modulates a large number of plant responses, including transmission of light signals from sunlight to the internal plant molecular clock (Osugi, et al., 2011). In this case, SC170 has an allele of PhyA which leads to an increase in the number of days to flowering. This locus and allele appear to contribute to greater leaf area as well. Sequencing of the PhyA gene from both SC170 and M35-1 would help

determine whether this candidate gene is the actual gene responsible for the variation attributed to the QTL on chromosome one.

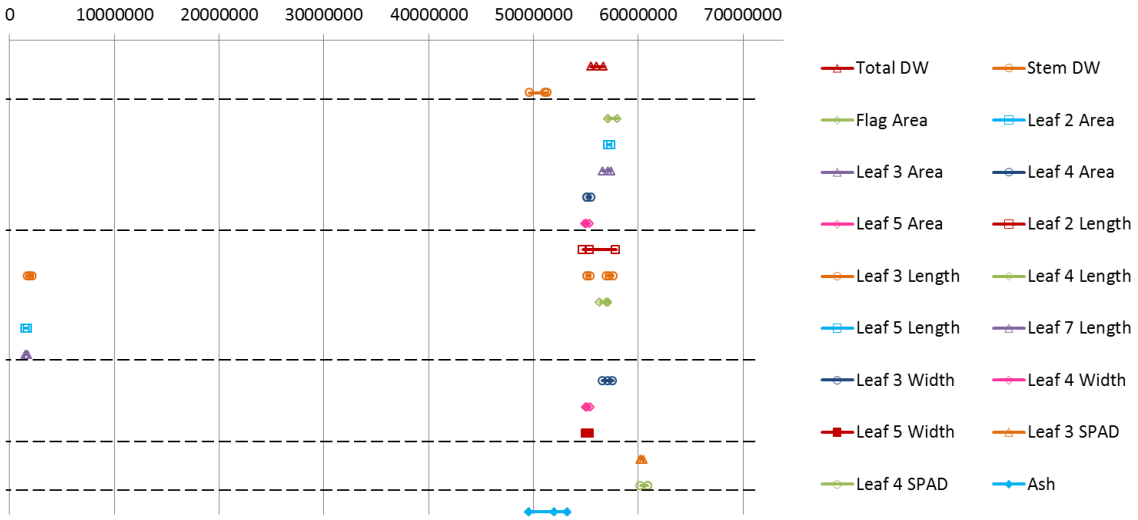

Figure 43: Combined QTL maps for chromosome 10 of SC170xM35-1 F5 generation. All QTL identified in this study on chromosome 10 are shown for comparison. Closed markers indicate increased phenotype values caused by the allele from M35-1 for that trait. Open markers indicate increased phenotype values caused by the allele from SC170 for that trait. X-axis represents entire length of chromosome 10. Dashed lines separate groups of traits. In descending order, trait groups are: macro traits, leaf areas, leaf lengths, leaf widths, leaf SPAD levels, biomass component percentages.

Additional macro traits can provide useful data on plant growth and development such as the number of stem nodes and LAR. There is only one location in the genome where control of both LAR and the number of nodes per plant are colocated. This is on chromosome eight, in a position where leaf FW and DW, total GLA, upper internode diameters, and individual leaf areas have QTL as well (Fig. 44). At this position, an allele from M35-1 contributes to a

greater number of days per leaf produced while the allele from SC170 contributes to an increased value for every QTL that maps to this location. The increased stem diameters may be explained by the greater number of nodes, because if more nodes are being made than others in a given period of time, there is necessarily less time for elongation of those nodes and a thicker node diameter may result. The thicker nodes may in turn contribute to increased individual leaf areas. Each leaf is connected to the stem via a leaf sheath, which emerges from the stem at a node. A larger nodal diameter gives greater circumference for the leaf sheath to emerge from and hence there is a wider span of tissue for the leaf blades to emerge from.

Future directions

Taken together, the QTL identified in this chapter provide a comprehensive description of the genetic control underlying variation in a large number of traits that are segregating in this population. SC170 and M35-1 are highly diverse genotypes of *S. bicolor* and the population created by crossing these two segregates for every trait that was measured.

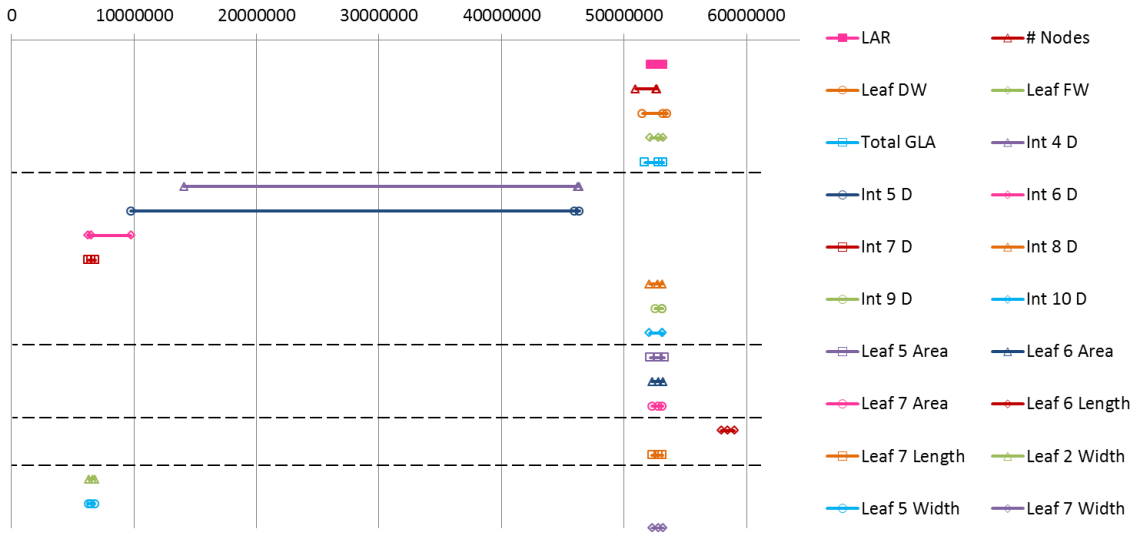

Figure 44: Combined QTL maps for chromosome 8 of SC170xM35-1 F5 generation. All QTL identified in this study on chromosome 8 are shown for comparison. Closed markers indicate increased phenotype values caused by the allele from M35-1 for that trait. Open markers indicate increased phenotype values caused by the allele from SC170 for that trait. X-axis represents entire length of chromosome 8. Dashed lines separate groups of traits. In descending order, trait groups are: macro traits, internode diameters, leaf areas, leaf lengths, leaf widths.

Future work with this population should advance the findings of the work presented here in multiple ways. First, it will be important to assess the leaf tissue of this population for biomass composition traits. These data will be useful in comparisons with the stem tissue composition data. Additionally, multiple candidate genes have been postulated in this chapter that could and should be verified. Fine mapping, coupled with expression studies and direct sequencing, will be invaluable in determining whether candidate genes identified herein do in fact contribute to measured phenotypic variation. Following candidate gene verification, such genes can be used in targeted breeding

programs to increase the potential biomass yield of *S. bicolor* for use in generation of biomass for biofuels.

Materials and methods

Parental line selection and population construction

The two parental lines used to construct this mapping population are SC170 and Maldandi 35-1 (M35-1). These are highly divergent *S. bicolor* genotypes. SC170 is a zerazera variety from the caudatum race of *S. bicolor* and was first identified growing in well-watered lowland river deltas in Ethiopia (Hart, et al., 2001). This variety is easily recognizable by its short stature and thick stem, extremely wide, large leaves, and very long, compact panicles. M35-1 comes from India, where it is grown as a commercial grain type, used in the jowar (Winter) growing season. This variety exhibits tall, slender stems with short, narrow leaves and large panicles with long, expanded branches (Starks and Doggett, 1970).

The population used here was created by making a cross that used SC170 as the female parent and pollen from M35-1. The F1s were verified and all F2 seed generated by the confirmed F1 plants was grown out in the field in Weslaco, Texas. These plants were self-pollinated and the resultant F3 seed was planted in greenhouses in College Station, Texas. F3 and F4 plants were grown and advanced using single-seed descent in pots in the greenhouses to generate F5 seed. This F5 seed was planted in the field in College Station, Texas in the spring of 2011.

Field plots, planting, and growth conditions

Prior to field planting, germination assays were carried out for each F5. Ten seed were placed on moistened germination paper and allowed to germinate for four days in a growth chamber (30^{o}C day/22^{o}C night, 16 h day) with constant moisture. The proportion of seeds that germinated in that time is the germination rate. Based on these values, the number of seed used for planting each F5 family was adjusted to lead to 100 effective plants in the field. For example, an F5 family with 0.8 germination rate would need 125 seeds planted to yield 100 viable plants. This way, an even, robust stand of plants was guaranteed in the field. Just prior to planting, seeds were treated with a thin coating of acrylic paint as a protection against pre-emergent herbicides in use in the field soil.

There is considerable variation in height within this population, so shading between plots was a concern. To mitigate the effects of intra-population shading, the F5 families were planted in order of increasing height of the F4 plant that was the progenitor to that F5 family. The field plot was 16 rows by eight ranges, located in field 214 of the Texas Agricultural Experiment Station farm located on Hwy 60. There was a border row of RTx430 planted on each side of this population and one plot each of SC170 and M35-1 was planted within this plot.

Planting of the plots occurred on March 23, 2011. At approximately 14 days after emergence (DAE), when seedlings were under 20 cm tall, thinning of

plots was carried out. The plots were thinned to a planting density of 10 plants per meter (one plant every 10 cm). Following planting, the field was irrigated at regular intervals except in cases where precipitation provided sufficient water.

Plant sampling technique

Sampling of plants was carried out over a series of harvests from June 2-22 of 2011. Each family was sampled within five days of anthesis. For these purposes, the day that at least half of the plants in a plot had half pollen shed is determined to be the day of anthesis for that family. For each F5 family, five representative plants were selected for sampling. Loppers were used to cut each plant out of the ground below the soil level. This method maximizes the likelihood of recovering the first internodes, which are often below soil level.

Once plants were cut out of the ground, they were transported from the field to the campus of Texas A&M University, where various morphometric measurements were taken immediately. Care was taken to minimize the time between harvesting of plants and making measurements in order to decrease the likelihood of any of those measurements changing as a result of the plants having been cut away from their roots.

Each plant was processed and measured separately. First, a sample of leaf tissue was taken for any necessary subsequent DNA analysis. Then, leaf blades were removed, numbered in descending order from the flag leaf, and placed into a cup containing water. This practice keeps leaves from drying out or curling up prior to assessment of leaf area and leaf mass measurements.

Then, the leaf sheaths were stripped from the stems and collected separately. The immature panicle was then removed from the stem for measurement. The stem was then measured as will be described later. Following morphometric measurements, all leaf, leaf sheath, panicle, and stem tissue was dried in an oven at 160^{o}C with continuous air flow for three days before additional measurements and analyses were performed.

Morphometric measurements

Leaf blades were measured using a LI-300C planimetric leaf area meter (Li-COR). This device records individual leaf area, length, and width. SPAD readings were performed using a SPAD-502 Chlorophyll meter (Spectrum Technologies, Inc.). Three measurements were taken for each leaf and the average of these three is reported as the SPAD reading for that leaf. Then, the total biomass of the leaf tissue (FW) was measured and recorded. Following the drying technique described above, the leaf DW was also measured and recorded.

Leaf sheath tissue was weighed (FW) and then dried and measured again for DW. The length and FW of the immature panicle were recorded. Following drying, the DW of the immature panicle was recorded as well. The QTL maps produced for these traits did not show any control in common with the other traits described, so these results have been omitted for brevity.

The Stem was measured in multiple ways. The first internode was identified based on visual inspection of the base of the stem. Then, each

internode was measured for both length and diameter. Length was measured using a tape measure and recorded in cm. Diameter was measured using micro-calipers and was recorded in mm. The total number of nodes was counted as well. Then, the whole stem was weighed and FW was recorded. Following drying, stem DW was also measured and recorded.

Composition analysis

The composition of stem tissue was assessed in this population. Following drying of the stem, stem tissue was ground using a Wiley rotary mill to a uniform particle size of 0.2 mm. Then, ground tissue was analyzed using near infrared spectroscopy (NIR). This was carried out using an XDS-rapid contend analyzer solids module (Foss). Subsequent analysis of NIR spectra was performed using a standard curve developed by Dr. Ed Wolfrum at the National Renewable Energy Laboratory (NREL). The composition is reported in terms of relative percentages of each component measured.

Genotyping

For each F5 family in this study, DNA was extracted for genotyping using a FastDNA Spin Kit and FastPrep apparatus (MP Biomedicals). The tissue used for each extraction was a sample of 5-10 seedlings for each family. Resulting DNA samples were quantitated in triplicate using the Qubit fluorometer and DNA-BR standards kit (Invitrogen). Only samples with concentrations between 80 and 120 ng/µL were considered acceptable for continued processing.

The remainder of sample preparation was carried out according to Morigishige et al., 2012 (in prep). Prepared DNA was then analyzed using an Illumina $GAII_x$ instrument (Illumina). Generation of this genotype data was carried out by the Laboratory for Plant Genome Technologies (LPGT) at Texas A&M University. This sequencing platform generates large quantities of short sequences (72 bp), which were used to create a haplotype map of the entire genome for each F5 family. These haplotypes are based in single nucleotide polymorphisms (SNPs) and short insertions or deletions (InDels). Any sequence polymorphism identified between the two parents, SC170 and M35-1, is considered a locus or marker. Each F5 is then analyzed to determine the parental identity at each of these markers. Analysis of genotype data for this population was carried out by Dr. Patricia Klein at Texas A&M University. The final output of this genotyping process is a haplotype map which indicates identity by descent to one parent or the other for each F5 family at every possible locus across the genome. This information is then used in subsequent genetic map construction and QTL mapping.

Genetic map construction

Genetic maps for each chromosome were created based on the genotype data collected. A freely available piece of software, MapMaker 3.0B, was used to construct these genetic maps. This program uses user-input data about the generation level of the population as well as genotype information to determine the genetic distance between each pair of markers in centimorgans (cM).

Genetic distances are based on the number of crossing-over events (or parental genotype changes) that occur within a population between a pair of loci. The calculation of genetic distance is based on statistics and is of great importance for use in the following step of this process.

QTL mapping

Once a genetic map was made for each chromosome, and the phenotypes of interest were all measured, quantitative trait locus (QTL) maps were generated. This is a process which combines user-input genotype and phenotype data to identify QTL throughout the genome on a trait-by-trait basis. The free software WinQTLCartographer (v2.5.010) was used to generate QTL maps. The composite interval mapping module with standard parameters was used.

Following generation of QTL maps, random permutations were carried out (1000 iterations) to determine the threshold LOD score for each trait. For traits where permutation tests were not done, the threshold was assumed to be 3.0. The positions of QTL are reported in terms of genetic position, or cM. However, in order to make these results more easily relatable to QTL mapping studies done on other populations, QTL positions were converted from genetic to physical positions manually. In cases where the genetic position of the start, peak, or end of a QTL was located between two physically located markers, the physical position was estimated based on the distance between each of the markers and the reported genetic position.

Genome analysis

Once QTL were identified for a trait in the SC170 x M35-1 F5 population, potential candidate genes were identified based on the annotated genome sequence (Goodstein, et al., 2012). The genome browser and the BioMart application that are accessible through the Phytozome website were used to generate a list of potential candidate genes from all of the annotated genes located within the genomic regions associated with QTL for the traits assessed in this study (Guberman, et al., 2011).

CHAPTER V

FLOWERING IN *SORGHUM BICOLOR*: LOCATING AND IDENTIFYING MA2

Background and introduction

Control of flowering time: relevance to biomass generation

Regulation of flowering time has been established to be an important consideration for development of energy crops (Rooney, et al., 2007; Rooney and Aydin, 1999). In plants like *Sorghum bicolor*, flowering is a terminal state that is accompanied by a cessation of vegetative growth and biomass accumulation (Smith and Frederiksen, 2000; Gerik, et al., 2003). Therefore, maximal biomass accumulation can be achieved in part by maximizing the duration of vegetative growth through delaying initiation of flowering.

Sorghum bicolor *is induced by short day conditions*

The initiation of flowering in *S. bicolor* is regulated by a complex network of signals, both internal and external, which allows the plant to flower at the precisely correct time based on the plant's own age, its nutrition status, and the environment in which it is growing (Blázquez, 2000). Initiation of flowering occurs at the end of the vegetative phase of the plant, when this group of signals comes together to induce the plant's vegetative shoot apical meristem to transition into a floral shoot apical meristem. Photoreceptors communicate the light conditions and photoperiod to the plant internal clock, which tracks the circadian rhythm of the plant (Murphy, et al., 2011). Sucrose levels within the

plant communicate the nutrition status of the plant, and meristem identity genes like LEAFY (LFY) physically communicate the signal for transition into the meristem (Childs, et al., 1997). *S. bicolor* is induced to flower by short-day conditions. The homologs of *Arabidopsis thaliana* genes *LHY* and *TOC1* in *S. bicolor* make up the core oscillator of the circadian clock (Foster, et al., 1994). The afternoon peak in this oscillator, when coincident with the absence of light sensed by photoreceptors, provides the signal that short day conditions are occurring (Murphy, et al., 2011).

The maturity loci in Sorghum bicolor

In *S. bicolor*, four original maturity loci were identified that had an effect on the number of days to anthesis for *S. bicolor* plants (Quinby and Karper, 1945; Quinby, 1966). These loci are named *Ma1*, *Ma2*, *Ma3*, and *Ma4*, and at each locus late flowering is dominant to early flowering (Quinby, 1972). Further research into photoperiod sensitive genotypes of *S. bicolor*, which are relevant to development of high-biomass varieties of *S. bicolor*, identified two additional maturity loci, *Ma5* and *Ma6*, which can contribute to a photoperiod sensitive response in *S. bicolor* (Rooney and Aydin, 1999). Each of these loci, when dominant, contributes to late flowering for plants grown in long day conditions.

Ma1

The strongest effect on flowering time of all of the known maturity loci comes from Ma1 (Quinby and Karper, 1945). This locus is known to have some degree of interaction with the locus Ma2 (Quinby, 1972). Recent work has

identified *PSEUDORESPONSE REGULATOR 37* (*PRR37*) as the gene corresponding to the *Ma1* locus (Murphy, et al., 2011). This gene encodes a protein which represses *FLOWERING LOCUS T* (*FT*), and therefore flowering, in *S. bicolor* grown in long day conditions.

Ma3

The gene that corresponds to the maturity locus *Ma3* has been identified as *PHYTOCHROME B* (*PHYB*) (Childs, et al., 1997). This was the first of the *S. bicolor* maturity loci to be definitively identified. The experiments which identified *PHYB* as the gene that corresponds to *Ma3* included cloning and investigation of other phytochrome genes (*PHYTOCHROME A* and *PHYTOCHROME C*). Neither of the other light-sensing phytochrome genes could be mapped to coincide with the action of *Ma3*, which determined conclusively that the action of *Ma3* is due to the action of *PHYB* only (Foster, et al., 1994; Childs, et al., 1997).

Ma4

The gene that corresponds to *Ma4* has not yet been identified. This locus was first identified in the *S. bicolor* genotype Hegari, which will be used in experiments presented in this chapter. The *Ma4* locus was first identified by Quinby (1966), at which time it was hypothesized to be influenced to some degree by high temperature growing conditions.

Ma5/Ma6

These two loci were identified later than the original four maturity loci in *S. bicolor*. It was not until 1999, as photoperiod sensitive, long-growing, high-

biomass hybrid genotypes of *S. bicolor* were being bred, that the *Ma5* and *Ma6* loci were identified (Rooney and Aydin, 1999). These two loci are not linked, and as yet the relationships between *Ma5*/*Ma6* and *Ma1* through *Ma4* are not well-understood. Further characterization of the *Ma6* locus is described in brief later in this chapter and will be fully described in a later publication.

Ma2

Prior to the experiments described in this dissertation, there was no genomic location ascribed to *Ma2*, though it was known that *Ma2* contributed to delayed anthesis in a dominant fashion (Quinby and Karper, 1945). This chapter will elucidate the physical position of the gene that corresponds to the Ma2 locus in the genome of *S. bicolor*. This will be accomplished through the generation of multiple populations by genetic mating, quantitative trait locus (QTL) mapping, and sequencing.

Results

80M x 100M segregates for Ma2 only

When the F2 generation of 80M x 100M was grown out in the field in College Station, Texas, the phenotypes for days to anthesis were clearly split into only two groups. There was a clear separation between the early-flowering cohort of plants and the late-flowering cohort (Fig. 45). There is a span of 14 days between anthesis of the last member of the early cohort and anthesis of the first member of the late cohort. The ratio of early- to late-flowering F2 plants

in this population was ~1:3 (358 early: 900 late), which is consistent with the maturity-related genotypes of the progenitors of this population (Quinby, 1972).

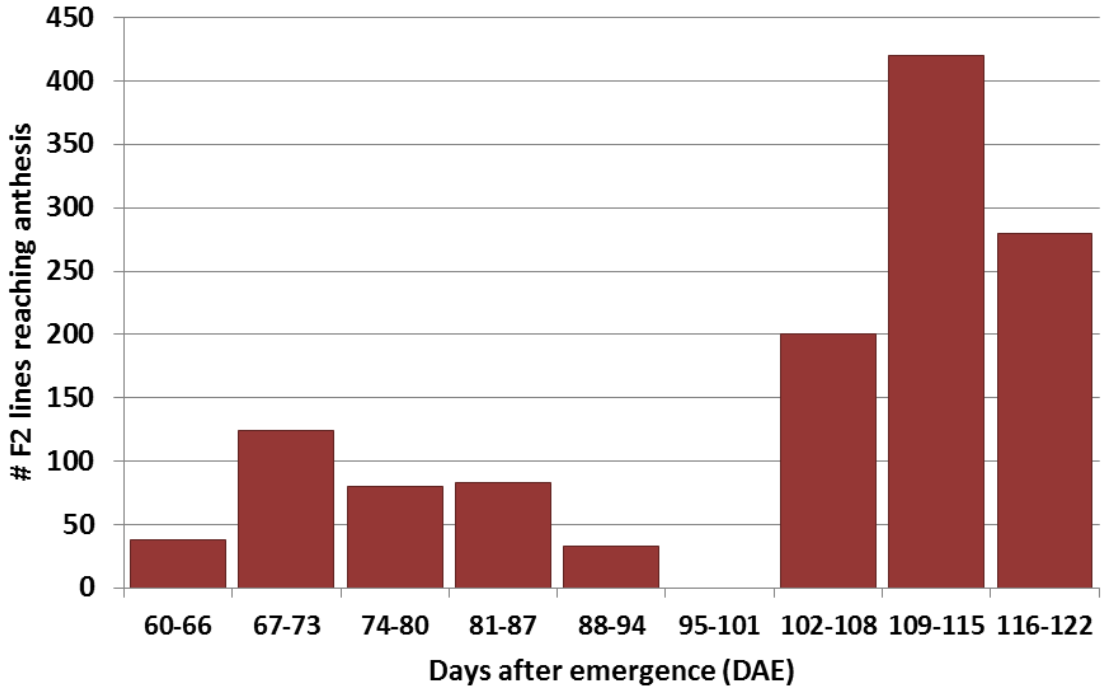

Figure 45: Distribution of days to anthesis for 80Mx100M F2 generation. Plants were grown in the field in College Station, Texas in the summer of 2007. All plants in a bulk plot of F2s were assessed for days to anthesis. Each bin of the histogram represents one week, beginning with the first day that any F2 was observed to reach anthesis. There is a 14-day gap in days to anthesis between the last F2 in the early cohort and the first F2 in the late cohort. Furthermore, the ratio of early flowering to late flowering F2 plants is almost exactly 1:3, indicating a single gene segregating for this trait.

In order to begin the QTL mapping process in the 80M x 100M F2 population, the parental lines were genotyped using digital genotyping. For every location across the genome, this method determines whether there is a sequence polymorphism (SNP or small InDel) between the parental genome

sequences. Each polymorphism discovered is referred to as a marker and can be used in QTL mapping with the offspring of the population. 80M and 100M do not have a large number of polymorphic markers between them (Fig. 46). The majority of the identified polymorphisms are concentrated on the end of chromosome one, while all the other chromosomes have five or fewer polymorphic markers across the entire chromosome.

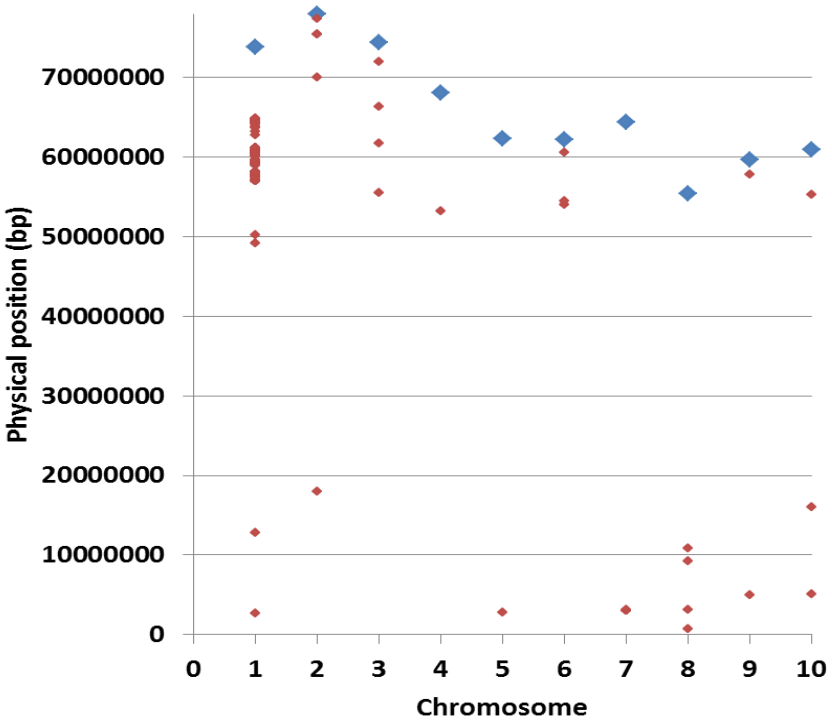

Figure 46: Distribution of genetic polymorphisms between 80M and 100M. Blue markers represent ends of chromosomes, based on chromosome lengths contained in the Phytozome genomic database. Each maroon marker represents the physical position of a polymorphism identified between 80M and 100M using digital genotyping analysis (DGA). While there are many polymorphic markers concentrated on chromosome 1, most chromosomes have only a small number of polymorphic markers, so mapping in this background will not be worthwhile.

Based on these results it was determined that 80M and 100M are too genetically similar to be appropriate for use in QTL mapping. No further genetic analyses were carried out on this population. Rather, it was determined that Ma2 would not be mapped successfully by using parental genotypes with extremely similar genomes like 80M and 100M.

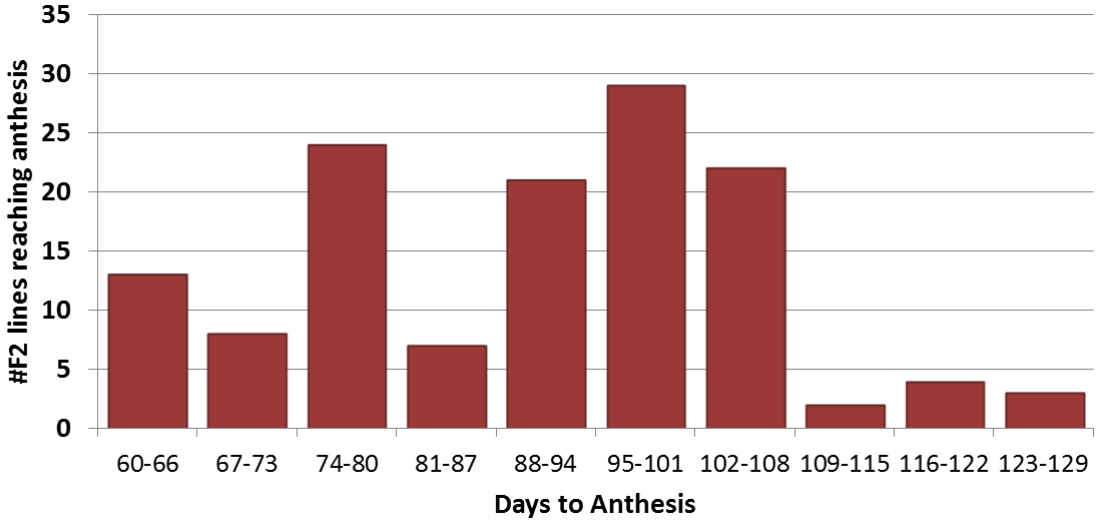

Figure 47: Distribution of days to anthesis for 80MxR.07007 F2 generation. Plants were grown in long day greenhouse conditions in 2009. There is no clear separation between the early and late flowering cohorts. Rather, multiple overlapping cohorts suggest multiple genes are underlying this variation in days to anthesis.

R.07007 has a recessive ma2 allele

A second attempt to map Ma2 was carried out using a population constructed from very different parental genotypes. 80M and R.07007 are two genotypes which come from different geographical origins and have dissimilar

morphology when grown (Quinby, 1972; Rooney and Aydin, 1999). 80M is an early-flowering genotype of *S. bicolor* which is characterized by short main stem growth (Quinby and Karper, 1945). R.07007 is a tall, highly tillering, late-flowering genotype of *S. bicolor* (Rooney and Aydin, 1999). Because of its late-flowering phenotype, R.07007 was hypothesized to express the dominant Ma2 allele. A genetic mating of 80M x R.07007 was carried out and the F2 generation was used to attempt to map Ma2.

When grown in a greenhouse, the 80M x R.07007 F2 plants flowered in 60 – 131 days (Fig. 47). Unlike the 80M x 100M F2 population, there was not a clear separation between early- and late-flowering plants in this case. There appeared to be multiple cohorts of plants flowering in multiple intervals, which suggests that there are multiple genes for days to anthesis that are segregating in this population.

Once the phenotypes were gathered for the F2s, genotyping and QTL mapping were carried out for days to anthesis in the 80M x R.07007 F2 population. The comparative analysis of the genomes of 80M and R.07007 was carried out prior to QTL mapping (Fig. 48). While there is adequate marker coverage over most of the genome, it was not complete. The end of chromosome two is completely devoid of polymorphisms and there are only three markers identified to be polymorphic on chromosome nine. If the gene corresponding to Ma2 is located in one of these genomic regions, it will not be identified in this population.

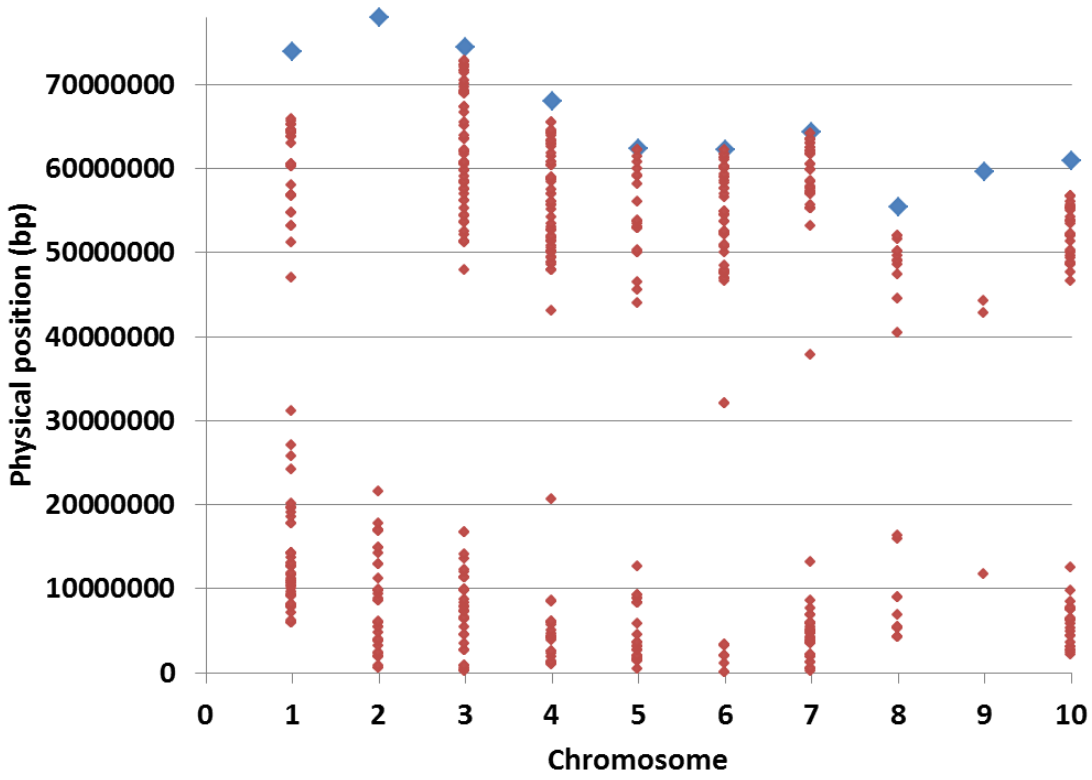

Figure 48: Distribution of genetic polymorphisms between 80M and R.07007. Blue markers represent ends of chromosomes, based on chromosome lengths contained in the Phytozome genomic database. Each maroon marker represents the physical position of a polymorphism identified between 80M and R.07007 using digital genotyping analysis (DGA). While there is good coverage over the majority of the genome, the end of chromosome two is completely devoid of polymorphisms and there are only three polymorphic markers on chromosome 9.

In spite of the small genomic regions of similarity between the parental genotypes, QTL mapping was carried out. Three QTL were identified (Fig. 49 and Table 14). The locus with the greatest effect on days to anthesis is located on chromosome six. A second locus, also located on chromosome six, contributes the second-strongest effect on days to anthesis. The third locus,

with the smallest effect of the three, is located on chromosome one. For both of the QTL on chromosome six, the allele from R.07007 contributes to an increase in the number of days to anthesis. For the QTL on chromosome one, the allele from 80M contributes to an increase in the number of days to anthesis. In total,

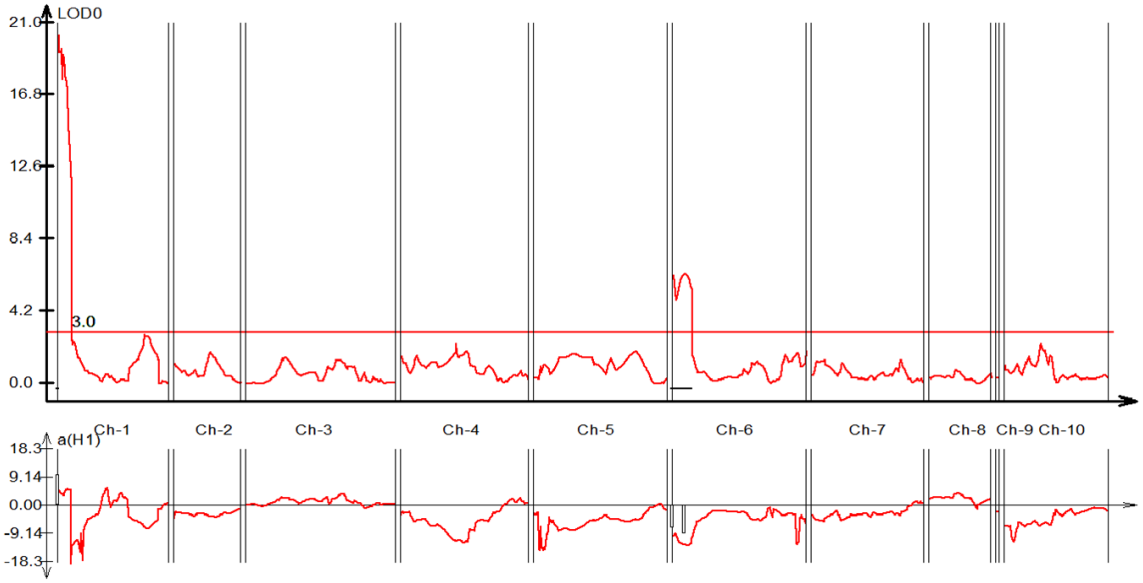

Figure 49: Mapping of QTL for days to anthesis in 80MxR.07007 F2 generation. It had been hypothesized that R.07007 would be Ma2, making it an ideal candidate for QTL mapping of that locus when crossed with 80M. However, no QTL for Ma2 was identified in this background. Rather, multiple other QTL were identified, including the maturity locus *Ma6*, located on chromosome six.

these three loci explain 50.1% of the total variation in days to anthesis in the 80M x R.07007 F2 generation.

Parallel experiments in QTL mapping of days to anthesis in other populations of *S. bicolor* elucidated the same QTL on chromosome six that was

found in this study. In the other populations, Ma2 is known not to be segregating, so this locus on chromosome six was eliminated as a potential QTL corresponding to the action of Ma2. Since the QTL identified on chromosome one contributes to an increase in days to anthesis in the presence of the allele from 80M, this locus cannot correspond to Ma2 either.

Table 14: Position and R^2 for each QTL identified for days to anthesis in 80MxR.07007 F2 population. The QTL identified on chromosome one is aligned with one of the QTL for days to anthesis that was identified in the SC170xM35-1 F5 background. Of the two QTL located on chromosome 6, the latter has been identified as Ma6, while the former is as yet unidentified.

Days to Anthesis			Peak		bp	
Chr.	R^2	Add Var	bp	cM	left	right
1	5%	9.6	5,948,243	0.01	0	6,500,000
6	19%	-7.3	12,923	0.01	0	100,000
6	26%	-9.5	800,000	10.8	200,000	1,100,000

While none of these three QTL corresponds to Ma2, this result is not without impact. The locus on chromosome six with the greatest effect on days to anthesis in this population has been identified as *Ma6* (data unpublished). The recessive ma6 allele, which decreases days to anthesis, is present in 80M. The Ma6 locus has since been identified in additional mapping populations of *S. bicolor*. Comprehensive characterization of this locus will be described in a future publication.

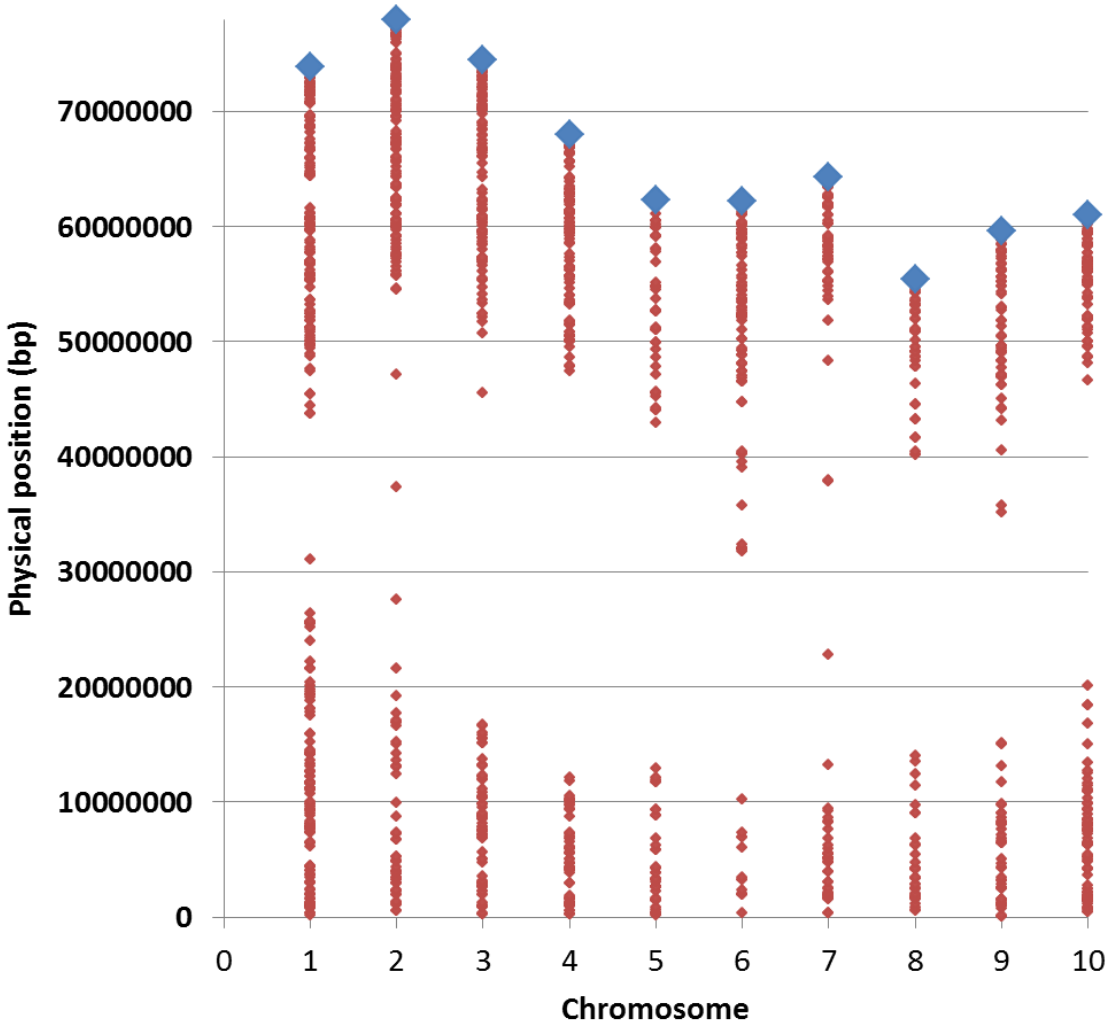

Figure 50: Distribution of genetic polymorphisms between Hegari and 80M. Blue markers represent ends of chromosomes, based on chromosome lengths contained in the Phytozome genomic database. Each maroon marker represents the physical position of a polymorphism identified between Hegari and 80M using digital genotyping analysis (DGA). These genotypes were compared to ensure a high degree of polymorphism spanning the entire genome before this cross was made. Such a practice leads to better odds of successfully mapping QTL in the offspring.

Whole genome scanning identifies ideal genotypes for mapping Ma2

An alternative approach to mapping Ma2 was used following the experiments with both 80M x 100M and 80M x R.07007. In this third case, the genome sequences of parental genotypes known to express the dominant Ma2 allele were compared to the genomic sequence of 80M. This protocol identified polymorphisms that existed between each dominant Ma2 parental genotype and 80M. This allowed for generation of a marker distribution map prior to genetic mating or advancing of a population. One such comparison, between 80M and Hegari, revealed that a distribution of polymorphic markers spanned each of the ten chromosomes (Fig. 50). Hegari expresses a recessive ma4 allele, but is dominant for *Ma1*, *Ma2*, and *Ma3* (Quinby, 1966). This means that only two maturity loci, *Ma2* and *Ma4*, were expected to segregate in a background of Hegari and 80M. Based on this result, a genetic mating was carried out between Hegari and 80M by Daryl Morishige and the resulting F2 plants were used for QTL mapping of days to anthesis.

Hegari x 80M F2s segregate for flowering time

Over 400 F2 plants were grown from the Hegari x 80M population. There is no clear separation in number of days to anthesis between the early- and late-flowering cohorts of plants (Fig. 51). There appear to be multiple overlapping peaks in the distribution of days to anthesis for this population, which is consistent with the anticipated maturity genotypes of both parental genotypes. Both Ma2 and Ma4 were expected to be segregating in this population. The

flowering phenotypes of the Hegari x 80M F2 plants were collected by growing the plants in a greenhouse under 14 hour days, tissue was collected for DNA extraction, and QTL mapping was carried out.

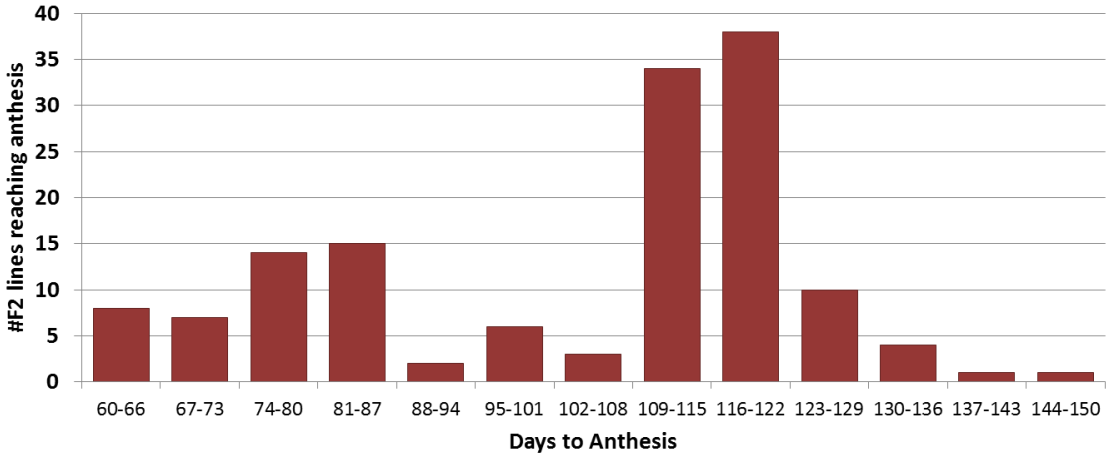

Figure 51: Distribution of days to anthesis for Hegarix80M F2 generation. Plants were grown in long day greenhouse conditions in 2011. There is no clear separation between the early and late flowering cohorts. Rather, multiple overlapping cohorts suggest multiple genes may be underlying this variation in days to anthesis. Based on known haplotypes of 80M and Hegari, both Ma2 and Ma4 should be segregating.

Ma2 is located on chromosome two

The QTL map of days to anthesis in the Hegari x 80M F2 population identified three loci that are controlling this phenotype (Fig. 52 and Table 15). For two of these loci, located on chromosomes two and nine, the allele from Hegari contributes to an increase in days to anthesis. The reverse is true of the third QTL; the presence of the allele from 80M contributes in an increase in days

to anthesis. The QTL on chromosome two explains over 40% of the total variance for days to anthesis in the Hegari x 80M F2 population.

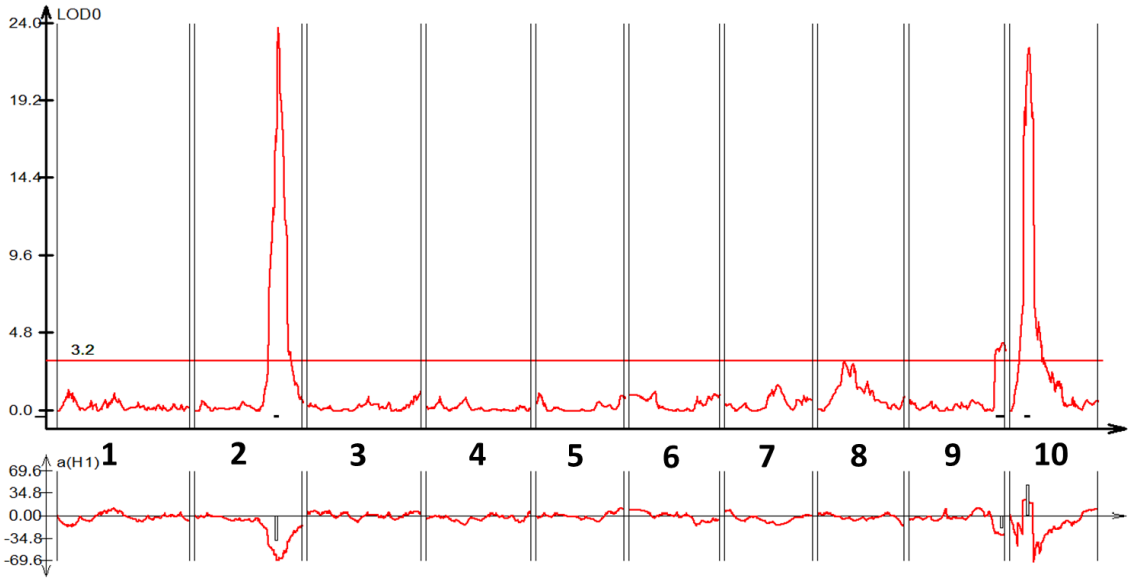

Figure 52: Mapping of QTL for days to anthesis in Hegarix80M F2 generation. Three loci are identified in this background. The strongest effect, explaining 43.8% of the total variance observed, is located on chromosome two. This is hypothesized to be the genomic location of Ma2. The other two loci have much less phenotypic effect. The locus on chromosome nine is aligned with the QTL identified in chapter IV for number of nodes, total green leaf area, etc. The QTL on chromosome 10 is hypothesized to represent the historic flowering time locus Ma4.

Based on the allelic contribution by Hegari to delayed flowering at the locus on chromosome two, it can be concluded that this QTL corresponds to the maturity locus Ma2. The dominant Ma2 allele from Hegari at this locus contributes to an increase in the number of days to anthesis, while the recessive ma2 allele from 80M contributes to a decrease in the number of days to anthesis in this population. Conversely, the locus on chromosome 10 corresponds to

Ma4 based on the contribution of the dominant Ma4 allele from 80M to an increase in the number of days to anthesis. The QTL on chromosome nine has not been identified with respect to known maturity loci.

Table 15: Position and R^2 for each QTL identified for days to anthesis in Hegarix80M F2 population. The locus on chromosome 2 has been identified as Ma2, while the locus on chromosome 10 is hypothesized to be Ma4. The remaining locus, located on chromosome 9, has not yet been identified.

Days to Anthesis			Peak		bp range	
Chr.	R^2	Add Var	bp	cM	left	right
2	44%	-40.6	68,500,000	111	67,720,252	69,585,439
9	8%	-21.2	58,819,358	126	57,956,650	59,034,562
10	6%	47.6	4,200,000	24.5	3,607,821	5,029,546

Discussion

Control of flowering time is a critical component of energy crop design and improvement. This is especially true of annually-grown crops like *S. bicolor* where flowering is a terminal state that includes cessation of vegetative growth and biomass accumulation. Delayed flowering will be an important attribute of any ideal energy crop. Identification of the location of Ma2 provides the data necessary to achieve better control of days to flowering in *S. bicolor*. With this knowledge, it will now be possible to conclusively identify the Ma2 allele (dominant or recessive) expressed by any parental genotype using molecular techniques.

R.07007 expresses a recessive ma2 allele

Based on the QTL mapping results, it can now be concluded that R.07007 contains a recessive ma2 allele. The position of *Ma2* is known to be near the end of chromosome two. In the mapping of the 80M x R.07007 F2 population, that entire arm of chromosome two is devoid of polymorphic markers, including the region that contains the Ma2 locus. While R.07007 was hypothesized to encode a dominant allele of *Ma2*, these results demonstrate that R.07007 actually expresses a recessive ma2 allele, obviating the possibility of mapping *Ma2* in the 80M x R.07007 F2 population. Instead, a different maturity locus, *Ma6*, was identified in this population.

Cop9 and myb are candidate genes for Ma2

The genomic region that corresponds to *Ma2* was examined using Genome Browser interface (Goodstein, et al., 2012). Approximately 200 genes are located within this QTL (Guberman, et al., 2011). Based on this annotated genome sequence, two genes were identified as candidates for the gene that corresponds to the Ma2 locus (Fig. 53).

The first of these candidate genes, annotated as COP9, putatively encodes a protein that is a member of the COP9 signalosome complex (CSN). In particular, this gene encodes the Complex subunit number 7A. This action of the CSN complex decreases the ubiquitin ligase activity of the SCF-type E3 ubiquitin ligase complex, which is an important post-transcriptional regulator of gene expression (Wei and Deng, 1992). If this gene corresponds to the maturity

locus Ma2, then the action of the dominant allele will be to delay anthesis, potentially by decreasing the level of ubiquitination of a downstream gene that acts to negatively regulate initiation of anthesis.

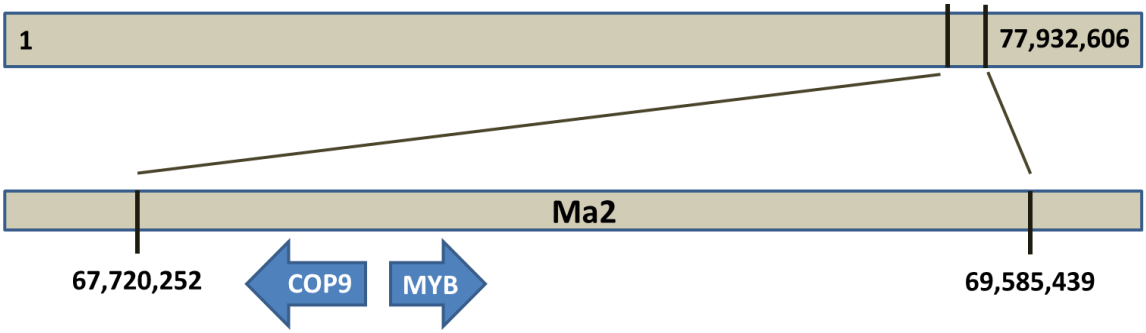

Figure 53: Ma2 locus within chromosome two. The positions of MYB and COP9, the two likeliest candidate genes identified so far at this locus, are shown in their correct orientations. All genes within this region are listed in the appendices.

The second candidate gene is annotated as a MYB transcription factor. This gene sequence encodes two MYB-like DNA binding domains and is a putative MYB-like transcription factor which will act in the nucleus (Chen, et al., 2006). If this gene corresponds to the maturity locus Ma2, the dominant allele will act to delay anthesis, likely by entering the nucleus and affecting the expression of a gene or genes related to control of flowering time. This transcription factor would act at the level of transcription, either by increasing transcription or decreasing transcription of another gene.

Future directions

Initial sequencing of both the MYB transcription factor and COP9 yielded no sequence differences between the cDNA sequences of Hegari (Ma2) and 80M (ma2). Future work will include sequencing of the entire genomic sequence for each of these genes. It is possible that a sequence polymorphism exists in the sequence of one of the introns of either gene that could explain the action of Ma2. It is also possible that the action of Ma2 is actually due to differential expression of one of these genes between Hegari and 80M. Circadian cycling experiments, involving plants grown in short day and long day conditions, will be instrumental in further characterization of each of these candidate genes with respect to the action of Ma2.

It is also worth noting that there are many other genes within the locus identified as Ma2 that have not yet been considered as candidate genes. A search in BioMart shows that there are 212 genes annotated within this locus (Guberman, et al., 2011). Further fine mapping work, which will be carried out using additional F2 lines from the Hegari x 80M population, will help to narrow the bounds of the Ma2 locus, which will exclude a portion of the 212 genes. This will narrow the field of candidate genes and it may become possible to identify the gene that corresponds to the action of Ma2 through fine mapping of the mapped region.

Materials and methods

Parental genotypes used in this study

Four parental genotypes were used to create the populations in this study (Table 16). 100M is a milo genotype that is not recessive for any of the four original maturity loci. 80M is a milo genotype that is recessive for only one original maturity locus, Ma2. It is likely that the 80M genotype arose as an early-flowering mutant from a 100M background (Quinby, 1967), making these two genotypes a lot like nearly-isogenic lines (NILs).

R.07007 is an R-line that has been used in relation to bioenergy hybrid *S. bicolor* genotype development. The maturity genotype of R.07007 was hypothesized to be dominant for Ma2, though experimental findings presented here demonstrate that R.07007 actually expresses the recessive ma2 allele. Hegari is a genotype that was initially identified for its expression of a recessive allele of ma4, which is relatively unique in the *S. bicolor* germplasm.

Table 16: Maturity genotypes of the parental lines used in this study.

Parent	Maturity Genotype
80M	*Ma1ma2Ma3Ma4Ma5ma6*
100M	*Ma1Ma2Ma3Ma4Ma5ma6*
R.07007	*Ma1ma2Ma3Ma4ma5Ma6*
Hegari	*Ma1Ma2Ma3ma4Ma5ma6*

Genetic mating and population creation

The 80M x 100M genetic mating and that of 80M x R.07007 were both carried out by Dr. Bill Rooney at Texas A&M University. In both cases, the initial cross, growing of the putative F1 plants, and verification of F1 status was carried out by Dr. Rooney's group. Confirmed F1 plants were grown in the field in Puerto Rico and self-pollinated to generate the F2 seed that was used in these experiments.

The genetic mating of Hegari and 80M took place following genotype assessment. Many different parental genotypes have been assessed using digital genotyping analysis and these data were used to assess the level of polymorphism between genotypes. On the basis of this comparison, Hegari was chosen as an acceptable genotype to mate with 80M to generate an F2 population that would segregate for Ma2 (and Ma4) and exhibit significant genetic polymorphism throughout the genome.

Hegari was crossed to 80M by Dr. Daryl Morishige at Texas A&M University. Dr. Morishige carried out the cross, growth and confirmation of F1 plants, and threshing of F2 seeds.

Flowering time assessment

80M x 100M F2s

All of the F2 seed that was generated from the F1 plants was bulked for field planting and assessment of flowering time phenotypes. This bulked seed was planted in field plots in April, 2008, at the Texas Agricultural Experiment

Station in College Station, Texas. The plots were not thinned after planting. Plots were irrigated as necessary from emergence through grain filling for all plants.

The first F2 plant reached anthesis on June 3, 2008. The final F2 plant to flower reached anthesis on August 8, 2008. A plant was considered to be at anthesis when pollen shed was visible anywhere on the exserted panicle. The plants were checked for pollen shed every other day following the first plant reaching anthesis on June 3rd. When pollen shed was observed on a panicle, a paper pollinating bag (Lawson Pollinating Bags, Northfield, IL) was placed over the head and stapled shut around the peduncle. The bag was labeled with the date of anthesis for that plant. The bag is used to ensure that self-pollination occurs, as well as to mitigate the effects of insects and birds on the developing seeds.

80M x R.07007 F2s

Following genetic mating of 80M and R.07007, a set of 100 F2 seeds were planted for phenotype assessment. Seeds were planted in 5-gallon nursery pots in a greenhouse at Texas A&M University. The soil used was Sunshine MVP (SunGro Horticulture, Bellevue, Wa). Planting occurred on July 21, 2009. Pots were watered regularly and fertilization was carried out every 14 days using Peter's Professional fertilizer solution (Scotts Professional, The Netherlands). The greenhouse conditions were long days, with 14 hours of light and 10 hours of darkness.

Two F2 seeds were planted per pot. Once the plants had reached the five-leaf stage, a small tissue sample was taken from each plant for genotype analysis. The first 80MxR.07007 F2 plant reached anthesis on September 2, 2009, at 64 days after emergence (DAE). Plants were checked for pollen shed every day. When pollen shed was evident on a panicle, that plant was determined to be at anthesis and it was bagged in the same way that the 80Mx100M F2 plants were. The final F2 in this population to reach anthesis did so on November 8, 2009, at 131 DAE.

Hegari x 80M F2s

The F2 seed generated from the cross of Hegari and 80M was planted in pots in a greenhouse at Texas A&M University. The seeds were planted in two different sets during 2011. Both sets were planted in 5-gallon nursery pots. The soil used to grow these plants was a 2:1 mixture of Coarse Vermiculite (SunGro Horticulture, Bellevue, Wa) to brown pasture soil (American Stone and Turf, College Station, TX). Plants were irrigated as necessary following emergence and were fertilized every 14 days using Peter's Professional fertilizer solution (Scotts Professional, The Netherlands). The greenhouse conditions were long days, with 14 hours of light and 10 hours of dark.

The first set of Hegarix80M F2 plants was planted on April 14, 2011. This set included 286 F2 plants. The second set of F2 plants was planted on September 21, 2011. This set included 146 F2 plants. When the plants had reached the five-leaf stage, a small sample of tissue was taken from each plant

for genotype analysis. Each plant, upon reaching anthesis, was bagged in the same way that was described for the 80Mx100M F2 population.

For the first set of Hegarix80M F2 plants, the first to reach anthesis did so on May 26, 2011, at 42 DAE. Some of the plants in this set did not reach anthesis due to extremely high temperatures in the greenhouse during the summer weather of 2011.

The second set of Hegarix80M F2 plants, which was planted in the fall, was not subjected to the extreme heat of the first set. The first to reach anthesis did so on November 19, 2011, at 59 DAE. The last plant of this set to reach anthesis did so on February 14, 2012, at 146 DAE.

Genotyping

For each F2 plant used in this study, DNA was extracted for genotyping using a FastDNA Spin Kit and FastPrep apparatus (MP Biomedicals). The tissue used for each extraction was a single leaf from each F2 plant, taken once the plant had reached the five-leaf stage. Resulting DNA samples were quantitated in triplicate using the Qubit fluorometer and DNA-BR standards kit (Invitrogen). Only samples with concentrations between 50 and 120 ng/µL were considered acceptable for continued processing.

The remainder of sample preparation was carried out according to Morigishige et al., 2012 (in prep). Prepared DNA was then analyzed using an Illumina $GAII_x$ instrument (Illumina). Generation of this genotype data was carried out by the Laboratory for Plant Genome Technologies (LPGT) at Texas

A&M University. This sequencing platform generates large quantities of short sequences (72 bp), which were used to create a haplotype map of the entire genome for each F5 family. The haplotypes are based on single nucleotide polymorphisms (SNPs) and short insertions or deletions (InDels). Any sequence polymorphism identified between the two parents, either 80M and R.07007, or Hegari and 80M, was considered a locus or marker. Each F2 DNA sample was then analyzed to determine the parental identity at each of these markers. Analysis of genotype data for this population was carried out by Dr. Patricia Klein at Texas A&M University. The final output of this genotyping process is a haplotype map which indicates identity by descent to one parent or the other for each F2 plant at every possible locus across the genome. This information is then used in subsequent genetic map construction and QTL mapping.

Genetic map construction

Genetic maps for each chromosome were created based on the genotype data collected. A freely available piece of software, MapMaker 3.0B, was used to construct these genetic maps. This program uses user-input data about the generation level of the population as well as genotype information to determine the genetic distance between each pair of markers in centimorgans (cM). Genetic distances are based on the number of crossing-over events (or parental genotype changes) that occur within a population between a pair of loci. The calculation of genetic distance is based on statistics and is of great importance for use in the following step of this process.

QTL mapping

Once a genetic map was made for each chromosome, and the phenotypes of interest were all measured, quantitative trait locus (QTL) maps were generated. This is a process which combines user-input genotype and phenotype data to identify QTL throughout the genome on a trait-by-trait basis. The free software WinQTLCartographer (v2.5.010) was used to generate QTL maps. The composite interval mapping module with standard parameters was used.

Following generation of QTL maps, random permutations were carried out (1000 iterations) to determine the threshold LOD score for each trait. For traits where permutation tests were not done, the threshold was assumed to be 3.0. The positions of QTL are reported in terms of genetic position, or cM. However, in order to make these results more easily relatable to QTL mapping studies done on other populations, QTL positions were converted from genetic to physical positions manually. In cases where the genetic position of the start, peak, or end of a QTL was located between two physically located markers, the physical position was estimated based on the distance between each of the markers and the reported genetic position.

Genome analysis

Once QTL were identified in these populations, potential candidate genes were identified based on the annotated genome sequence available at http://phytozome.org (Goodstein, et al., 2012). The genome browser and the

BioMart application that are accessible through the Phytozome website were used to generate a list of potential candidate genes from all of the annotated genes located within the genomic region identified as Ma2 (Guberman, et al., 2011).

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The experiments described in this dissertation have yielded a large quantity of data which, when taken together, make a clear case for the ideal nature of *Sorghum bicolor* for use in generation of biomass for biofuels. Many of the attributes of bioenergy hybrid genotypes of *S. bicolor* are beneficial in an energy crop, and there are additional attributes of other *S. bicolor* genotypes that can be incorporated into bioenergy hybrid genotypes to further increase the applicability of these genotypes in biomass generation.

Bioenergy hybrid Sorghum bicolor *generates high biomass*

In chapter one it was established that bioenergy hybrid *S. bicolor* genotypes can generate very large amounts of biomass. This high yield is due to rapid canopy closure by juvenile plants, high radiation use efficiency, long duration of vegetative growth, and a very high stem-to-leaf ratio. Based on these results, the bioenergy hybrid *S. bicolor* genotype TX08001 exceeds biomass yields of grain type *S. bicolor* as well as reported biomass yields for many other biomass crops (Byrt, et al., 2011; Dohleman and Long, 2009; Heaton, et al., 2004; McLaughlin and Kszos, 2005). Irrigation of growing *S. bicolor* plants is necessary throughout the growing season to achieve maximum yield. The genotype tested in this experiment demonstrated a cessation of biomass accumulation when subjected to dryland conditions with very limited

precipitation, but biomass accumulation resumed later in the growing season when rainfall rehydrated the soil. The information presented in chapter one demonstrate that bioenergy hybrid *S. bicolor* has the biomass generation capacity to fulfill the renewable fuels standards established for the United States (US Environmental Protection Agency, 2010).

One of the most important attributes of any crop that is to be used to generate biomass for biofuels is the ability of that crop to accumulate a large quantity of biomass. The data presented in chapter two of this dissertation demonstrate the significant genetic yield potential of TX08001, a bioenergy hybrid genotype of *S. bicolor* (Fig. 9). This genotype of *S. bicolor*, and other hybrid genotypes like it, will be ideal candidates as energy crops.

NUE is high in bioenergy hybrid Sorghum bicolor

Nitrogen use efficiency (NUE) is a measurement of the amount of biomass accumulated by a crop per amount of nitrogen (N) contained within the tissues of that plant. Bioenergy hybrid *S. bicolor* exhibits high NUE compared to grain type *S. bicolor* as well as many other crops (Table 4). This high level of NUE has been attributed to a long duration of vegetative growth accompanied by the recycling of N from older leaf tissue to newer leaf tissue (van Oosterom, et al., 2010; Hirel, et al., 2007).

As was previously discussed, fertilization with N accounts for a significant portion of the financial investment required to produce biofuels from biomass. An ideal energy crop will have a high level of NUE such that no additional N

fertilizer will be required for optimal growth of the energy crop beyond what would be required for any other grain crop. Bioenergy hybrid *S. bicolor* can accumulate the reported high biomass yield when grown in soil that is fertilized at a level optimal for growth of grain type *S. bicolor*. Additionally, the NUE of bioenergy hybrid *S. bicolor* increases constantly throughout the growing cycle, suggesting that earlier planting, and hence a longer duration of vegetative growth, could lead to even higher NUE. On the basis of NUE, bioenergy hybrid *S. bicolor* is an ideal energy crop; its NUE ranks among the highest levels reported for candidate biofuel crops (Table 4).

Many loci contribute to plant size in Sorghum bicolor

When considering a crop's ability to generate biomass for biofuels, the size of the plant is an important determinant of the eventual biomass yield (Heaton, et al., 2004; Bhandari, et al., 2011). While increased biomass yield can be achieved through targeted breeding approaches, greater gains will be possible with the use of genetic tools once the genes underlying plant size traits have been identified. The genetic regulation of plant size in *S. bicolor* occurs through the action of a complex network of genes. In a population of SC170 x M35-1 F5 plants, QTL were identified that modulate a multitude of morphometric traits. Traits analyzed included a selection of macro traits, such as days to anthesis and whole plant biomass yield (total DW), as well as many individual component traits, like the lengths of individual internodes and the widths of individual leaves. While some loci were identified that were unique to individual

traits, many of the QTL identified that modulate individual traits were also identified to be modulating macro traits as well.

Through examination of the coincidence of QTL for related traits, it was possible to identify putative regulators that modulate complex macro traits. For example, three loci were identified for stem length that were also identified to be modulating various other component traits in this population. Located on chromosome nine, Dw1 modulates stem length as well as total DW and individual internode lengths (Fig. 42). This locus increases stem length without resulting in a decrease in biomass yield, which is an invaluable trait for any biofuel crop. Dw3 on chromosome seven increases stem length and individual internode length, but not total DW. The action of this locus is different than that of Dw1, however, as increases in length are accompanied by a decrease in internode diameter, which is why the action of Dw3 can increase length but not DW of the stem. Dwx, which is located on chromosome one, contributes to an increase in stem length accompanied by a decrease in secondary cell wall components.

These loci and the others identified by the experiments described in this dissertation can have significant effects on the biomass yield of *S. bicolor* plants. Biomass yield is a critical component of biofuel yield from biomass. *S. bicolor* is an ideal energy crop not only because of the high biomass yield it can attain, but also because of the considerable number of genetic loci contributing to biomass

yield identified here, which can be used to further increase biomass yield in *S. bicolor* grown for biomass for biofuels.

Composition of Sorghum bicolor *biomass is ideal*

Many biomass components warrant examination when considering the applicability of a crop for use in generation of biomass for biofuels. Cellulose is an important component for conversion to fuel through fermentation (Dale, 1987). Approximately 25% of the DW of a bioenergy hybrid *S. bicolor* stem is cellulose, making biomass from this genotype ideal for generation of biofuels through cellulose conversion. In addition, *S. bicolor* biomass contains a large lignin component, which is very important for generation of ethanol from lignocellulosic biomass (Dale, 1987; McDermitt and Loomis, 1981). Other component percentages are also relevant to generation of biofuels, as components like ash and other minerals can negatively impact the biofuel yield from a given source of biomass (Monti, et al., 2008). The ash composition of *S. bicolor* reported in this dissertation is on the low end of the ranges reported for biomass that is appropriate for use in generation of biofuels, which helps to make the case that *S. bicolor* is an ideal candidate for use in generation of biomass for biofuels.

Delayed flowering due to Ma2 can increase biomass yield

As floral maturity is a terminal state in *S. bicolor*, control of flowering time is a viable method of manipulating biomass yield in this species (Rooney and Aydin, 1999; Rooney, 2004). Of the maturity loci identified in *S. bicolor*, *Ma2*

had not been located prior to the work reported in this dissertation. This locus, like the other loci controlling maturity in *S. bicolor*, acts to delay flowering when a dominant allele is present (Quinby and Karper, 1945). *Ma2*, which is located near the end of chromosome two, was identified in an F2 population of Hegari x 80M, where Hegari contains a dominant allele of *Ma2* and 80M contains a recessive *ma2* allele (Fig. 52). While the individual gene acting at this locus is not yet known, the identification of the genetic position of this maturity gene is a significant step forward in understanding the overall genetic control of flowering time in *S. bicolor*.

Taken together, the conclusions put forward in this dissertation make a strong case for the applicability of *S. bicolor* for generation of biomass for biofuels. On the basis of high biomass yield, limited requirement of inputs like nitrogen and water, and the composition of biomass generated, *S. bicolor* is an ideal energy crop.

Future directions

While the data presented in this dissertation make a compelling case for the ideal nature of *S. bicolor* for use in generation of biomass for biofuels, the examination of this crop is not yet complete. Additional efforts, built upon the conclusions made here, have the potential to provide additional strong evidence for the applicability of *S. bicolor* as an energy crop. Following are suggestions for future directions which would make a significant contribution to furthering the conclusions set forth in this dissertation.

Improving Sorghum bicolor *biomass accumulation*

The high biomass yield of bioenergy hybrid *S. bicolor* makes it an ideal energy crop. However, additional gains in biomass yield could further increase the potential fuel yield from bioenergy hybrid *S. bicolor*. As was discussed in chapter two, the work presented in this dissertation demonstrates that there are clear opportunities for biomass yield improvement.

First, it was shown that limited water availability in mid-August in rainfed plots led to a temporary cessation of biomass accumulation. The arrival of rain in September brought about renewed growth in these plants. This growth pattern suggests that incorporating well-documented quiescence adaptations from other genotypes of *S. bicolor* could lead to improved biomass yield (Tuinstra, et al., 1997; Mutava, et al., 2011). By creating hybrid *S. bicolor* genotypes that have an enhanced ability to grow in water-limiting conditions, it will be possible to generate more biomass than before even when water is limited, as is often the case for crops grown in rainfed plots. Not only would such an effort increase the potential biomass yield of bioenergy hybrid *S. bicolor*, it would also decrease the water input requirement, making it less costly to produce biomass from this crop. Both of these attributes would make *S. bicolor* an even more ideal biomass crop.

In addition, the bioenergy hybrid *S. bicolor* genotype used in these studies is photoperiod sensitive and will not reach anthesis until October in College Station, Texas (Rooney, et al., 2007; Rooney and Aydin, 1999). The

length of the vegetative phase is determined by the flowering time of the crop as well as the planting time. It will be important to determine, through future experiments, how early in the year bioenergy hybrid *S. bicolor* can be planted in order to establish a strong stand for growth until October. It is possible that, through earlier planting, this crop will be able to accumulate even more biomass prior to anthesis and cessation of vegetative growth in October.

NUE in bioenergy hybrid Sorghum bicolor

There are significant questions which remain to be answered pertaining to the NUE of bioenergy hybrid *S. bicolor*. The experiments reported in this dissertation provide a basic framework for understanding this parameter, but it will be necessary to study NUE of *S. bicolor* further in order to take full advantage of this attribute with respect to growing *S. bicolor* as an ideal energy crop. First, it will be important to determine the minimum N fertilization level necessary for optimum growth of bioenergy hybrid *S. bicolor*. In chapter three, it was demonstrated that bioenergy hybrid plants achieved very large biomass yields when grown in a field fertilized with 100 kg Ha^{-1} N. It will be important to determine, through future growth experiments, whether the same biomass yield can be achieved when plants are grown in soil with less N content. A decrease in the N application rate for a field will decrease the total cost of production of biomass from that field and as such, decrease the eventual cost of biofuel produced from that biomass (Hons, et al., 1986).

It will also be important to determine whether any gains in genetic yield potential can be made through manipulation of leaf size, appearance rate, or N remobilization of bioenergy hybrid *S. bicolor* plants. Since it has been demonstrated that extremely long duration of vegetative growth is achieved in *S. bicolor* through remobilization of N from lower leaves to upper young leaves, it will be critical to determine how much of the N in a leaf can be remobilized by the plant prior to senescence and detachment of that leaf. It may be possible to capitalize on the N remobilization of *S. bicolor* to further increase NUE of long-growing bioenergy hybrid *S. bicolor* genotypes.

Finally, the roots of *S. bicolor* plants warrant further inspection. While this portion of a plant grows below the soil and is not a part of the harvestable biomass, the root system does participate in the uptake and partitioning of N (Takei, et al., 2002; Richard-Molard, et al., 2008). Limited studies of the root systems of *S. bicolor* have been presented in this dissertation, but more in-depth study will be necessary for full understanding of this system. Biomass composition of the root system needs to be determined in order to assess the sink strength of the roots in terms of N and other components. Also, the size of the root system needs to be further examined, as it may be possible to breed bioenergy hybrid *S. bicolor* genotypes with decreasing root biomass in favor of partitioning greater biomass into the shoots without sacrificing stem stability or drought survival.

Identification of genes corresponding to plant size QTL

The large number of QTL identified that modulate plant size traits in *S. bicolor* necessitate further study. As is the case in most QTL mapping studies, the next step will be to carry out fine mapping and candidate gene identification for the most promising of the loci identified. Dw1 and Dwx merit the most attention, as these two loci have the capacity to affect both stem length and total DW in mature plants. The population used in this study was the F5 generation of SC170 x M35-1. The F6 generation of this population is available and could be used to carry out fine mapping of the loci identified. Once fine mapping is successfully carried out, it will be possible to identify and validate candidate genes for these loci. Once the genes underlying the action of identified loci can be identified, it will be possible to include those genes in future breeding efforts for bioenergy hybrid genotypes of *S. bicolor*. This will make it possible to further increase the biomass yield potential of such genotypes, which will make *S. bicolor* a more ideal energy crop candidate.

Identifying the gene underlying Ma2 on chromosome two

Finally, it will be important to further characterize the *Ma2* maturity locus described in this dissertation. Control of maturity contributes to control of biomass yield in crops like *S. bicolor*, where floral maturity is a terminal state (Quinby, 1972). The genetic location of *Ma2* was identified in this work, but further refinement of this position is possible through further experimentation. Additional genetic data for fine mapping may be obtained through digital

genotyping and QTL mapping using the additional 200 lines that make up the Hegari x 80M F2 population that have not yet been genotyped. By carrying out fine mapping, it will be possible to narrow the bounds of the *Ma2* locus identified in chapter five of this dissertation.

Two candidate genes, COP9 and MYB, have been identified in this dissertation. While these are only two genes of the 200 possible genes located within the *Ma2* locus, further study of these genes is merited. The sequences of these two genes have not yielded any apparent polymorphisms between Hegari and 80M, so further information may be gained through expression studies. Circadian cycling experiments like those described in Murphy et al. (2011) need to be undertaken to determine whether a difference in expression of one of these genes can be correlated to the action of *Ma2*.

Taken together, these proposed future directions would contribute significantly to the status of *S. bicolor* as an ideal energy crop. The conclusions presented in chapters two through five of this dissertation address the ideal nature of *S. bicolor* for use in generation of biomass for biofuels on the basis of high biomass yield, ideal biomass composition, and minimal input requirements. While there are many crops which may be useful as energy crops, this dissertation makes a compelling case for *S. bicolor* as an ideal energy crop.

NOMENCLATURE

DAE	Days after emergence
ε_c	Radiation use efficiency
ε_i	Radiation interception efficiency
ε_p	Partitioning efficiency (harvest index)
PAR	Photosynthetically active radiation
S_t	Total PAR incident within a given time period
dT	Dry tons
N	Nitrogen
NUE	Nitrogen use efficiency
QTL	Quantitative trait locus
NIR	Near infrared reflectance spectroscopy

LITERATURE CITED

Agrama H, Zakaria A, Said F, Tuinstra M (1999) Identification of quantitative trait loci for nitrogen use efficiency in maize. Mol Breed **5:** 187-195

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration - guidelines for computing crop water requirements. FAO Irrigation and Drainage **56:** 10-85

Anderson J, Churchill G, Autrique J, Tanksley S, Sorrells M (1993) Optimizing parental selection for genetic-linkage maps. Genome **36:** 181-186

Apariciotejo P, Boyer J (1983) Significance of accelerated leaf senescence at low water potentials for water-loss and grain-yield in maize. Crop Sci **23:** 1198-1202

Asins M (2002) Present and future of quantitative trait locus analysis in plant breeding. Plant Breed **121:** 281-291

Beale C, Long S (1997) Seasonal dynamics of nutrient accumulation and partitioning in the perennial C-4-grasses Miscanthus x giganteus and Spartina cynosuroides. Biomass Bioenergy **12:** 419-428

Bégué A (1993) Leaf area index, intercepted photosynthetically active radiation, and spectral vegetation indices: a sensitivity analysis for regular-clumped canopies. Remote Sens Environ **46:** 45-59

Bennett AS, Anex RP (2009) Production, transportation and milling costs of sweet sorghum as a feedstock for centralized bioethanol production in the upper Midwest. Bioresour Technol **100:** 1595-1607

Bhandari HS, Saha MC, Fasoula VA, Bouton JH (2011) Estimation of genetic parameters for biomass yield in lowland switchgrass (Panicum virgatum L.). Crop Sci **51:** 1525-1533

Blázquez MA (2000) Flower development pathways. J Cell Sci **113:** 3547-3548

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol **54:** 519-546

Borrell A, Hammer G (2000) Nitrogen dynamics and the physiological basis of stay-green in sorghum. Crop Sci **40:** 1295-1307

Bouton JH (2007) Molecular breeding of switchgrass for use as a biofuel crop. Curr Opin Genet Dev **17:** 553-558

Bowman WD (1991) Effect of nitrogen nutrition on photosynthesis and growth in C4Panicum species. Plant, Cell Environ **14:** 295-301

Brown PJ, Rooney WL, Franks C, Kresovich S (2008) Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics **180:** 629-637

Brown R (1978) Difference in N use efficiency in C3 and C4 plants and its implications in adaptation and evolution. Crop Sci **18:** 93-98

Buchanan C, Lim S, Salzman R, Kagiampakis L, Morishige D, Weers B, Klein R, Pratt L, Cordonnier-Pratt M, Klein P, Mullet J (2005) *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol **58:** 699-720

Byrt CS, Grof CPL, Furbank RT (2011) C(4) Plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspectivefree access. J Integr Plant Biol **53:** 120-135

Campbell L, Casady A, Crook W (1975) Effects of a single height gene (Dw3) of sorghum on certain agronomic characters. Crop Sci **15:** 595-597

Caravetta GJ, Cherney JH, Johnson KD (1990) Within-row spacing influences on diverse sorghum genotypes: I. morphology. Agronomy Journal **82:** 206-218

Carena MJ (2009) Cereals. Springer Science, New York, pp 182-253

Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci **13:** 415-420

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci **48:** 30-40

Chaves M, Maroco J, Pereira J (2003) Understanding plant responses to drought - from genes to the whole plant. Funct Plant Biol **30:** 239-264

Chen Y, Yang X, He K, Liu M, Li J, Gao Z, Lin Z, Zhang Y, Wang X, Qiu X, Shen Y, Zhang L, Deng X, Luo J, Deng X, Chen Z, Gu H, Qu L (2006) The MYB transcription factor superfamily of arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol **60:** 107-124

Childs K, Miller F, CordonnierPratt M, Pratt L, Morgan P, Mullet J (1997) The sorghum photoperiod sensitivity gene, Ma(3), encodes a phytochrome B. Plant Physiol **113:** 611-619

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971

Clifton-Brown JC, Lewandowski I (2000) Water use efficiency and biomass partitioning of three different miscanthus genotypes with limited and unlimited water supply. Annals of Botany **86:** 191-200

Collard B, Jahufer M, Brouwer J, Pang E (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica **142:** 169-196

Cui Z, Zhang F, Mi G, Chen F, Li F, Chen X, Li J, Shi L (2009) Interaction between genotypic difference and nitrogen management strategy in determining nitrogen use efficiency of summer maize. Plant Soil **317:** 267-276

Dale BE (1987) Lignocellulose conversion and the future of fermentation biotechnology. Trends Biotechnol **5:** 287-291

Dewet J, Huckabay J (1967) Origin of *Sorghum bicolor* .2. distribution and domestication. Evolution **21:** 787-813

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS (2007) Domestication to crop improvement: genetic resources for sorghum and saccharum (Andropogoneae) RID B-5824-2008. Ann Bot **100:** 975-989

Dodds K, Ball R, Djorovic N, Carson S (2004) The effect of an imprecise map on interval mapping QTLs. Genet Res **84:** 47-55

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell **127:** 1309-1321

Dohleman FG, Heaton EA, Leakey ADB, Long SP (2009) Does greater leaf-level photosynthesis explain the larger solar energy conversion efficiency of Miscanthus relative to switchgrass? Plant, Cell Environ **32:** 1525-1537

Dohleman FG, Long SP (2009) More productive than maize in the midwest: how does miscanthus do it? Plant Physiology **150:** 2104-2115

Dugas DV, Monaco MK, Olsen A, Klein RR, Kumari S, Ware D, Klein PE (2011) Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics **12:** 514-534

Eilrich GL, Long RC, Stickler FC, Pauli AW (1964) Stage of maturity, plant population, and row width as factors affecting yield and chemical composition of Atlas forage sorghum. Kansas Agricultural Experiment Station Technical Bulletin 138

Fazio S, Monti A (2011) Life cycle assessment of different bioenergy production systems including perennial and annual crops. Biomass & Bioenergy **35:** 4868-4878

Foster K, Miller F, Childs K, Morgan P (1994) Genetic-regulation of development in sorghum-bicolor .7. shoot growth, tillering, flowering, gibberellin biosynthesis, and phytochrome levels are differentially affected by dosage of the Ma(3)(r) allele. Plant Physiol **105:** 941-948

Foyer, C.H., Hanma, Z. (2011) Nitrogen metabolism in plants in the post-genomic era. Wiley-Blackwell, Hoboken, pp 123-142

Frank AB, Berdahl JD, Hanson JD, Liebig MA, Johnson HA (2004) Biomass and carbon partitioning in switchgrass. Crop Science **44:** 1391-1396

Frink CR, Waggoner PE, Ausubel JH (1999) Nitrogen fertilizer: retrospect and prospect. Proc Natl Acad Sci U S A **96:** 1175-1180

Gaju O, Allard V, Martre P, Snape JW, Heumez E, Le Gouis J, Moreau D, Bogard M, Griffiths S, Orford S, Hubbart S, Foulkes MJ (2011) Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crops Res **123:** 139-152

George-Jaeggli B, Jordan DR, van Oosterom EJ, Hammer GL (2011) Decrease in sorghum grain yield due to the Dw3 dwarfing gene is caused by reduction in shoot biomass. Field Crops Res **124:** 231-239

Gepts P (2002) A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Sci **42:** 1780-1790

Gerik T, Bean B, Vanderlip R (2003) Sorghum growth and development. Kansas State University, Manhattan, pp 3-16

Gerik TJ, Neely CL (1987) Plant density effects on main culm and tiller development of grain sorghum. Crop Sci **27:** 1225-1230

Ghannoum O, Evans J, Chow W, Andrews T, Conroy J, von Caemmerer S (2005) Faster rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C(4) grasses. Plant Physiol **137:** 638-650

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40:** D1178-D1186

Guan Y, Wang H, Qin L, Zhang H, Yang Y, Gao F, Li R, Wang H (2011) QTL mapping of bio-energy related traits in sorghum. Euphytica **182:** 431-440

Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, Baldock R, Chelala C, Croft D, Cros A, Cutts RJ, Di Genova A, Forbes S, Fujisawa T, Gadaleta E, Goodstein DM, Gundem G, Haggarty B, Haider S, Hall M, Harris T, Haw R, Hu S, Hubbard S, Hsu J, Iyer V, Jones P, Katayama T, Kinsella R, Kong L, Lawson D, Liang Y, Lopez-Bigas N, Luo J, Lush M, Mason J, Moreews F, Ndegwa N, Oakley D, Perez-Llamas C, Primig M, Rivkin E, Rosanoff S, Shepherd R, Simon R, Skarnes B, Smedley D, Sperling L, Spooner W, Stevenson P, Stone K, Teague J, Wang J, Wang J, Whitty B, Wong DT, Wong-Erasmus M, Yao L, Youens-Clark K, Yung C, Zhang J, Kasprzyk A (2011) BioMart central portal: an open database network for the biological community. Database **11:** 1-9

Hackett C (2002) Statistical methods for QTL mapping in cereals. Plant Mol Biol **48:** 585-599

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

Hartley BE, Gibson JD, Alex Thomasson J, Searcy SW (2011) Moisture loss and ash characterization of high-tonnage sorghum. Trans ASABE **2011:** 38-49

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of Miscanthus. Global Change Biol **14:** 2000-2014

Heaton E, Voigt T, Long S (2004) A quantitative review comparing the yields of two candidate C-4 perennial biomass crops in relation to nitrogen, temperature and water. Biomass Bioenergy **27:** 21-30

Hedge B, Major D, Wilson D, Krogman K (1976) Effects of row spacing and population-density on grain-sorghum production in southern alberta. Can J Plant Sci **56:** 31-37

Himmel ME, Ding S, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:** 804-807

Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot **58:** 2369-2387

Hons F, Moresco R, Wiedenfeld R, Cothren J (1986) Applied nitrogen and phosphorus effects on yield and nutrient-uptake by high-energy sorghum produced for grain and biomass. Agron J **78:** 1069-1078

International Crops Research Institute for the Semi-Arid Tropics (2011) ICRISAT (2012) Sorghum crop, http://www.icrisat.org/crop-sorghum.htm.

Jorgensen U (2011) Benefits versus risks of growing biofuel crops: the case of Miscanthus. Curr Opin Environ Sustain **3:** 24-30

Kearsey M (1998) The principles of QTL analysis (a minimal mathematics approach). J Exp Bot **49:** 1619-1623

Kim J, Islam-Faridi M, Klein P, Stelly D, Price H, Klein R, Mullet J (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics **171:** 1963-1976

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci **48:** 212-226

Lawlor D (2002) Carbon and nitrogen assimilation in relation to yield: mechanisms are the key to understanding production systems. J Exp Bot **53:** 773-787

Lawlor D, Cornic G (2002) Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. Plant Cell and Environment **25:** 275-294

Lewandowski I, Schmidt U (2006) Nitrogen, energy and land use efficiencies of miscanthus, reed canary grass and triticale as determined by the boundary line approach. Agric Ecosyst Environ **112:** 335-346

Lin Y, Schertz K, Paterson A (1995) Comparative-analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population. Genetics **141:** 391-411

Mace ES, Jordan DR (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. Theor Appl Genet **123:** 169-191

Makino A, Sakuma H, Sudo E, Mae T (2003) Differences between maize and rice in N-use efficiency for photosynthesis and protein allocation. Plant Cell Physiol **44:** 952-956

Mall TK, Dweikat I, Sato SJ, Neresian N, Xu K, Ge Z, Wang D, Elthon T, Clemente T (2011) Expression of the rice CDPK-7 in sorghum: molecular and phenotypic analyses. Plant Mol Biol **75:** 467-479

Mastrorilli M, Katerji N, Rana G, Steduto P (1995) Sweet sorghum in mediterrenean climate: radiation use and biomass water use efficiencies. Ind Crops Prod **3:** 253

McDermitt D, Loomis R (1981) Elemental composition of biomass and its relation to energy content, growth efficiency, and growth-yield. Ann Bot **48:** 275-290

McLaughlin S, Kszos L (2005) Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States. Biomass Bioenergy **28:** 515-535

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Toerjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proc Natl Acad Sci U S A **104:** 4759-4764

Ming R, Del Monte T, Hernandez E, Moore P, Irvine J, Paterson A (2002) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polypiold genomes. Genome **45:** 794-803

Monk R, Miller F, McBee G (1984) Sorghum improvement for energy-production. Biomass **6:** 145-153

Monti A, Di Virgilio N, Venturi G (2008) Mineral composition and ash content of six major energy crops. Biomass Bioenergy **32:** 216-223

Muchow R, Sinclair T (1994) Nitrogen response of leaf photosynthesis and canopy radiation use efficiency in field-grown maize and sorghum. Crop Sci **34:** 721-727

Multani D, Briggs S, Chamberlin M, Blakeslee J, Murphy A, Johal G (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. Science **302:** 81-84

Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc Natl Acad Sci U S A **108:** 16469-16474

Murray SC, Rooney WL, Mitchell SE, Sharma A, Klein PE, Mullet JE, Kresovich S (2008a) Genetic improvement of sorghum as a biofuel feedstock: ii. qtl for stem and leaf structural carbohydrates. Crop Sci **48:** 2180-2193

Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S (2008b) Genetic improvement of sorghum as a biofuel feedstock: i. qtl for stem sugar and grain nonstructural carbohydrates. Crop Sci **48:** 2165-2179

Murty BR, Govil JN (1967) Description of 70 groups in genus sorghum based on a modified snowden's classification. Indian J Genet **27:** 75 -78

Mutava RN, Prasad PVV, Tuinstra MR, Kofoid MD, Yu J (2011) Characterization of sorghum genotypes for traits related to drought tolerance. Field Crops Res **123:** 10-18

National Research Council (U.S.). Committee on Economic and Environmental Impacts of Increasing Biofuels Production., National Research Council (U.S.). Board on Agriculture and Natural Resources., National Research Council (U.S.). Board on Energy and Environmental Systems. (2011) Renewable fuel standard : potential economic and environmental effects of U.S. biofuel policy. National Academies Press, Washington, D.C.

Nguyen HT, (2004) Physiology and biotechnology integration for plant breeding. Marcel Dekker, Inc, New York, pp 81-102

Osugi A, Itoh H, Ikeda-Kawakatsu K, Takano M, Izawa T (2011) Molecular dissection of the roles of phytochrome in photoperiodic flowering in rice. Plant Physiol **157:** 1128-1137

Owen Reece H (1999) Near infrared spectroscopy. Br J Anaesth **82:** 418

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551-556

Perlack R, Wright L, Turhollow A, Graham R, Stokes B, Erbach D (2005) Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. U S Department of Energy, Oak Ridge, pp 13-50

Perry LG, Blumenthal DM, Monaco TA, Paschke MW, Redente EF (2010) Immobilizing nitrogen to control plant invasion. Oecologia **163:** 13-24

Powell J, Hons F, McBee G (1991) Nutrient and carbohydrate partitioning in sorghum stover. Agron J **83:** 933-937

Price H, Dillon S, Hodnett G, Rooney W, Ross L, Johnston J (2005) Genome evolution in the genus Sorghum (Poaceae). Ann Bot **95:** 219-227

Quinby JR (1966) 4th maturity gene locus in sorghum. Crop Sci **6:** 516-518

Quinby JR (1967) The maturity genes of sorghum. Adv Agron **19:** 267-271

Quinby JR (1972) Influence of maturity genes on plant-growth in sorghum. Crop Sci **12:** 490-492

Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station, pp 56-62

Quinby JR, Karper R (1945) The inheritance of 3 genes that influence time of floral initiation and maturity date in milo. J Am Soc Agron **37:** 916-936

Quinby JR, Karper R (1954) Inheritance of height in sorghum. Agron J **46:** 211-216

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick Jr. WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science **311:** 484-489

Rahall N (2007) H.R. 6 (110th): Energy independence and security act of 2007.

Raneses AR, Glaser LK, Price JM, Duffield JA (1999) Potential biodiesel markets and their economic effects on the agricultural sector of the United States. Ind Crops Prod **9:** 151-162

Ranjith S, Meinzer F (1997) Physiological correlates of variation in nitrogen-use efficiency in two contrasting sugarcane cultivars. Crop Sci **37:** 818-825

Rice J, Saccone N, Corbett J (2001) The LOD Score Method. Genetic Dissection of Complex Traits, Academic Press Inc, San Diego, pp 99-113

Richard-Molard C, Krapp A, Brun F, Ney B, Daniel-Vedele F, Chaillou S (2008) Plant response to nitrate starvation is determined by N storage capacity matched by nitrate uptake capacity in two Arabidopsis genotypes. J Exp Bot **59:** 779-791

Roberts CA, Houx,J.H.,,III, Fritschi FB (2011) Near-infrared analysis of sweet sorghum bagasse. Crop Sci **51:** 2284-2288

Robinson RG, Bernat LA, Nelson WW, Thompson RL, Thompson JR (1964) Row spacing and plant population for grain sorghum in the humid north. Agron J **56:** 189-191

Rooney W (2004) Sorghum improvement - integrating traditional and newtechnology to produce improved genotypes. Adv Agron, Vol 83 **83:** 37-109

Rooney W, Aydin S (1999) Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) moench. Crop Sci **39:** 397-400

Rooney WL, Blumenthal J, Bean B, Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. Biofuels Bioprod Biorefining **1:** 147-157

Sage RF, Monson RK (1999) C4 plant biology. Academic Press, San Diego, pp 12-15

Sage R, Pearcy R (1987a) The nitrogen use efficiency of c-3 and c-4 plants .1. leaf nitrogen, growth, and biomass partitioning in chenopodium-album (l) and amaranthus-retroflexus(l). Plant Physiol **84:** 954-958

Sage R, Pearcy R (1987b) The nitrogen use efficiency of c-3 and c-4 plants .2. leaf nitrogen effects on the gas-exchange characteristics of chenopodium-album (l) and amaranthus-retroflexus (l). Plant Physiol **84:** 959-963

Sainju UM, Whitehead WF, Singh BP, Wang S (2006) Tillage, cover crops, and nitrogen fertilization effects on soil nitrogen and cotton and sorghum yields. Eur J Agron **25:** 372-382

Sanderson M, Agblevor F, Collins M, Johnson D (1996a) Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy. Biomass Bioenergy **11:** 365-370

Sanderson M, Reed R, McLaughlin S, Wullschleger S, Conger B, Parrish D, Wolf D, Taliaferro C, Hopkins A, Ocumpaugh W, Hussey M, Read J, Tischler C (1996b) Switchgrass as a sustainable bioenergy crop. Bioresour Technol **56:** 83-93

Sang T (2011) Toward the domestication of lignocellulosic energy crops: learning from food crop domestication. J Integr Plant Biol **53:** 96-104

Schmer MR, Vogel KP, Mitchell RB, Perrin RK (2008) Net energy of cellulosic ethanol from switchgrass. Proc Natl Acad Sci U S A **105:** 464-469

Schmitt M, Edwards G (1981) Photosynthetic capacity and nitrogen use efficiency of maize, wheat, and rice - a comparison between c-3 and c-4 photosynthesis. J Exp Bot **32:** 459-466

Semagn K, Bjornstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. EJB **13:** 14

Simmons BA, Loque D, Blanch HW (2008) Next-generation biomass feedstocks for biofuel production. Genome Biol **9:** 242-247

Sinclair T, Muchow R (1999) Radiation use efficiency. Adv Agron **65:** 215-265

Smirnoff N (1993) Tansley review 52. the role of active oxygen in the response of plants to water-deficit and desiccation. New Phytol **125:** 27-58

Smith CW, Frederiksen RA (2000) Sorghum: origin, history, technology, and production. John Wiley and Sons, New York, pp 201-318

Snowdon JD (1936) The cultivated races of sorghum. Adlard and Son, London, pp 150-274

Starks KJ, Doggett (1970) Breeding for resistance to the sorghum shoot fly. Crop Sci **10:** 528-531

Stockle CO, Kemanian AR (2008) On the use of radiation- and water-use efficiency for biomass production models. In LR Ahuja, VR Reddy, SA Saseendran, Q Yu, eds, Response of crops to limited water, ed advances in agricultural systems modeling 1. American Society of Agronomy, Inc., Madison, pp 39-51

Stockle CO, Kemanian AR (2009) Crop radiation capture and use efficiency: a framework for crop growth analysis. In V Sadras, D Calderini, eds, Crop physiology. Academic Press, Elsevier Inc.,pp 145-167

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen J, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res **14:** 1916-1923

Takei K, Takahashi T, Sugiyama T, Yamaya T, Sakakibara H (2002) Multiple routes communicating nitrogen availability from roots to shoots: a signal transduction pathway mediated by cytokinin. J Exp Bot **53:** 971-977

Tamang PL, Bronson KF, Malapati A, Schwartz R, Johnson J, Moore-Kucera J (2011) Nitrogen requirements for ethanol production from sweet and photoperiod sensitive sorghums in the southern high plains. Agron J **103:** 431-440

Taub D, Lerdau M (2000) Relationship between leaf nitrogen and photosynthetic rate for three NAD-ME and three NADP-ME C-4 grasses. Am J Bot **87:** 412-417

Teshome A, Baum B, Fahrig L, Torrance J, Arnason T, Lambert J (1997) Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in north Shewa and south Welo, Ethiopia. Euphytica **97:** 255-263

Thomas H, Ougham H, Canter P, Donnison I (2002) What stay-green mutants tell us about nitrogen remobilization in leaf senescence. J Exp Bot **53:** 801-808

Thornton P, Blairfish J, Wilson S (1991) Crop simulation modeling using a transputer-based parallel computer. Agric Syst **35:** 321-337

Tuinstra M, Grote E, Goldsbrough P, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in *Sorghum bicolor* (L.) Moench. Mol Breed **3:** 439-448

United States Department of Agriculture (2011) Web Soil Survey, http://websoilsurvey.nrcs.usda.gov/app/HomePage.htm.

United States Energy Information Agency (2009) Annual energy outlook. DOE Press, Washington DC, pp 1-56

United States Environmental Protection Agency (2010) Regulations of Fuels and Fuel Additives: Changes to Renewable Fuels Standard Program; Final Rule. Federal Register **5:** 14670-14722

van Oosterom EJ, Borrell AK, Chapman SC, Broad IJ, Hammer GL (2010) Functional dynamics of the nitrogen balance of sorghum: I. N demand of vegetative plant parts. Field Crops Res **115:** 19-28

van Oosterom EJ, Borrell AK, Deifel KS, Hammer GL (2011) Does increased leaf appearance rate enhance adaptation to postanthesis drought stress in sorghum? Crop Sci **51:** 2728-2740

van Oosterom EJ, Carberry P, Muchow R (2001) Critical and minimum N contents for development and growth of grain sorghum. Field Crops Res **70:** 55-73

Vanderlip R (1993) How a Sorghum Plant Develops. Kansas State University Press, Manhattan, pp 3-14

Vassilev SV, Baxter D, Andersen LK, Vassileva CG, Morgan TJ (2012) An overview of the organic and inorganic phase composition of biomass. Fuel **94:** 1-33

Vermerris W (2011) Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. J Integr Plant Biol **53:** 105-119

Wang D, Bean S, McLaren J, Seib P, Madl R, Tuinstra M, Shi Y, Lenz M, Wu X, Zhao R (2008) Grain sorghum is a viable feedstock for ethanol production. J Ind Microbiol Biotechnol **35:** 313-320

Wang MQ (1996) Development and use of the GREET model to estimate fuel-cycle energy use and emissions of various transportation technologies and fuels. DOE, Oak Ridge, pp 1-27

Wei N, Deng X (1992) COP9 - a new genetic-locus involved in light-regulated development and gene-expression in arabidopsis. Plant Cell **4:** 1507-1518

Weng J, Li X, Bonawitz ND, Chapple C (2008) Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. Curr Opin Biotechnol **19:** 166-172

White B (2006) Mississippi corn for grain and grain sorghum hybrid trials, Mississippi State University, Starkville, pp 1-19

Wilson J, Mertens D, Hatfield R (1993) Isolates of cell-types from sorghum stems - digestion, cell-wall and anatomical characteristics. J Sci Food Agric **63:** 407-417

Wolfe K, Gouy M, Yang Y, Sharp P, Li W (1989) Date of the monocot dicot divergence estimated from chloroplast DNA-sequence data. Proc Natl Acad Sci U S A **86:** 6201-6205

Wu X, Zhao R, Bean SR, Seib PA, McLaren JS, Madl RL, Tuinstra M, Lenz MC, Wang D (2007) Factors impacting ethanol production from grain sorghum in the dry-grind process. Cereal Chem **84:** 130-136

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol **8:** 103

Xu W, Subudhi P, Crasta O, Rosenow D, Mullet J, Nguyen H (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome **43:** 461-469

Yamoah CF, Varvel GE, Waltman WJ, Francis CA (1998) Long-term nitrogen use and nitrogen-removal index in continuous crops and rotations. Field Crops Res **57:** 15-27

Zhao YL, Dolat A, Steinberger Y, Wang X, Osman A, Xie GH (2009) Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel. Field Crops Res **111:** 55-64

Zhu X, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol **61:** 235-261

APPENDIX

Table 17: Monthly precipitation for College Station, Texas. All values are reported in mm. Data taken from NWS.

Precipitation (mm)	2008	2009
April	75	127.3
May	56.4	22.4
June	16.6	29.5
July	56.2	37.3
August	74.2	30.4
September	75.2	-
October	54.2	-

Table 18: Biomass accumulation (g m^{-2}) by 84G62 and TX08001. Values reported are for 2008 study.

Harvest Date	84G62	TX08001
April (15 DAE)	1.5	2.5
May (30 DAE)	15.2	28.1
June (60 DAE)	282.6	435.9
July (90 DAE)	540.3	1189.1
Aug (120 DAE)	76.2	1472.6
Sep (150 DAE)	-	794.0
Oct (180 DAE)	-	1031.7

Table 19: Total biomass yield for each harvest (g DW m^{-2}) for 84G62 and TX08001 during the 2008 growing season. For each harvest, N = 9 plants for each genotype.

Harvest Date	84G62	TX08001	p-value	Sig.
May (30 DAE)	17 ± 5	31 ± 6	3.8 E -5	***
Jun (60 DAE)	299 ± 165	466 ± 142	1.8 E -2	**
Jul (90 DAE)	1022 ± 70	1656 ± 491	5.2 E -4	***
Aug (120 DAE)	1168 ± 224	3128 ± 697	3.5 E -6	***
Sep (150 DAE)	-	3922 ± 512	-	-
Oct (180 DAE)	-	5084 ± 1716	-	-

Table 20: Total GLA (m^2 $plant^{-1}$) for each harvest for 84G62 and TX08001 during the 2008 growing season. For each harvest, N = 9 plants for each genotype.

Harvest Date	84G62	TX08001	p-value	Sig.
May (30 DAE)	0.03 ± 0.01	0.03 ± 0.01	0.19	-
June (60 DAE)	0.25 ± 0.11	0.33 ± 0.08	3.1 E -2	**
July (90 DAE)	0.27 ± 0.05	0.52 ± 0.11	3.2 E -5	***
Aug (120	0.19 ± 0.03	0.65 ± 0.11	3.5 E -7	***
Sep (150	-	0.62 ± 0.14	-	-
Oct (180 DAE)	-	0.30 ± 0.12	-	-

Table 21: Positions and Additive Effects of QTL for macro traits in SC170 x M35-1 F5 population.

Macro traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Total DW	1	10.82	10.4%	69.8	54,600,000	53,100,000	55,200,000
	9	12.48	13.9%	75.5	54,089,703	53,250,000	54,750,000
	9	11.04	11.4%	86.1	57860000	57,746,975	58,300,000
	10	-9.95	9.0%	72.1	56,032,596	55,500,000	56,700,000
StemDW	1	6.40	7.6%	53.6	26,300,000	24,026,964	49,450,000
	1	6.27	7.5%	69.8	54,600,000	53,100,000	55,300,000
	7	7.39	10.7%	61.9	58400000	57,900,000	59,300,000
	9	10.50	21.1%	80.9	55,800,000	55,200,000	56,100,000
	10	-6.84	8.6%	55.7	51,068,697	49,600,000	51,358,889
LeafDW	1	-3.34	32.2%	16.6	6,200,000	3,607,567	6,500,000
	1	-2.40	18.2%	27.4	11,610,000	11,229,963	13,150,000
	8	-1.57	7.1%	73.3	53200000	51,500,000	53,522,669
	9	2.05	12.1%	89.6	58,471,776	57,850,000	58,650,000
LAR	1	0.15	7.8%	104.8	68,214,539	67,100,000	70,720,000
	2	-0.21	15.8%	81.1	64,700,000	63,000,000	65,500,000
	8	0.14	7.0%	71.9	52,800,000	52,200,000	53,220,000
DTF	1	-2.22	26.1%	21.6	8,165,000	8,162,000	9,180,000

Table 21: Continued

Macro Traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
	4	0.96	5.0%	91.4	62800000	61,800,000	62,900,000
	5	1.09	6.6%	53.8	54,900,000	54,550,000	56,500,000
	9	1.15	7.1%	50.6	9,700,902	7,550,000	47,800,000
# Nodes	1	-0.51	25.2%	25.6	10,900,000	10,195,438	25,717,446
	8	-0.33	10.0%	71.4	52,600,000	50,938,116	52,715,418
	9	0.32	10.0%	88.5	58,200,000	57,984,457	58,976,796
Total GLA	1	-355.34	21.7%	21.6	8,165,000	7,800,000	10,800,000
	8	-242.79	10.2%	71.9	52,800,000	51,700,000	53,200,000
	9	210.88	7.6%	90.6	58,500,000	57,850,000	57,200,000
Stem L	1	28.20	23.3%	62.9	51,500,000	50,789,360	55,501,169
	7	40.35	49.0%	62.9	58,500,000	57,115,857	58,895,006
	9	18.76	20.7%	82.8	56,292,749	55,656,464	57,746,975

Table 22: Positions and additive variance of QTL identified for leaf traits in SC170 x M35-1 F5 population.

Leaf traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Leaf FW	1	-9.49	16.6%	12.4	2,300,000	2,150,000	2,400,000
	8	-7.84	11.9%	71.9	52,800,000	52,150,000	53,200,000
Flag Leaf Area	10	-21.30	16.5%	78.0	57,140,000	57,100,000	58,000,000
Leaf 2 Area	10	-31.99	19.6%	78.0	57,140,000	57,140,000	57,400,000
Leaf 3 Area	10	-43.73	29.6%	78.0	57,140,000	56,600,000	57,400,000
Leaf 4 Area	7	-30.05	15.7%	53.7	57,200,000	56,300,000	57,950,000
Leaf 5 Area	1	-25.82	9.1%	27.4	11,630,000	11,229,963	13,150,000
Leaf 6 Area	1	-44.75	22.7%	27.4	11,630,000	11,229,963	13,150,000
	7	-27.27	8.1%	61.3	58,350,000	57,830,000	58,500,000
	8	-31.46	11.1%	71.9	52,800,000	52,300,000	53,200,000
	9	27.07	8.0%	88.5	59,790,000	58,100,000	60,200,000

Table 22: Continued

Leaf traits				Peak		Bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Leaf 7 Area	1	-52.51	25.4%	21.6	8,165,000	7,870,000	10,800,000
	8	-43.33	17.1%	71.9	52,800,000	52,300,000	53,150,000
	9	30.54	8.5%	89.5	58,400,000	57,800,000	58,650,000
Flag Leaf Length	3	-3.20	13.1%	60.7	56,000,000	53,750,000	57,900,000
Leaf 2 Length	6	3.48	15.4%	65.4	50,221,734	48,900,000	52,200,000
	10	-3.43	14.9%	66.9	54,969,836	57,880,000	55,380,000
Leaf 3 Length	6	2.53	8.5%	63.7	48,861,138	48,200,000	51,600,000
	10	-3.94	11.1%	10.8	1,992,631	1,740,000	2,178,615
	10	-2.84	10.0%	68.1	55,123,255	55,100,000	55,400,000
	10	-3.08	11.1%	78.8	57,293,920	57,000,000	57,600,000
Leaf 4 Length	6	3.21	15.0%	65.4	50,221,734	48,400,000	51,500,000
	10	-2.68	9.4%	77.0	56,944,310	56,327,575	57,140,000
Leaf 5 Length	1	-4.32	20.9%	17.5	6,300,000	4,300,000	6,550,000
	6	2.46	7.2%	65.4	50,221,734	47,600,000	51,700,000
	9	2.53	7.4%	89.6	58,471,776	57,820,000	58,700,000
	10	-2.65	7.1%	9.0	1,732,221	1,500,000	1,750,000
Leaf 6 Length	1	-5.60	30.9%	27.4	11,630,000	11,229,963	13,140,000
	7	-2.60	6.6%	61.9	58,450,000	57,950,000	59,000,000
	8	-3.33	11.0%	71.9	52,800,000	52,250,000	53,200,000
Leaf 7 Length	1	-6.10	29.4%	18.5	6,400,000	6,220,000	7,870,000
	8	-3.96	12.2%	71.9	52,800,000	52,300,000	53,150,000
	10	-3.52	9.1%	6.7	1,628,671	1,550,000	1,700,000
Flag Leaf Width	1	0.25	10.1%	34.6	14,800,000	13,200,000	15,000,000
	8	-0.30	15.3%	45.0	6,350,000	5,900,000	6,600,000
Leaf 2 Width	8	-0.21	10.3%	46.2	6,650,000	6,300,000	6,800,000

Table 22: Continued

Leaf traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Leaf 3 Width	7	-0.20	9.8%	48.9	56,200,000	55,220,000	57,500,000
	10	-0.25	14.7%	78.0	57,140,000	56,600,000	57,580,000
	10	-34.78	20.7%	68.1	55,123,255	55,100,000	55,500,000
Leaf 4 Width	7	-0.17	8.6%	50.9	56,400,000	56,250,000	57,900,000
	10	-0.26	19.2%	68.1	55,123,255	55,000,000	55,400,000
	7	-36.22	17.8%	59.5	58,000,000	57,980,000	58,550,000
	8	-29.14	11.7%	71.9	52,800,000	52,150,000	53,300,000
	10	-27.64	10.2%	67.9	55,050,000	54,902,813	55,350,000
Leaf 5 Width	7	-0.22	12.7%	54.7	57,450,000	56,120,000	57,900,000
	8	-0.20	10.4%	46.2	6,500,000	6,300,000	6,800,000
	10	0.23	13.4%	68.1	55,123,255	55,000,000	55,350,000
Leaf 6 Width	7	-0.22	9.6%	51.9	56,900,000	56,600,000	57,850,000
Leaf 7 Width	1	-0.23	8.8%	18.9	6,530,000	2,900,000	7,870,000
	8	-0.34	19.5%	71.9	52,800,000	52,300,000	53,200,000
Flag Leaf SPAD	.	.	.	.	.	.	.
Leaf 2 SPAD	3	1.29	12.7%	87.6	67,512,498	65,300,000	68,800,000
Leaf 3 SPAD	10	-1.70	18.8%	93.8	60,300,000	60,200,000	60,500,000
Leaf 4 SPAD	10	-1.45	13.7%	93.8	60,300,000	60,223,765	60,500,000
Leaf 5 SPAD	7	1.04	8.5%	73.8	61,750,000	60,500,000	61,900,000
Leaf 6 SPAD	1	-1.22	10.9%	16.6	6,200,000	2,250,000	6,500,000
	2	1.34	13.7%	82.4	66,113,122	64,720,000	67,500,000
	3	1.10	9.1%	77.7	60,500,000	59,400,000	62,300,000
Leaf 7 SPAD	1	-1.74	17.2%	14.1	2,900,000	2,400,000	6,200,000

Table 23: Positions and additive variance of QTL identified for stem traits in SC170 x M35-1 F5 population.

Stem traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Stem FW	1	32.36	9.7%	69.8	54,650,000	50,789,360	55,501,169
	7	35.80	11.9%	61.9	58,500,000	57,115,857	58,895,006
	9	53.78	26.9%	81.9	55,800,000	55,656,464	57,746,975
Int 4 L	1	1.15	7.6%	74.1	55,300,000	54,706,017	57,032,744
	6	1.06	6.5%	63.7	48,861,138	48,186,751	51,018,662
	7	2.32	32.4%	61.3	58,350,000	57,801,919	58,395,121
Int 5 L	1	1.83	12.7%	75.1	55,600,000	55,501,169	56,252,115
	4	1.15	5.2%	68.5	54,629,843	53,450,000	55,650,000
	7	3.42	46.0%	61.9	58,450,000	58,310,534	58,895,006
	9	1.72	11.3%	82.8	56,292,749	55,656,464	58,471,776
Int 6 L	1	4.06	24.8%	75.1	55,600,000	54,706,017	55,780,288
	7	3.57	37.6%	61.9	58,400,000	58,310,534	58,895,006
	9	1.50	6.7%	77.8	54,791,810	54,328,355	55,656,464
Int 7 L	1	4.98	25.5%	74.1	55,300,000	54,706,017	55,780,288
	2	1.40	5.7%	77.6	62,486,170	61,953,374	64,766,587
	7	6.36	60.6%	62.9	58,395,121	58,395,121	58,895,006
	9	1.60	7.5%	82.8	56,292,749	56,292,749	57,984,457
Int 8 L	1	1.64	8.6%	2.6	840,000	382,869	920,559
	1	1.86	11.0%	74.1	55,300,000	54,706,017	55,780,288
	7	3.02	29.0%	62.9	58,600,000	58,395,121	58,895,006
	9	1.54	7.4%	82.8	56,292,749	56,292,749	57,984,457
Int 9 L	1	1.55	10.4%	74.1	55,300,000	54,706,017	55,780,288
	7	2.38	25.2%	63.9	58,600,000	58,395,121	58,895,006
	9	1.81	13.8%	82.8	56,292,749	56,292,749	57,984,457
Int 10L	1	1.12	7.4%	4.6	1,250,000	920,559	1,601,930
	1	1.38	11.1%	74.1	55,300,000	54,706,017	55,780,288
	7	1.92	21.9%	52.7	57,150,000	57,115,857	57,432,096
	9	1.63	14.8%	86.5	57,984,457	57,746,975	57,984,457
Int 4 D	7	-0.87	17.5%	61.9	58,400,000	58,310,534	58,895,006
	8	-0.67	10.2%	55.6	46,200,000	14,076,848	46,371,482

Table 23: Continued

Stem traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Int 5 D	7	-0.81	16.9%	61.9	58,400,000	58,310,534	58,895,006
	7	-0.63	10.3%	72.5	61,400,000	60,484,978	61,741,986
	8	-0.67	11.4%	54.6	46,000,000	9,757,739	46,371,482
Int 6 D	7	-0.86	20.2%	61.9	58,400,000	58,310,534	58,895,006
	7	-0.67	12.9%	73.5	61,745,000	60,484,978	61,741,986
	8	-0.59	9.6%	46.2	6,500,000	6,276,347	9,757,739
Int 7 D	7	-0.69	14.0%	53.7	57,200,000	57,115,857	57,801,919
	7	-0.77	17.5%	66.6	59,000,000	58,895,006	59,242,274
	8	-0.76	16.8%	46.0	6,500,000	6,276,347	6,833,807
Int 8 D	7	-0.63	12.9%	53.7	57,200,000	57,115,857	57,801,919
	7	-0.55	10.2%	73.5	61,750,000	61,050,067	61,791,728
	8	-0.67	15.5%	71.9	52,750,000	52,076,135	53,134,252
Int 9 D	8	-0.64	14.0%	72.7	53,100,000	52,531,494	53,134,252
Int 10 D	8	-0.55	10.5%	73.3	53,100,000	52,076,135	53,134,252

Table 24: Positions and additive variance of QTL identified for stem biomass composition traits in SC170 x M35-1 F5 population.

Composition traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Cellulose	1	0.81	14.3%	20.7	7,880,000	6,300,000	9,200,000
	1	0.81	14.6%	27.4	11,610,000	11,200,000	13,200,000
Lignin	1	0.69	36.4%	29.4	13,200,000	11,630,000	13,400,000
	1	-0.44	14.1%	61.9	51,300,000	50,900,000	51,700,000
	3	0.34	8.5%	47.5	12,700,000	12,500,000	15,600,000
Xylan	1	0.60	26.8%	25.6	10,900,000	10,737,638	11,610,000
	1	-0.37	9.9%	61.9	51,300,000	49,600,000	52,800,000
	3	0.31	7.3%	47.5	12,700,000	12,800,000	15,547,653

Table 24: Continued

Composition traits				Peak		bp bounds	
Trait	Chr	Add. Var.	R^2	cM	bp	left	right
Galactan	1	-0.02	10.1%	53.6	26,300,000	24,500,000	48,000,000
	5	0.02	9.8%	61.5	56,900,000	56,000,000	58,300,000
	7	-0.02	19.1%	52.7	57,150,000	56,600,000	57,432,096
	7	-0.02	19.7%	61.3	58,390,000	57,801,919	58,600,000
Arabinan	1	0.11	11.7%	25.6	10,900,000	9,300,000	11,652,471
	1	-0.16	7.9%	61.9	51,300,000	49,500,000	53,000,000
	2	-0.17	10.5%	77.6	62,486,170	61,700,000	64,626,219
	7	-0.23	18.6%	62.9	58,500,000	58,000,000	59,600,000
Protein	1	-0.22	11.1%	53.6	26,100,000	25,187,750	47,689,265
	1	-0.21	10.3%	58.6	49,900,000	48,745,130	50,789,360
	7	-0.20	9.9%	61.9	58,400,000	57,701,793	59,242,274
Sucrose	1	-1.91	25.2%	25.6	10,900,000	9,281,848	13,150,000
	1	1.19	9.0%	61.9	51,300,000	49,500,000	52,800,000
	2	1.12	8.6%	78.8	63,422,576	61,600,000	66,300,000
	7	1.02	7.1%	67.6	59,242,274	57,900,000	60,484,978
Ash	2	-0.18	9.7%	81.6	64,766,587	62,208,664	67,469,780
	10	0.17	8.9%	56.9	51,975,628	49,561,463	53,231,571
Extractives	1	-2.35	38.6%	29.4	13,100,000	11,600,000	13,400,000
	1	1.36	11.9%	58.6	49,900,000	49,500,000	50,900,000
	3	-1.05	7.8%	51.8	50,800,000	15,600,000	51,800,000
G cellulose	7	2.02	16.8%	62.9	58,550,000	58,000,000	59,010,000
	9	2.38	23.2%	81.9	56,000,000	54,800,000	56,600,000
G lignin	1	0.65	8.9%	69.8	54,600,000	53,100,000	55,900,000
	7	0.88	16.6%	64.6	58,900,000	58,400,000	59,010,000
	9	0.96	19.9%	80.9	55,700,000	54,200,000	56,740,209
G xylan	1	0.76	7.8%	53.6	26,200,000	24,000,000	48,000,000
	6	0.79	8.9%	45.1	38,500,000	35,766,835	44,000,000
	7	0.76	8.2%	61.9	58,450,000	57,900,000	59,600,000
	9	1.21	20.4%	69.0	51,880,862	51,381,303	52,685,980
	9	1.37	26.8%	80.9	55,700,000	54,839,052	55,900,000
	10	-0.86	9.9%	55.7	51,068,697	49,590,000	51,358,889

Table 25: Genes located within the *Ma2* locus. Gene names and positions are taken from Phytozome database. PFAM descriptions, where available, are taken from BioMart.

Gene Name	Gene Start	Gene End	PFAM Description, if any
Sb02g033310	67,981,670	67,984,776	MYND finger
Sb02g034500	69,077,130	69,079,184	Microtubule associated protein 1A/1B, light chain 3
Sb02g033360	68,016,887	68,019,033	
Sb02g034540	69,096,666	69,097,432	Plastocyanin-like domain
Sb02g033130	67,767,283	67,772,283	Cyclophilin peptidyl-prolyl cis-trans isomerase/CLD
Sb02g033460	68,086,393	68,087,447	
Sb02g033680	68,233,702	68,238,431	PCI domain
Sb02g034340	68,917,941	68,919,266	Ribosomal RNA adenine dimethylase
Sb02g034340	68,917,941	68,919,266	Methyltransferase domain
Sb02g033170	67,885,097	67,888,815	Serine carboxypeptidase
Sb02g034830	69,296,949	69,298,712	ATPase family assoc. with various cellular activities
Sb02g034160	68,660,128	68,661,546	
Sb02g033800	68,328,490	68,332,181	Arabidopsis proteins of unknown function
Sb02g034920	69,374,400	69,383,266	Leucine Rich Repeat
Sb02g034920	69,374,400	69,383,266	Patatin-like phospholipase
Sb02g035030	69,534,787	69,536,262	
Sb02g034070	68,575,352	68,580,722	Di-glucose binding within endoplasmic reticulum
Sb02g034070	68,575,352	68,580,722	Leucine rich repeat N-terminal domain
Sb02g034070	68,575,352	68,580,722	Leucine Rich Repeat
Sb02g034780	69,274,431	69,276,174	ATPase family assoc. with various cellular activities
Sb02g033940	68,469,058	68,478,741	RNA helicase (UPF2 interacting domain)
Sb02g034720	69,218,784	69,221,644	MatE
Sb02g034630	69,142,110	69,148,946	Methyltransferase TYW3
Sb02g034630	69,142,110	69,148,946	Kelch motif

Table 25: Continued

Gene Name	Gene Start	Gene End	PFAM Description, if any
Sb02g034960	69,458,085	69,462,279	Receptor family ligand binding region
Sb02g034630	69,142,110	69,148,946	Met-10+ like-protein
Sb02g033230	67,922,970	67,924,845	Aldo/keto reductase family
Sb02g034650	69,161,577	69,166,727	RNA recognition motif. (a.k.a. RRM, RBD, or RNP)
Sb02g033990	68,514,105	68,518,346	Tim17/Tim22/Tim23 family
Sb02g034800	69,283,171	69,285,239	
Sb02g034736	69,236,931	69,237,697	RNA recognition motif. (a.k.a. RRM, RBD, or RNP)
Sb02g034240	68,725,020	68,728,712	Protein phosphatase 2C
Sb02g034020	68,538,851	68,540,053	Chalcone and stilbene synthases, N-terminal domain
Sb02g034020	68,538,851	68,540,053	Chalcone and stilbene synthases, C-terminal domain
Sb02g034210	68,696,577	68,700,124	NB-ARC domain
Sb02g033750	68,279,283	68,281,384	
Sb02g033770	68,294,743	68,295,897	
Sb02g033690	68,243,464	68,249,907	GRAM domain
Sb02g033150	67,874,345	67,877,333	Zinc finger, C3HC4 type (RING finger)
Sb02g033590	68,176,892	68,178,609	2Fe-2S iron-sulfur cluster binding domain
Sb02g033270	67,956,826	67,958,227	Cathepsin propeptide inhibitor domain (I29)
Sb02g033270	67,956,826	67,958,227	Papain family cysteine protease
Sb02g034380	68,936,151	68,947,790	Metallopeptidase family M24
Sb02g034270	68,753,513	68,755,978	Plant mobile domain
Sb02g033830	68,370,267	68,371,629	Myb-like DNA-binding domain
Sb02g033540	68,151,510	68,156,982	RhoGAP domain
Sb02g034120	68,628,595	68,630,091	
Sb02g033500	68,109,887	68,110,624	Protein of unknown function (DUF3123)

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g034530	69,089,159	69,089,767	Protein of unknown function (DUF3123)
Sb02g034570	69,112,941	69,114,543	ATP synthase
Sb02g033400	68,051,110	68,054,801	Calcineurin-like phosphoesterase
Sb02g034100	68,610,029	68,611,976	
Sb02g034740	69,239,545	69,242,862	DEAD/DEAH box helicase
Sb02g034740	69,239,545	69,242,862	Helicase conserved C-terminal domain
Sb02g033906	68,442,073	68,442,444	MULE transposase domain
Sb02g033080	67,723,299	67,723,655	Auxin responsive protein
Sb02g033630	68,206,092	68,210,802	Ubiquinol-cytochrome C chaperone
Sb02g034320	68,851,980	68,853,044	Dof domain, zinc finger
Sb02g034680	69,181,492	69,183,057	BNR/Asp-box repeat
Sb02g033880	68,423,093	68,425,010	Ubiquitin-2 like Rad60 SUMO-like
Sb02g033880	68,423,093	68,425,010	Ubiquitin family
Sb02g033880	68,423,093	68,425,010	Ribosomal L40e family
Sb02g034050	68,568,768	68,571,587	NB-ARC domain
Sb02g034050	68,568,768	68,571,587	Leucine Rich Repeat
Sb02g033550	68,158,726	68,160,814	
Sb02g034640	69,150,066	69,154,180	Protein kinase domain
Sb02g034640	69,150,066	69,154,180	EF hand
Sb02g034890	69,345,514	69,352,745	PBS lyase HEAT-like repeat
Sb02g034890	69,345,514	69,352,745	HEAT repeat
Sb02g034435	68,990,110	68,991,148	
Sb02g033200	67,900,730	67,901,286	
Sb02g034490	69,068,354	69,073,043	Peptidase family C1 propeptide
Sb02g034490	69,068,354	69,073,043	Papain family cysteine protease
Sb02g034190	68,676,958	68,677,558	
Sb02g033740	68,275,961	68,277,238	Protein kinase domain
Sb02g034990	69,510,403	69,511,304	
Sb02g034725	69,224,977	69,225,402	
Sb02g033670	68,231,152	68,232,088	

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g034670	69,171,936	69,174,687	
Sb02g035020	69,530,262	69,531,234	Late embryogenesis abundant protein
Sb02g034480	69,066,811	69,068,835	NC domain
Sb02g035020	69,530,262	69,531,234	Transmembrane alpha-helix domain
Sb02g033600	68,183,962	68,185,002	Pollen allergen
Sb02g034710	69,206,479	69,206,850	
Sb02g034550	69,098,892	69,101,611	GRAS family transcription factor
Sb02g035060	69,577,753	69,582,259	TBC domain
Sb02g033890	68,428,747	68,431,109	Ubiquitin-2 like Rad60 SUMO-like
Sb02g033890	68,428,747	68,431,109	Ubiquitin family
Sb02g033890	68,428,747	68,431,109	Ribosomal L40e family
Sb02g033130	67,767,302	67,772,283	Cyclophilin peptidyl-prolyl cis-trans isomerase/CLD
Sb02g033515	68,116,586	68,118,193	
Sb02g034230	68,718,357	68,722,371	WD domain, G-beta repeat
Sb02g033570	68,166,723	68,168,186	
Sb02g033755	68,287,250	68,290,811	
Sb02g033700	68,251,429	68,253,190	
Sb02g034930	69,422,282	69,426,508	NB-ARC domain
Sb02g034930	69,422,282	69,426,508	Leucine Rich Repeat
Sb02g034775	69,264,700	69,272,294	Calponin homology (CH) domain
Sb02g034775	69,264,700	69,272,294	IQ calmodulin-binding motif
Sb02g034775	69,264,700	69,272,294	Armadillo/beta-catenin-like repeat
Sb02g034090	68,596,284	68,604,098	MatE
Sb02g034660	69,170,003	69,171,244	Protein phosphatase 2C
Sb02g034000	68,520,758	68,533,406	UvrD/REP helicase
Sb02g033160	67,880,826	67,884,898	PQ loop repeat
Sb02g033950	68,482,544	68,485,315	ATP synthase, Delta/Epsilon chain, beta-sandwich dom
Sb02g034630	69,142,110	69,148,933	Methyltransferase TYW3
Sb02g034630	69,142,110	69,148,933	Kelch motif

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g033430	68,072,427	68,075,203	Carbon-nitrogen hydrolase
Sb02g035010	69,528,254	69,528,841	Late embryogenesis abundant protein
Sb02g033780	68,296,525	68,301,648	Zinc finger, C3HC4 type (RING finger)
Sb02g034260	68,739,461	68,749,014	Protein kinase domain
Sb02g034260	68,739,461	68,749,014	Protein tyrosine kinase
Sb02g034080	68,585,011	68,591,394	Zinc finger, C3HC4 type (RING finger)
Sb02g034940	69,428,193	69,431,168	NB-ARC domain
Sb02g034940	69,428,193	69,431,168	Leucine Rich Repeat
Sb02g034630	69,142,110	69,148,933	Met-10+ like-protein
Sb02g034350	68,920,597	68,921,765	Ribosomal RNA adenine dimethylase
Sb02g034350	68,920,597	68,921,765	Methyltransferase domain
Sb02g033450	68,083,787	68,085,493	EamA-like transporter family
Sb02g033090	67,734,503	67,736,351	Eukaryotic rRNA processing protein EBP2
Sb02g034873	69,328,803	69,332,115	Myb-like DNA-binding domain
Sb02g033900	68,433,116	68,434,144	Phosphate-induced protein 1 conserved region
Sb02g034150	68,655,063	68,656,511	
Sb02g033840	68,381,416	68,398,030	Transglutaminase-like superfamily
Sb02g034040	68,548,925	68,555,112	
Sb02g033910	68,443,921	68,446,718	FAR1 DNA-binding domain
Sb02g033910	68,443,921	68,446,718	MULE transposase domain
Sb02g033910	68,443,921	68,446,718	SWIM zinc finger
Sb02g033560	68,162,346	68,165,715	Protein of unknown function (DUF3123)
Sb02g034490	69,068,354	69,073,144	Peptidase family C1 propeptide
Sb02g034490	69,068,354	69,073,144	Papain family cysteine protease
Sb02g034900	69,370,639	69,371,130	
Sb02g034310	68,793,665	68,798,184	Variant SH3 domain
Sb02g034310	68,793,665	68,798,184	SH3 domain
Sb02g033190	67,895,763	67,898,592	

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g034200	68,688,325	68,694,488	Domain of unknown function
Sb02g034675	69,175,887	69,179,238	BNR/Asp-box repeat
Sb02g034975	69,482,800	69,483,360	Ribosomal protein L44
Sb02g034770	69,257,614	69,264,069	
Sb02g034980	69,489,957	69,492,335	EamA-like transporter family
Sb02g034980	69,489,957	69,492,335	Triose-phosphate Transporter family
Sb02g034180	68,669,877	68,673,679	SBP domain
Sb02g033390	68,040,346	68,048,101	PHD-finger
Sb02g033390	68,040,346	68,048,101	Zinc finger, C3HC4 type (RING finger)
Sb02g033390	68,040,346	68,048,101	CUE domain
Sb02g033650	68,219,221	68,224,519	Permease family
Sb02g034650	69,161,577	69,165,297	RNA recognition motif. (a.k.a. RRM, RBD, or RNP)
Sb02g034840	69,301,513	69,303,136	ATPase family assoc. with various cellular activities
Sb02g033490	68,107,371	68,108,601	
Sb02g035050	69,560,896	69,563,496	EamA-like transporter family
Sb02g034620	69,136,763	69,139,051	PPR repeat
Sb02g033100	67,737,850	67,747,082	ACT domain
Sb02g033100	67,737,850	67,747,082	Protein tyrosine kinase
Sb02g033100	67,737,850	67,747,082	Protein kinase domain
Sb02g033220	67,911,109	67,913,578	Amidase
Sb02g034220	68,713,181	68,717,871	Plant protein of unknown function (DUF869)
Sb02g034300	68,788,763	68,792,886	
Sb02g034425	68,978,025	68,982,344	
Sb02g033340	67,996,703	68,000,880	NB-ARC domain
Sb02g033340	67,996,703	68,000,880	Leucine Rich Repeat
Sb02g034730	69,230,867	69,231,327	
Sb02g033710	68,254,058	68,256,297	Nucleoside diphosphate kinase
Sb02g033250	67,948,999	67,950,288	Pyridoxal-dep decarboxylase, pyridoxal binding dom

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g033820	68,358,079	68,361,020	
Sb02g033530	68,143,796	68,146,148	Ribosomal L15
Sb02g033210	67,905,226	67,907,493	Amidase
Sb02g034670	69,171,936	69,174,687	Endoribonuclease L-PSP
Sb02g034600	69,133,458	69,134,531	
Sb02g034520	69,086,556	69,088,645	DHHC zinc finger domain
Sb02g034295	68,774,419	68,783,436	Helicase conserved C-terminal domain
Sb02g034295	68,774,419	68,783,436	Helicase associated domain (HA2)
Sb02g034295	68,774,419	68,783,436	Domain of unknown function (DUF1605)
Sb02g034316	68,802,963	68,805,924	
Sb02g033870	68,417,073	68,421,209	
Sb02g033250	67,948,999	67,950,288	Pyridoxal-dep decarboxylase, C-term sheet domain
Sb02g033970	68,501,187	68,506,767	Cupin domain
Sb02g033930	68,461,028	68,464,585	Ras family
Sb02g033930	68,461,028	68,464,585	Miro-like protein
Sb02g034560	69,105,679	69,110,532	FAR1 DNA-binding domain
Sb02g034560	69,105,679	69,110,532	MULE transposase domain
Sb02g034560	69,105,679	69,110,532	SWIM zinc finger
Sb02g034580	69,114,934	69,121,702	3'-5' exonuclease
Sb02g034580	69,114,934	69,121,702	HRDC domain
Sb02g034140	68,641,482	68,642,930	
Sb02g033903	68,439,874	68,440,392	
Sb02g033440	68,075,970	68,081,103	OB-fold nucleic acid binding domain
Sb02g033440	68,075,970	68,081,103	tRNA synthetases class II (D, K and N)
Sb02g034850	69,304,589	69,306,461	ATPase family assoc. with various cellular activities
Sb02g034360	68,923,258	68,924,436	ATP-dependent protease La (LON) domain
Sb02g033660	68,225,294	68,228,531	Zn-finger in Ran binding protein and others

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g033790	68,326,064	68,327,876	Peroxidase
Sb02g034733	69,234,088	69,234,567	
Sb02g034880	69,334,858	69,342,435	Myb-like DNA-binding domain
Sb02g034790	69,279,808	69,281,496	ATPase family assoc. with various cellular activities
Sb02g033920	68,450,528	68,457,988	Malic enzyme, N-terminal domain
Sb02g033920	68,450,528	68,457,988	Malic enzyme, NAD binding domain
Sb02g033240	67,934,272	67,936,233	Glycosyl hydrolases family 16
Sb02g033240	67,934,272	67,936,233	Xyloglucan endo-transglycosylase (XET) C-terminus
Sb02g033660	68,225,294	68,228,583	Zn-finger in Ran binding protein and others
Sb02g034110	68,620,876	68,622,534	
Sb02g034870	69,324,928	69,326,463	Cytochrome P450
Sb02g035000	69,518,201	69,526,104	Thioredoxin
Sb02g035000	69,518,201	69,526,104	Endoplasmic reticulum vesicle transporter
Sb02g034280	68,764,886	68,765,837	
Sb02g034950	69,441,162	69,454,079	ABC transporter
Sb02g034950	69,441,162	69,454,079	ABC-2 type transporter
Sb02g034950	69,441,162	69,454,079	Plant PDR ABC transporter associated
Sb02g033730	68,267,823	68,269,100	Protein kinase domain
Sb02g034690	69,193,188	69,193,951	Protein of unknown function, DUF584
Sb02g034410	68,960,227	68,962,608	D-mannose binding lectin
Sb02g034410	68,960,227	68,962,608	PAN-like domain
Sb02g034410	68,960,227	68,962,608	Protein kinase domain
Sb02g034410	68,960,227	68,962,608	Protein tyrosine kinase
Sb02g034010	68,533,830	68,535,032	Chalcone and stilbene synthases, N-terminal domain
Sb02g034010	68,533,830	68,535,032	Chalcone and stilbene synthases, C-terminal domain
Sb02g033620	68,189,250	68,198,270	CG-1 domain
Sb02g033620	68,189,250	68,198,270	Ankyrin repeat

Table 25: Continued

Gene name	Gene start	Gene end	PFAM Description, if any
Sb02g034370	68,926,236	68,929,466	Thioredoxin
Sb02g033510	68,112,494	68,114,742	
Sb02g033410	68,056,276	68,064,812	Protein of unknown function (DUF1395)
Sb02g033960	68,494,229	68,498,330	Casein kinase II regulatory subunit
Sb02g034130	68,636,006	68,637,433	
Sb02g034820	69,292,406	69,293,246	ATPase family assoc. with various cellular activities
Sb02g033580	68,172,202	68,174,750	
Sb02g034060	68,573,231	68,575,071	C2 domain
Sb02g034640	69,150,066	69,153,592	Protein kinase domain
Sb02g034640	69,150,066	69,153,592	EF hand
Sb02g034480	69,066,757	69,069,147	NC domain
Sb02g033300	67,975,828	67,977,411	F-box domain
Sb02g033300	67,975,828	67,977,411	Protein of unknown function (DUF295)
Sb02g034440	69,007,297	69,013,542	SNF2 family N-terminal domain
Sb02g034440	69,007,297	69,013,542	Zinc finger, C3HC4 type (RING finger)
Sb02g034440	69,007,297	69,013,542	Helicase conserved C-terminal domain
Sb02g033620	68,189,250	68,198,270	IQ calmodulin-binding motif
Sb02g034370	68,926,236	68,929,466	Phosphoadenosine phosphosulfate reductase family

Table 26: Average component percentages and p-values for differences between stem and leaf component percentages from TX08001 tissue at 120 DAE.

Component	Average		p-value	Sig.
	Stem	Leaf		
Glucan	**29.5%**	23.2%	8.0E-27	***
Extractives	**27.7%**	16.3%	6.9E-14	***
Ash	6.9%	**12.5%**	5.5E-18	***
Protein	4.0%	**8.1%**	1.2E-16	***

Table 26: Continued

Component	Average		p-value	Sig.
	Stem	Leaf		
Lignin	**12.3%**	11.2%	1.4E-10	***
Starch	**-1.0%**	-3.0%	1.3E-08	***
Sucrose	**10.4%**	1.3%	3.1E-18	***
Xylan	16.0%	**17.1%**	3.4E-10	***
Galactan	0.8%	**0.9%**	1.0E-11	***
Arabinan	2.0%	**2.5%**	9.0E-09	***
Cellulose	**30.5%**	26.2%	4.7E-19	***

VITA

Name: Sara Nicole Olson

Address: MS 2128 TAMU
Room 417 Bio/Bio
College Station, TX 77843

Email address: saranolson@gmail.com

Education: B.S. Biology, Texas A&M University, 2006

Ph.D., Biochemistry, Texas A&M University, 2012