GENETIC AND PHENOTYPIC DIVERSITY IN SORGHUM FOR IMPROVEMENT AS A BIOFUEL FEEDSTOCK

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Seth Calder Murray

August 2008

GENETIC AND PHENOTYPIC DIVERSITY IN SORGHUM FOR IMPROVEMENT AS A BIOFUEL FEEDSTOCK

Seth Calder Murray, Ph. D.

Cornell University 2008

Sorghum [*Sorghum bicolor* (L.) Moench] is the sixth most important grain crop worldwide. In addition to grain varieties, diversifying selection in sorghum breeding has produced elite types for forage, silage, sugar, syrup, fiber, and broom production. Sorghum, however, has yet to undergo selection for whole plant energy production. Recent initiatives and improvements in grain, stem sugar, and lignocellulosic ethanol suggest that all plant parts may be valuable for energy production in the future. The primary focus of this dissertation was to understand the genetics of tradeoffs in yield and composition of grain, stem sugar and lignocellulosic biomass, and identify genetic diversity in grain and sweet sorghum types.

A biparental recombinant inbred line population, derived from a cross between a sweet and a grain sorghum, was used to identify QTL for non-structural carbohydrates in the first study and structural carbohydrates in the second. Including both studies, a total of 303 QTL were identified across 69 reported traits. From these studies, a major finding was that tradeoffs are minimal between carbohydrates under favorable conditions. This supports a sink limited model of whole plant energy production. Identified tradeoffs co-localized with QTL for height, flowering time and stand density suggesting some pleiotropic effects. Another important finding in the second study was that the genetic control of structural composition was found to be different between stem and leaf tissue. This was also true for protein, with QTL for grain, stem

and leaf crude protein failing to co-localize. These results suggest independent selection on tissues would be most effective. Additional products of these first two studies were a genetic map and NIRS equations for grain, stem and leaf tissue.

The third study used a diverse panel of 125 sweet, grain and landrace sorghums to examine genetic relationships within the sweet sorghums and between sweet and grain sorghums. Using principal coordinate analysis, three main populations within sweet sorghum were identified, syrup types, sugar and energy types, and amber types. These had some correspondence with grain races kafir, bicolor and caudatum. Association mapping detected four major height QTL, and one QTL for brix.

BIOGRAPHICAL SKETCH

Seth Calder Murray was born in Lansing, Michigan in January, 1980 to Mrs. Janet and Dr. Dennis Murray. His parents nurtured his interest in science at a young age, his mother through gardening and his father with his lab research in human infectious diseases. Seth's interests in plant improvement and genetics initiated at Okemos High School through exposure to greenhouse work and a video on biotechnology for improving salt tolerance in rice and growing plants in zero gravity. In his senior year, literature teacher Ms. Lisa Holmes gave him a book by Wes Jackson, "Becoming Native to This Place". It was in this book where he learned about breeding plants with the goal of environmentally friendly agriculture. Dr. Jackson developed ideas of a perennial polyculture that mimicked the natural systems of the prairie, for Seth this was an introduction to sustainable agriculture and a "paradigm shift" from the annual monoculture commonly observed and practiced.

Seth received his B.S. in Crop and Soil Sciences from Michigan State University in December of 2001 with minors in the Bailey Scholars Program, Biotechnology, and the Honors College. While at MSU he conducted an honors research project on the genetic inheritance of color and other traits in *Beta vulgaris* (sugar beet) with Dr. J. Mitchell McGrath. He also led efforts to start the Michigan State Student Organic Farm, which is now a 45 member Community Supported Agriculture (CSA) farm where students can gain hands on experience in agricultural production and research. During two summers Seth was a historic interpreter at Philmont Boy Scout Ranch where he instructed thousands of scouts on the differences in historic and modern agriculture and how science has been important in changing agriculture and agricultural productivity.

It was the combined interests in plant genetics and sustainable agriculture coupled with experiences in the Bailey Scholars Program and Boy Scouts that led Seth to his long term goal; to improve plants for an agriculture that is more economically and environmentally sustainable in the long term. As this can only be accomplished using genetic diversity, it was this desire, that led to Seth to work with Dr. Stephen Kresovich. Seth had wonderful and wide ranging experience in his PhD program and could not imagine doing anything with his life that shall have a more worthy, beneficial, or rewarding outcome.

I dedicate this dissertation to God, my supportive wife Andrea, my loving parents and sister, and all the mentors, friends, and collogues that I have met along the way.

ACKNOWLEDGMENTS

Dr. Stephen Kresovich accepted me as a student, allowed me the freedom and flexibility to design my own projects, and helped me understand the importance of genetic diversity. He continues to be an inspiration for his leadership, hard work, and collaborative ability to get along with most anyone. He understands how the university and scientific communities really operate and graciously imparts this knowledge to his students.

Without **Dr. William Rooney** (Texas A&M) this research would have been impossible. Dr. Rooney created the population that was used for much of this dissertation, planted the material, and helped to provide the labor, facilities, and equipment in Texas to evaluate it. Dr. Rooney also imparted a great deal of knowledge on sorghum, the Texas lifestyle, and how beef brisket *really* should taste.

My brilliant committee members, **Dr. Edward Buckler IV** and **Carlos Bustamante**. Dr. Buckler's program has been an inspiration for much of this work, his quest for high throughput is essential for the statistical power to conduct quantitative genetics. I would not have been able to complete this work without the knowledge of his lab, especially regarding NIRs and barcode technology. Dr. Bustamante was an inspiration for my desire to gain a greater statistical knowledge. His work and his lab has helped shape how I look at population genetic diversity and statistical analysis.

So many people at the Institute for Genomic Diversity have made meaningful contributions to education and productivity necessary to complete this dissertation. **Dr. Sharon Mitchell** facilitated much of this work, is the best proof reader I have ever met, and made sure we were not lacking for anything, most importantly was a softie with a great heart. **Dr. Alexandra Casa** introduced me to lab techniques and was the

most instrumental single person in training for the lab work at IGD. **Dr. Hong Sun** also deserves a significant amount of credit for my training in PCR techniques, and as helping me get through my introductory statistical classes. **Dr. Martha Hamblin** was a piller at IGD who imparted great wisdom on proper lab technique, population genetics, and the state of the democratic party, without her I would not understand population genetics. **Dr. Gael Pressior** with his original voluminous opinions and guidance of everything to do with science and life were often helpful and always amusing. His contribution to the analysis in this dissertation is noted. **Charlotte Acharya** provided excellent lab management, and was very helpful in facilitating important activities. **Wen Yen Zhu** also was excellent in her collection of data, and unforgettable in her humorous demeanor. All of the SSR data in the fouth chapter was collected by her. **Maria Gaudalupe Salas Fernendez** was helpful in processing of material in College Station 2005, more importantly, without her humor I never would have made it through that terrible field location. **Caitlin Mooney**, an undergraduate student who helped in the summer of 2006 contributed to much of the seed phenotyping and addition of genetic markers. **Alejandro Zamora's** contribution as a TA to Steve Kresovich's class was beneficial. Most importantly, **Patrick James Brown** and **Dr. Randall Jerome Wisser** must be acknowledged for my positive attitude towards graduate school and collaborative learning. Our discussions, in the lab and over beer at the **Big Red Barn**, were critical to my understanding of science and helped me avoid making many mistakes. They are truly "good guy's".

The Cornell Faculty, both past and present, have been key to making this institution so great to learn and be a graduate student in, especially in the plant sciences. **Dr. Mark Sorrell's** should be lauded for his contributions to graduate education. Dr. Sorrell's, through his two classes gave me a grounded education in plant breeding past and present. His interaction with the graduate students group was

greatly appreciated, and his program has served as a model for the integration of breeding and genetics. The faculty, staff, and students in the **Plant Breeding and Genetics Department** for accepting me, and allowing me distract myself and others with social activities.

The large amount of manual labor that was needed to harvest and strip the sorghum used in this research deserves many other thanks. At the Weslaco, TX TAMU Station; **Harold 'John' Drawe** for being my facilitator, **Gualberto 'Beto' Garza Jr.** for help with flowering time notes and field management, **Lori Gregg** for helping strip leaves and being good humored enough to get me through down there, **Samuel Perales** for harvest and strip help**, Eduardo 'Eddie' Hernandez**, **Robert 'Dusty' McGee**, and the **Cowley Sugar Mill** for letting me work after hours in a high security facility. At the College Station TAMU Sorghum Breeding Program led by **Dr. William L. Rooney; Karen Prihoda** for her station management, flowering time notes, critical tasks that get overlooked and good humor. **Delroy Collins** for his quality crew organization and crop management. **Stephen Labar** for being a immense help with so many things including equipment setup and repair, guidance on correct processing methods for the project, harvesting and striping, and fermentation technology par excellence. **Mr. Bill Lyles** for being worth 10 other people in harvesting, stripping and any other task, his historical stories taught me a lot. Many other graduate, college and high school students assisted in the harvesting and processing of the material **Katie Deno**, **Terry Felderhoff**, **Daniel Packer**, **Dr. Rafael Mateo**, **Dr. Les Kuhlman**, **Ryan Bading**, **Walter Dee Grant II**, and **Sean Grant.**

The genetic linkage map used in the first two studies was created in large part due to the efforts of **Dr. John Mullet, Dr. Patricia Klein, Dr. Arun Sharma, and Natalie Unruh** at Texas A&M University**.** The HPLC equipment was borrowed from

Dr. William Miller with assistance from **Susan Liou** and **Hi Ji Kim.** NIRS equipment and gracious help was provided by **Dr. Don Viands, Dr. Julie Hansen,** and **Jamie Neally** at Cornell. Sorghum grain **NIRS** equations for were greatly improved with a dataset from **Dr. Jeffrey Pedersen, Steve Masterson, and Dr. Kenneth Vogel at the USDA.**

Finally none of this work would be possible without the funding of vary expensive graduate students. For this I thank the **Department of Plant Breeding and Genetics**, **The Anonymous Donor** who funded The Institute for Genomic Diversity, and **The Department of Energy**/ **United States Department of Agriculture** for funding "Sorghum Biomass/ Feedstock Genomics Research for Bioenergy".

TABLE OF CONTENTS

Biographical sketch iii

Dedication v

Acknowledgements vi

Table of contents x

List of figures xi

List of tables xii

List of abbreviations xiii

Chapter 1. Introduction 14

Chapter 2. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem and grain nonstructural carbohydrates 29

Chapter 3. Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates 69

Chapter 4. Genetic diversity of and association mapping in sweet sorghums 111

Chapter 5. Conclusions 156

LIST OF FIGURES

Figure1.1 *Sorghum bicolor*: original range of wild and domesticated types 16

Figure 1.2 Sorghum domestication and crop improvement and the reduction of diversity 18

Figure 2.1. Genetic map derived from the Rio X BTx623 RIL population 52

Figure 2.2. Map positions and strength of QTL effects from data collected at three locations. 54

Figure 3.1. QTL positions in three locations 91

Figure 3.2. Yield, composition and theoretical ethanol for parents and eight selected RILs 104

Figure 4.1. PCoA plot of sweet sorghum panel genetic similarity 126

Figure 4.2. PCoA plot of the sweet sorghum panel combined with larger grain sorghum panel of Casa et al. (2008) 128

Figure 4.3 Genetic and physical position of chromosome 3 brix QTL and genetic markers 131

Figure 4.4. Relationship between brix and HPLC measured stem sugar for CS06 132

Figure 4.5 Relationship within and between brix and height across three locations 134

Figure 4.6. Results of population structure analysis using InStruct, Structure and PCoA 137

LIST OF TABLES

Table 2.1. Calibration treatment and statistics for grain NIRS 39

Table 2.2. Trait values for Rio X BTx623 recombinant inbred and parental lines at three locations 43

Table 2. 3. Trait heritability and variance component percentage attributable to genetic (G), environmental (E), genetic/environmental interaction (G*E) and other effects 45

Table 2.4. Pearson correlation coefficients of genetic and genetic by environmental effects 48

Table 2.5. Positions of QTLs identified by CIM and single marker analyses. 55

Table 3.1. Calibration treatment and statistics for leaf and stem NIRS 78

Table 3.2. Trait values for Rio X BTx623 recombinant inbred and parental lines at three locations 81

Table 3.3. Trait heritability and variance component percentage attributable to genetic (G), environmental (E), genetic/environmental interaction (G*E) and other effects 83

Table 3.4. Pearson correlation coefficients for corrected trait data 86

Table 3.5. Regrowth Trait values for Rio X BTx623 recombinant inbred 89

Table 3.6. Regrowth trait heritability and variance component percentage attributable to genetic (G), genetic/environmental interaction (G*E) and other effects 89

Table 3.7. Positions of QTLs identified by CIM and single marker analyses 92

Table 4.1. Sweet panel line names and associated information 117

Table 4.2. Polymorphism between lines with shared names 124

Table 4.3. TASSEL association of 378 markers across the genome for brix and height 136

Table 4.4. Markers with a significant p_value at 0.001 or in top 5% of Fst in each category 136

LIST OF ABBREVIATIONS

ADF	Acid detergent fiber
AFLP®	Amplified fragment length polymorphism
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
CIM	Composite interval mapping
CMS	Cytoplasmic male sterility
EST	Expressed sequence tag
HIF	Heterozygous inbred family
HPLC	High performance liquid chromatography
IM	Interval mapping
LD	Linkage disequilibrium
MAS	Marker assisted selection
NDF	Neutral detergent fiber
NIL	Near isogenic line
NIRS	Near infrared spectroscopy
PCoA	Principal coordinate analysis
QTL	Quantitative trait locus (Loci)
RILs	Recombinant inbred lines
SSR	Simple sequence repeat

CHAPTER 1

INTRODUCTION OF SORGHUM TYPES AND DIVERSITY FOR BIOFUEL OR INDUSTRIAL FEEDSTOCK USE

Thesis Statement

Sorghum [*Sorghum bicolor* L. Moench] is a water and nutrient efficient crop that can be grown throughout the entire US. In the developing world, landrace sorghums are often "dual-purpose" crops that are selected for both grain and stem biomass. In the developed world, crops are typically grown for a single product, primarily grain and forage, but also sugar and fiber. Elite sorghum varieties have been developed for grain, sugar or forage biomass production. Of all crops, sugar and forage sorghums have produced some of the highest biomass yields in the temperate US.

Recent increases in energy costs, coupled with demands for "green products" and concerns about greenhouse gases have increased demand for using crops as industrial and biofuel feedstocks. Crops, including sorghum, have not typically been improved for total energy yield and therefore little is known about crop potential, relevant traits or the genetics controlling this phenotype. Because all grain sorghums produce biomass and all biomass sorghums can produce grain, it appears likely that additional harvestable energy could be obtained by using biomass in addition to grain. To understand how to improve sorghums for energy production it is important to identify genetic variation available for improvement. This genetic variation is unlikely to be found in grain or forage sorghums alone.

Sorghum crop domestication and diversity

Sorghum [*Sorghum bicolor* L. Moench] is a hardy, drought tolerant and nutrient efficient C4 crop widely adapted throughout the world. Although there is some disagreement on timing, it is generally believed that sorghum was first domesticated from wild material in the area around Ethiopia and Sudan between 6000 and 3000 years ago (Kimber, 2000). These domesticated sorghums would have undergone selection for the loss of seed shattering and perhaps also for traits such as grain yield, grain taste, increased stem sugar and others. The domestication process would have reduced genetic diversity of cultivated sorghums when compared with wild sorghums.

It is believed that domesticated sorghum then radiated throughout Africa and Asia, undergoing selection for local adaptation and desirable phenotypes over time (Fig. 1.1 *Sorghum bicolor*: original range of wild and domesticated types). As domesticated sorghums dispersed, landraces were established from open pollinated populations by farmer selection for local adaptation and other desired traits. There was also likely outcrossing with local wild sorghums. Landraces from specific geographic regions tend to have similar phenotypes and these morphological similarities, specifically panicle and seed architecture, form the basis for classifying the five major sorghum races (bicolor, caudatum, guinea, kafir, and durra) (Harlan and deWet, 1972). The yellow arrows in Figure 1.1 show the hypothesized development and movement of each racial type from the area of domestication (Kimber, 2000).

Although it is impossible to determine the specific traits that were selected during the domestication of sorghum, many African and Indian landraces are currently grown for both biomass and grain. The biomass from these dual-purpose crops are for animal fodder or construction material and the grain is used for human consumption (Kelly et al., 1991; Bramel-Cox et al. 1995 Rai et al., 1999; House et al., 2000). These

landraces remain a promising source of novel traits and genes for sorghum biomass, grain, and ultimately, energy improvement.

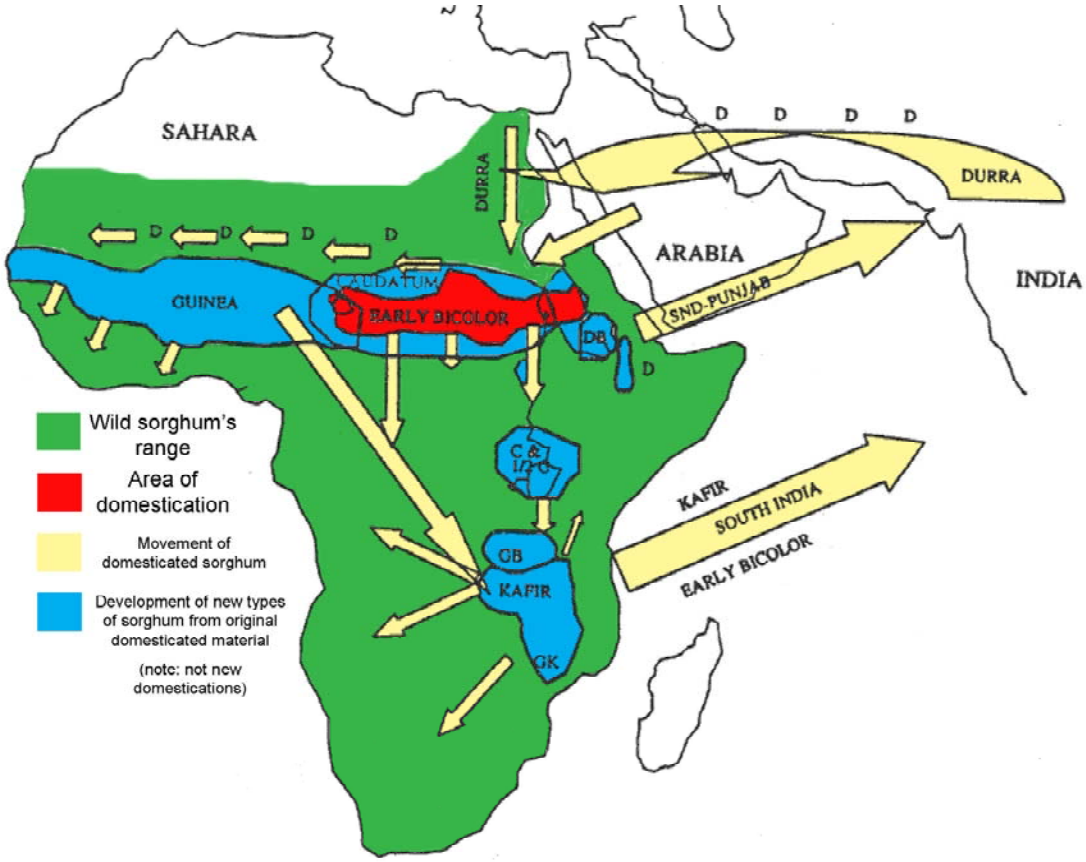


Figure1.1 *Sorghum bicolor*: original range of wild and domesticated types

The domestication of sorghum originally only used a small sample of the natural variation in the wild species occurring around the Sudan and Ethiopia (in red). As humans moved domesticated plants to new environments (yellow arrows) selection occurred for new phenotypes and local adaptation. This led to the creation of five main races (in blue). It is likely outcrossing with local wild sorghum (in green) occurred and continues to occur (Adapted from Kimber, 2000).

Sorghum was brought to the US very recently in evolutionary time. It is believed that Benjamin Franklin first brought broomcorn, tall sorghums with elongated panicle branches that are still used to make brooms, to the US in the late

1700's. Introduction of other sorghums for edible grain by African slaves, however, likely occurred earlier. In 1851 amber cane sorghums from China were introduced to the US for forage, syrup and sweet juicy stems good for chewing (Smith and Frederiksen, 2000). Syrup was produced by boiling off water and caramelizing the sweet juice from the stem (mostly comprised of sucrose). Sorghum syrup is still used today in cooking and as a condiment on biscuits. Throughout the late 1800s and early 1900s 'Chinese Amber' and other syrup producing sorghums were more important in the US than grain sorghums, producing millions of gallons of syrup annually (Winberry, 1980). During this time, a small number of landraces brought from South Africa by Leonard Wray dominated US sorghum production. These included varieties 'Sumac', 'Orange', 'Honey', 'White African', and 'Gooseneck'. These few landraces would be expected to have had far less genetic diversity than what might be found throughout Africa.

Until the advent of formal plant breeding and genetics, sorghum in the US was treated primarily as local landraces though some pure line selection occurred. There were distinctions, however, as to the type of sorghums: Milo's were sorghum selected and grown primarily for grain, sorgos or amber canes were those grown for syrup or forage. As knowledge of genetics and programs in plant breeding developed, an additional genetic bottleneck was created as elite sorghum types were selected from the landraces (Figure 1.2). These elite sorghum cultivars were inbred and further selected for a single product such as grain, stem sugar, forage, fuels, or brooms, unlike the dual-purpose landrace populations grown in Africa and Asia.

The first was the adoption of the combine in the 1940's, which necessitated dwarf plants that did not lodge. The second was the identification and use of cytoplasmic male sterility (CMS) for hybrid development in the mid 1950's (Smith and Frederiksen, 2000). These two improvements dramatically increased harvestable

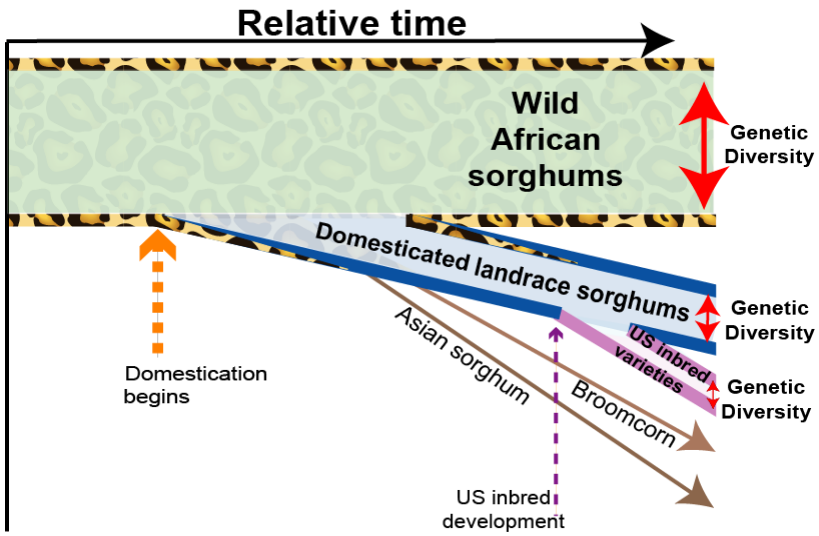


Figure 1.2 Sorghum domestication and crop improvement and the reduction of diversity

The figure shows a simplified and hypothetical model of how we might view reduction in sorghum diversity from domestication of wild material to produce landraces and the subsequent selection of elite inbred lines.

grain yields, the first by decreasing labor through mechanization the second by heterosis. As a crop, sorghum now has the sixth highest US acreage after corn, soybeans, wheat, alfalfa/forage, and cotton (USDA/NASS, 2008). Worldwide, sorghum is considered the fifth most important cereal crop after maize, rice, wheat and barley (FAO, 2007). Countries that produce over 1million metric tons (MT) of sorghum grain are, in order of the amount of grain produced: USA, Nigeria, India, Mexico, Sudan, Argentina, China, Ethiopia, Austrailia, Brazil, and Burkina Faso (FAO/ESS, 2008). These countries represent a mixture of developed nations that use mechanization to harvest grain for feed and developing nations that grow subsistence dual-purpose crops.

Since the introduction of hybrids in the US, large acreages of sorghum have been grown as grain. With a nutritional composition similar to corn sorghum is

primarily used as animal feed. (Hulse et al., 1980). Today in the US, seed companies use a limited number of inbred lines to create grain sorghum hybrids exclusively for grain as well as forage production. Forage sorghum (and sorghum-sudan grasses) acreage is not well tracked but forage seed outsells grain three to one. The use of hybrids has yet to be employed for sugar, syrup, broom or energy type sorghums as it has for grain and forage types. Although hybrid development would likely increase yields, it would also be expected to further decrease the genetic diversity for these types of sorghum grown in the US.

Breeding considerations in sorghum

For both breeding and genetics, it is important that the basic population genetic mechanisms of sorghum are understood. Sorghum is a self pollinating crop (87-100%) therefore lines are likely to be homozygous, but due to some outcrossing populations may be heterogeneneous with low mutational load (Pedersen and Toy 1998). When compared with a similar outcrossing species, *Zea mays* (corn) sequence data, sorghum has fourfold fewer polymorphisms and much more extensive LD (Hamblin et al., 2004; Hamblin et al. 2005).

Because sorghum is self pollinating, germplasm enhancement has primarily focused on pedigree selection methods of backcrossing and selecting from F_2 and bi-parental populations (Rooney and Smith 2000). Unlike corn, population improvement (recurrent selection) has not been a focus in sorghum breeding programs because of the difficulty in intermating large populations. A few programs have used population improvement methodology by incorporating recessive genes for genetic male sterility which allow a proportion of lines to be used as females without emasculation (House, 1985). The use of genetic male sterility differs from cytoplasmic male sterility used in hybrid production.

Pedigree selection, a primary breeding method for grain sorghums in the US, would be expected to reduce genetic diversity more than selection from a composite population. This reduction is caused by another genetic bottleneck, in addition to bottlenecks already experienced during domestication and development of elite material. For goals that will change whole plant architecture (such as cellulosic biofuels production, or whole plant energy production), landraces and wild plants should be a better source of breeding material than grain sorghums and population improvement procedures may be initially more successful than pedigree breeding.

Biofuels and rapid changes in lingo-cellulosic digestion

As fossil fuel (energy) prices have increased and global warming and environmental degradation have become concerns, there has been increased attention toward renewable biofuel production. This interest also extends to the plastic and chemical industries where manufacturers are seeking renewable feedstocks to replace petroleum. Current industrial processing for ethanol biofuel almost exclusively uses grain starch which is broken down to glucose, and fermented.Grain sorghum is the second most important source this starch after corn (NSP, 2007). Simple sugars found in sugarcane or sweet sorghum stem juice, the primary source of ethanol in Brazil and India, however, require less processing and could be a viable option for current US digestion systems (Rooney et al. 2007). On a per acre basis, grain starch and stem sugar are predicted to have far less energy potential than lignocellulosic biomass (Farrell et al., 2006). Although lignocellulosic ethanol production is less efficient than simple carbohydrate fermentation, there is much more biomass feedstock available. Until now it has not been technologically or economically feasible to build commercial cellulosic digestion plants and six are now under construction in the USA (USDOE, 2007)

Technological improvement for carbohydrate digestion and fermentation of grain, stem sugar, and lignocellulose is advancing rapidly. Yet, the ideal composition of grain, stem sugar, or lignocellulosic biomass to maximize biofuel production is currently unknown. It is reasonably assumed that carbohydrate content is linearly related to ethanol production but very few studies have carefully examined this relationship. Sorghum stem sugar content has been shown to demonstrate a good linear relationship with ethanol yield (Rooney et al. unpublished) but no such relationship exists between sorghum grain starch, which has a diverse composition, and ethanol yield (Wang et al. 2007). Recent research has suggested that the relationship between starch and ethanol yield may be adversely affected by protein or other seed component content and these would be targets to select against (Wu et al. 2007, Zhao et al. 2007). To my knowledge, nothing is known about the relationship between plant lignocellulose composition and ethanol production. The high cost of evaluating cellulose and the lack of a common digestion technology suggest that question will not be addressed in the near future.

NIRS technology

To improve the feedstock value of any crop, quantitative measurements on composition properties and/or end use value will be necessary. Although accurate measurements have been developed and are routinely used for many traits like grain starch, grain fiber, grain protein and biomass acid detergent fiber, neutral detergent fiber and acid detergent lignin these alone for large sample sizes needed in breeding and genetics programs are costly and time consuming (Van Soest, 1991; AOAC, 1990; Vogel et al., 1999) .

Near infrared spectroscopy (NIRS) is a technique that allows a rapid quantitative measurement of most organic and some inorganic compounds in samples.

NIRs has been shown to be an accurate, reliable and repeatable method across sorghum tissue types (Williams and Sobering, 1993; Hicks et al., 2002; deAlencarFigueiredo et al., 2006; Hooks et al., 2006; Murray et al. 2008a, 2008b). NIRS measures the reflectance over a range of near infrared wavelengths in less than one-minute per sample. Different types of chemical bonds present in each sample will absorb different wavelengths of light. Native software identifies the samples with the most informative spectra, which are then selected for measurement of composition trait of interest by wet chemistry. The software then creates calibrations between the composition trait of interest and wavelength absorption. These calibrations, after validation, may then be used on a variety of samples to quantitatively estimate the composition. The major advantages of this method are that it is high-throughput, fast, inexpensive (no consumables), and is open architecture allowing calibration to predict almost any tissue trait of interest. There have been studies on using NIRS for biofuel composition by CERES Inc. and the National Renewable Energy Lab (NREL) and the use of this technology is only likely to grow in the future (Hames et al., 2003; Kram, 2007).

Sorghum improvement for energy

Feedstocks will be best selected not only based on composition and digestibility, but also on yield, proximity to processing facilities (e.g. stem sugar near sugar mills), and economics of harvest and transportation. Agronomic experiments on planting, managing, harvesting, and post-harvest handling of biofuel crops have not been conducted. Additionally, experiments on whole plant digestion (sugar, grain and cellulose) have not been conducted. Until all of these many factors have been addressed it will be difficult to optimize composition for feedstock goals.

Another main challenge of breeding for energy production is to understand and utilize the appropriate genetic material for improvement. The bottlenecks of domestication, farmer selected landraces, pure lines and inbred grain varieties make it unlikely that sufficient diversity for biofuel improvement will be available in the most elite sorghum material. However, increased diversity may be found in elite material by looking across and using grain, stem sugar, and forage types together. Another strategy for increased diversity is to look to dual-purpose African landraces, material from the center of domestication, wild sorghums, and other sorghum species such as *S. halepense* and *S. propinquum*. It is important to note that material with more diversity usually has unfortunate drawbacks such as smaller grain, smaller panicles, seed shattering, smaller tillers or other undesirable traits.

Sorghum as a genetic model and molecular resources

In addition to its importance as a food, feed and forage crop, sorghum is used as a genetic model and has good genetic resources available. Sorghum is a close relative of corn (the primary starch feedstock crop) and sugarcane (the primary sugar feedstock crop) and is used as a genetic model for both. Sorghum has good gene conservation, co-linearity and a smaller genome then its relatives: Sorghum, diploid – up to 772 Mbp; Corn, diploid – up to 2716 Mbp; Sugarcane, anueo-hexapliod – up to 3605 Mbp; (Arumuganathan and Earle. 1991). Sorghum, as a C4 grass, may also be a good model for other biomass crops like *Miscanthus* and switchgrass, which have very large and complex genomes.

Many molecular genetic resources for sorghum are available. Two high density and numerous low density genetic maps (Menz et al. 2002, Bowers et al. 2003), a BAC based physical map (Bowers 2005), cross species comparative maps (Paterson et al. 2000), more than 240,000 EST sequences (NCBI, 2008), 1X coverage of the

methylation-filtered gene space (Bedell et al. 2005), and the entire *S. bicolor* genome sequence is now available for use (Phytozome, 2008). Many other genome metrics in sorghum have also been investigated such as gene organization, DNA sequence diversity, recombination rates, and linkage disequilibrium in a number of genomic regions have also been reported (Ilic et al., 2003; Hamblin et al., 2004; Casa et al., 2005, 2006; Hamblin et al., 2006; Casa et al., 2008). Over 40,000 sorghum accessions, wild, landrace, and elite types from a broad range of environments, are available from the U.S. National Plant Germplasm System (USDA-ARS, 2008). Taken in whole, these resources and studies have and will continue to help elucidate the genetic landscape of sorghum, how diversity is partitioned and how to improve the crop for food, fiber, and fuel.

REFERENCES

AOAC (Association of Official Analytical Chemists) 1990a. In K. Helrich (ed.) Official methods of plant analysts (15th ed.). AOAC, Arlington, VA

Arumuganathan, K., and E.D. Earle. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9:208-218.

Bedell J.A., M.A. Budiman, A. Nunberg, R.W. Citek, D. Robbins, J. Jones, E. Flick, T T. Rholfing, J. Fries, K. Bradford, J. McMenamy, M. Smith, H. Holeman, B.A. Roe, G. Wiley, I.F. Korf, P.D. Rabinowicz, N. Lakey, W.R. McCombie, J.A.Jeddeloh, R.A. Martienssen. 2005. Sorghum genome sequencing by methylation filtration. PLoS Biol. 3(1):e13.

Bowers J.E., C. Abbey, C. Anderson, C. Chang, X. Draye, A.H. Hoppe, R. Jessup, C. Lemke, J. Lennington, Z. Li , Y.R. Lin, S.C. Liu, L. Luo, B.S. Marler, R. Ming, S.E. Mitchell, D. Qiang, K. Reischmann, S.R. Schulze, D.N Skinner, Y.W. Wang, S. Kresovich, K.F. Schertz, and A.H. Paterson. 2003. A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367-386.

Bowers, J.E., M.A. Arias, R. Asher, J.A. Avise, R.T. Ball, G.A. Brewer, R.W. Buss, A.H. Chen, T.M. Edwards and J.C. Estill. *et al.* (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proc Natl Acad Sci USA* **102**: 13206–13211.

Casa, A.M.,S.E. Mitchell, M.T. Hamblin, H. Sun, J.E. Bowers, A.H. Paterson, C.F. Aquadro, S. Kresovich. 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor. Appl. Genet. 111:23-30.

Casa, A.M., S.E. Mitchell, J.D. Jensen, M.T. Hamblin, A.H. Paterson, C.F. Aquadro, S. Kresovich. 2006. Evidence for a selective sweep on chromosome 1 of cultivated sorghum. Plant Genome 1:S1-S15.

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community resources and strategies for association mapping in sorghum. Crop Sci 48:30-40.

Bramel-Cox, P. J., K. A. Kumar, J. D. Hancock, and D. J. Andrews, 1995: Sorghum and millets for forage and feed. In: D. A. V. Dendy (ed.). Sorghum and Millets Chemistry and Technology, 325-364. American Assoc. Cereal Chemists, Inc., St Paul.

Deaville, E.R., and P.C. Flinn.(2000). Near-infrared (NIR) spectroscopy: an alternative approach for the estimation of forage quality and voluntary intake. In: D.I. Givens, E. Owen, R.F.E. Axford and H.M. Omed, Editors, *Forage Evaluation in Ruminant Nutrition*, CAB International, pp. 301–320.

deAlencarFigueiredo, L.F., F. Davrieux, G. Fliedel, J.F. Rami, J. Chantereau, M. Deu, B. Courtois, and C. Mestres. 2006. Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. J. Agric. Food Chem. 54:8501-8509.

FAO. 2007. FAO Statistical Yearbook 2005 -2006. Vol. 1/2 (Food Agric. Org., U.N., Rome). P.82.

FAO/ESS. 2005. Major food and agricultural commodities and producers. [Online]. Available at http://www.fao.org/es/ess/top/commodity.html?lang=en&item=83&year=2005 (verified 05 May. 2008).

Farrell, A.E., R.J. Plevin, B.T. Turner , A.D. Jones, M. O'Hare, D.M. Kammen. 2006. Ethanol can contribute to energy and environmental goals. Science 27: 506-8.

Hamblin, M. T., S. E. Mitchell, G. M. White, J. Gallego, R. Kukatla, R. A. Wing, A. H. Paterson, and S. Kresovich. 2004. Comparative population genetics of the panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics* 167: 471–483.

Hamblin, M.T., M.G.S. Fernandez, A.M. Casa, S.E. Mitchell, A.H. Paterson, and S. Kresovich. 2005. Equilibrium processes cannot explain the high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genetics 171:1247–1256.

Hamblin, M.T., A.M. Casa, H. Sun, S.C. Murray, A.H. Paterson, C.F. Aquadro, and S. Kresovich. 2006. Challenges of detecting directional selection after a domestication bottleneck: lessons from sorghum bicolor. Genetics 173:953-964.

Hames, B.R., S.R. Thomas, A.D. Sluiter, C.J. Roth and D.W. Templeton. 2003. Rapid biomass analysis—new tools for compositional analysis of corn stover feedstocks and process intermediates from ethanol production. Applied Biochemistry and Biotechnology **105:**5–16.

Harlan, J.R., and J.W.J.deWet. 1972. A simplified classification of sorghum. Crop Sci. 12:172–176.

Hicks, C., M.R. Tuinstra, J.F. Pedersen, F.E. Dowell, and K.D. Kofoid. 2002. Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica 127:31-40.

Hooks, T., J.F. Pedersen, D.B. Marx, and K.P.Vogel. 2006. Variation in the US photoperiod insensitive sorghum collection for chemical and nutritional traits. Crop Sci. 46:751-757.

House, L. R., 1985: A Guide to Sorghum Breeding, 2nd edn. ICRISAT, *Patancheru, AP, India.*

House, H.R, M. Gomez, O.S. Murty, Y. Sun, and B.N. Verma. 2000. Development of Some Agricultural Industries in Several African and Asian Countries. p. 131-190. *In* Smith et al. (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, Inc., New York.

Hulse, JH, Laing EM, Pearson OE (1980) Sorghum and Millets: Their Composition and Nutritive Value. Academic Press, New York. 997 pp.

Ilic, K., P.J. SanMiguel, and J.L. Bennetzen. 2003. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Proc. Natl. Acad. Sci. USA. 100:12265-12270.

Kelly, T. G., and P. Parthasarathy Rao, and T. S. Walker. 1991. The relative value of cereal straw fodder in the semi-arid tropics of India: implications for cereal breeding programmes at ICRISAT. In: K. Dvorak (ed.). Social Science Research for Agricultural Technology Development. Proc. Int. Inst. Trop. Agric .-Rockefeller Foundation Workshop, Ibadan, Nigeria, 2-5 October 1990, CABI.

Kimber, C., 2000 Origins of domesticated sorghum and its early diffusion to India and China, pp. 3–98 in *Sorghum*, edited by C. W. SMITH and R. A. FREDERIKSEN. John Wiley & Sons, New York.

Krakowsky, M.D., M. Lee, and J.G. Coors. 2005. Quantitative trait loci for cell wall components in recombinant inbred lines of maize (Zea mays L.) I: Stalk tissue. Theor. Appl. Gen. 111:337-346.

Krakowsky, M.D., M. Lee, and J.G. Coors. 2006. Quantitative trait loci for cell wall components in recombinant inbred lines of maize (Zea mays L.) II: Leaf sheath tissue. Theor. Appl. Gen. 112:717-726.

Kram, J.W. 2007. The Need for Speed: Rapid Biomass Analysis Makes Better Breeding Possible. Biomass magazine. [Online]. Available at http://biomassmagazine.com/article.jsp?article_id=1232 (verified 05 May. 2008).

Menz, M.A., R.R. Klein, J.E. Mullet, J.A. Obert, N.C. Unruh, and P.E. Klein. 2002. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Molecular Biology Reporter 48:483-499.

Ming, R., S.C. Liu, P.H. Moore, J.E. Irvine, and A.H. Paterson. 2001. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 11:2075-2084.

National Sorghum Producers. 2007. [Online]. Available at http://www.sorghumgrowers.com (verified 12 Nov. 2007).

NCBI, National Center for Biotechnology Information. 2008. www.**ncbi**.nlm.nih.gov/ verified Feb. 19, 2008.

Paterson, A.H., J.E. Bowers, M.D. Burow, X. Draye, C.G. Elsik, C.X. Jiang, C.S. Katsar, T.H. Lan, Y.R. Lin, R. Ming, and R.J. Wright. 2000. Comparative Genomics of Plant Chromosomes. The Plant Cell. 12:1523-1540.

Pedersen, J.F., .J. Toy. 1998. Natural outcrossing of sorghum and sudangrass in the central great plains, Crop Sci. 38: 937–939.

Phytozome. 2008. www.phytozome.net/ verified Feb. 19, 2008.

Rai, K.N., D.S. Murty, D.J. Andrews, P.J. Bramel-Cox. 1999. Genetic enhancement of pearl millet and sorghum for the semi-arid tropics of Asia and Africa. Genome 42:617–628.

Rooney, W.L., and C.W. Smith. 2000. Techniques for developing new cultivars. p. 329–347. *In* C.W. Smith and R.A. Frederiksen (ed.) Sorghum: Origin, history, technology and production. John Wiley & Sons, New York.

Rooney, W.L., J.Blumenthal, B. Bean, J.E. Mullet. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod. Bioref. 1:147–157.

Smith, C.W., and R.A. Frederiksen. 2000. History of cultivar development in the United States: From "Memoirs of A.B. Maunder—Sorghum breeder." p. 191–223. *In* C.W. Smith and R.A. Frederiksen (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, New York.

USDA/NASS. (2008) Crop production 2008 summary. Cr Pr 2-1 (08). U.S. Department of Agriculture, National Agricultural Statistics Service, Washington, D.C.; 2001.

Van Soest, P.J., J.B. Robertson, and B.A. Lewis. 1991. Methods for dietary fiber, neutral detergent fiber, and non-starch polysaccharides in relation to animal nutrition. J. Dairy Sci. 74:3583–3597.

Vietor, D.M., F.R.Miller. 1990. Assimilation, partitioning, and nonstructural carbohydrate in sweet compared with grain sorghum. Crop Sci. 30:1109-1115.

Vogel, K.P., J.F. Pedersen, S.D. Masterson, and J.J. Toy. 1999. Evaluation of a filter bag system for NDF, ADF, and IVDMD forage analysis. Crop Sci. 39:276–279.

Wang, S., C.J. Basten, and Z.B. Zeng. 2007. Windows QTL Cartographer 2.5. Available at Department of Statistics, North Carolina State University, Raleigh, NC. [Online]. Available at http://statgen.ncsu.edu/ (verified 05 May. 2008)

Williams, P.C., and D.C. Sobering. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. J. Near Infrared Spectrosc. 1:25-33.

Winberry, JJ. 1980. The Sorghum Syrup Industry: 1854-1975. Agricultural History 54(2): 343-352.

Wu, X., R. Zhao, S.R. Bean, P.A. Seib, J.S. McLaren, R.L. Madl, M. Tuinstra, M.C. Lenz, and D. Wang. 2007. Factors impacting ethanol production from grain sorghum in the dry-grind process. Cereal Chem. 84:130-136.

Zhao, R., S. Bean, and D.Wang. 2007. The impacts of protein on grain sorghum ethanol fermentation efficiency. ASABE Meeting Presentation Paper:076158.

USDA-ARS, NGRP. 2008. National Genetic Resources Program. Germplasm Resources Information Network (GRIN). Nat. Germplasm Resour. Lab., Beltsville, MD. [Online Database] Available at http://www.ars-grin.gov [verified 5 May. 2008].

USDOE. 2006. Breaking the biological barriers to cellulosic ethanol: A joint research agenda. DOE/SC 0095. USDOE Office of Science and Office of Energy Efficiency and Renewable Energy, Washington, DC. [Online]. Available at http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml (verified 14 Nov. 2007)

USDOE. 2007. DOE Selects Six Cellulosic Ethanol Plants for Up to $385 Million in Federal Funding. [Online]. Available at http://www.energy.gov/news/4827.htm (verified 05 May. 2008)

CHAPTER 2

GENETIC IMPROVEMENT OF SORGHUM AS A BIOFUEL FEEDSTOCK: I. QTL FOR STEM SUGAR AND GRAIN NONSTRUCTURAL CARBOHYDRATES[1]

Abstract

Genetic improvement of sorghum [*Sorghum bicolor* (L.) Moench] has traditionally focused on a single non-structural carbohydrate; either grain starch, or stem sugar. Sorghum starch and sugar may both be used as feedstocks for biofuel production. To investigate genetic tradeoffs between grain and stem sugar, a population derived from sweet sorghum cultivar 'Rio' and grain sorghum 'BTx623' was evaluated for 28 traits related to grain and stem sugar yield and composition. Across three environments, a total of 145 QTL were identified. Tradeoffs identified between grain and stem sugar yield QTL co-localized with height and flowering time QTL. Most importantly, QTL were identified that increased yield and altered the composition of stem sugar and grain without pleiotropic effects. For example, a QTL on chromosome 3 that explained 28% of the genetic variance for stem sugar concentration did not co-localize with any grain QTL. These results suggest that total non-structural carbohydrate yield could be increased by selecting for major QTL from both grain and sweet sorghum types. We conclude that altering grain and stem sugar genetic potential for yield traits should lead to greater feedstock improvement than altering composition traits.

[1] Murray, S.C., A. Sharma, W.L. Rooney, P.E. Klein, J.E. Mullet, S.E. Mitchell, and S. Kresovich (under review) Genetic improvement of sorghum as a biofuel feedstock I: quantitative loci for stem sugar and grain nonstructural carbohydrates. Crop Science.

Introduction

There is renewed interest in using sugars derived from agricultural crops as feedstocks for biofuel production (U.S. DOE, 2006; Somerville, 2007; Farrell et al., 2006), large-scale manufacture of more complex molecules (Lichtenthaler and Peters, 2004), and *in planta* syntheses of harvestable bio-molecules (*i.e.*, nutraceuticals) (Mazur et al., 1999; Mohanty et al., 2002). Improving a species for use as biofuel feedstock requires a change in perspective; crops must be regarded as living systems for capturing and storing energy rather than simply as a sole source of food, feed, or fiber products. This changes the basic biological question from "how much of a crop's energy can be converted into food?" to "how can we maximize the total useable energy that can be produced and stored throughout the growing season?". Furthermore, characterization and quantification of environmental and post-harvest energy degradation in addition to genetics will be crucial for developing economically feasible biofuel feedstocks.

Sorghum [*Sorghum bicolor* (L.) Moench], a hardy drought tolerant and nutrient efficient C 4 grass, is widely adapted throughout the world. Sorghum is also closely related to other potential biofuel crops such as sugarcane, the principal sugar feedstock, and maize, the most important starch feedstock. Sweet sorghums accumulate up to 25% sugar, 1.4 to 2.7 times more whole-plant nonstructural carbohydrates than grain sorghums, in the parenchyma of juicy stalks (Vietor and Miller, 1990; Ming et al., 2001). The physiological mechanism of sugar accumulation, however, appears to differ between sorghum and sugarcane (Tarpley and Vietor 2007). Additionally, many of the enzymes associated with sugar accumulation in sugarcane (i.e. sucrose phosphate synthase and invertase) do not appear to play major roles for sugar accumulation in sorghum (Lingle et al. 1987; Tarpley et al. 1994). Sweet sorghums are more water and nutrient efficient than sugarcane and maize and can be

grown over a wide area of the US (Jackson et al., 1980; Hallam et al., 2001). To date, less than 50 US elite inbred sweet sorghum cultivars have been released. Among these cultivars are syrup types, lines selected for high quality and quantity of stem juice sugar to be boiled into syrup, and a few sugar types, lines selected for high sucrose yield only (Jackson et al., 1980). Because both sweet sorghum types were selected specifically for extractable stem sugar, these lines generally produce small amounts of grain with undesirable characteristics such as small seed and high tannin content. It is unclear whether there is a genuine physiological tradeoff between high stem sugar production and reduced grain yield or if the relationship is simply due to the fact that sweet sorghum cultivars have never been improved for grain traits. In sorghum, the mode of inheritance of increased stem sugar depends on the cross and has been shown to be either additive or dominant (Schlehuber, 1949; Clark, 1981). Genetic mapping experiments for sorghum stem sugar have identified one to a few loci (Natoli et al. 2002; Bian et al. 2006; Ritter 2007) but the small variance explained suggests that additional loci with complex interactions may also be involved.

In the U.S., most ethanol is produced from maize grain starch which is enzymatically converted to glucose and then fermented. The same process is used for grain sorghum; in fact, sorghum is the second most commonly used grain in ethanol production in the U.S. (National Sorghum Producers, 2007). In Brazil, ethanol is produced from sucrose extracted from sugarcane. This process is simpler as it eliminates the need for enzymatic degradation of starch and requires less processing. Sweet sorghum juice could certainly be used in a similar system as sugarcane. In addition, harvesting grain from sweet sorghum provides another important source of fermentable carbohydrates for conversion to ethanol (Jackson et al., 1980; Kresovich and Henderlong, 1984). The “dual-purpose” nature of sorghum raises the possibility that energy production could be maximized by concurrent improvement of both grain

and stem sugar yields. Because elite sorghum cultivars have traditionally been bred for a single use (*i.e.*, grain for human or animal consumption, stem sugar for syrup production, or forage/silage for animal feed) little is known about the physiological tradeoffs of simultaneously improving both grain and sugar traits.

In this study, we investigated the potential of developing high starch grain sorghums with increased stem sugar for the ultimate goal of improving sorghum as a dedicated feedstock crop. To accomplish this we identified and mapped quantitative trait loci (QTL) controlling yield and composition of sugar in the stem as well yield and composition of starch, fat, protein, fiber and phosphorus in grain. Specifically, we were interested in determining: (1) the genetic tradeoffs between grain and stem sugar yield; (2) the genetic tradeoffs between grain composition and yield of stem sugar or grain; and (3) if there were significant effects of harvest date and post-harvest handling on the production of fermentable carbohydrates.

Materials and Methods

Plant Material and Plot Design: A mapping population consisting of 176 $F_{4:5}$ recombinant inbred lines (RILs) from a cross between cultivar 'Rio', a high biomass sweet sorghum (Broadhead, 1972), and 'BTx623', an elite inbred grain sorghum (Frederiksen and Miller, 1972), was phenotyped for grain yield, grain composition and stem sugar traits. The RILs and the parental lines were planted during the summer growing season in 2005 at Weslaco, TX (WE05), and in 2005 at College Station, TX (CS05). In 2006, 165 $F_{5:6}$ RILs were planted in College Station (CS06) from self-pollinated seed produced at CS05. In each location, two replicates of 3.05 meter rows were planted in a randomized complete block design. Seeds were planted at a rate of 160 000 plants ha^{-1} with either 76 cm (CS05, CS06) or 102 cm row spacings (WE05). The WE05 site had five centimeters of pre-plant irrigation and received less than two

centimeters of rainfall for the remainder of the growing season. CS05 emerged based on available soil moisture; 43 cm of rain fell during the growing season primarily during flowering. In CS06, the total rainfall of 36 cm was distributed evenly throughout the growing season.

Phenotypic Measurement of Field Traits: In total, we measured 28 traits of agronomic and quality importance (Table 2.2). Plant height was measured either in the field (WE05) or at harvest, due to high lodging (CS05, CS06). Stand density and tillering were each estimated on a one to ten scale in the harvested area of each row. Flowering time was measured as 50% plot anthesis (WE05 and CS05).

Harvests were staggered over 16 days (WE05), 14 days (CS05), and 11 days (CS06) due to the volume of work and logistics of labor and equipment. Harvest date was used as one of the cofactors in subsequent statistical analyses to control for experimental error. From each row, a random meter of plant material was harvested in the morning from a central stand by cutting the plants within three centimeters of the ground. Each cut row was then bundled in clear plastic sheeting and taken to a shaded central processing facility within two hours of harvest. The bundled row was stripped into panicles, and stems and each was weighed (panicle fresh weight, stem fresh weight traits, respectively). Strip date was also recorded because only about half the plants could be processed on the day they were harvested. At CS05 and CS06, replications were harvested simultaneously; at WE05 harvest was sequential meaning harvest date and replicate sources of error would be nested for statistical analysis.

Fresh stem tissue was crushed in a three roller sugar mill (WE05, CS06), or a potato starch drier (CS05) to extract the juice. At this time, juice volumes and weights were recorded, brix was measured using a hand held refractometer (Atago U.S.A. Inc., Bellevue, WA), and aliquots of juice (15 mL per sample) were frozen for HPLC analysis of sugars.

For each experimental unit, random sub-samples of grain panicles, and pressed stems were collected. Sub-samples were then weighed and dried in a greenhouse (WE05) or in a forced air drier at 38°C (CS05, CS06). Dry stem and panicle sub-samples were weighed, and panicles were threshed; stem dry weight, panicle dry weight and grain dry yield were calculated from these measurements. All dried material and frozen juice samples were then shipped to Ithaca, NY for further analysis.

Measurement of Sugar Traits: To determine sugar composition and quality, frozen juice was evaluated by high performance liquid chromatography (HPLC) based on the instrument manufacturer's instructions (Dionex Corporation, 2006). Juice samples were thawed, lactose was added as an internal standard, samples were diluted 250X in water and filtered through a 0.45µm filter (PALL acroprepTM96 , Pall Life Sciences, Ann Arbor, MI). Samples were analyzed on a Dionex HPLC with EP50 gradient pump, AS40 autosampler, ED40 HPAE-PAD detector, and CarboPac PA1 analytical and guard columns. Results were evaluated using a software package, PeakNet (Dionex Corporation, Sunnyvale, CA). Stem juice samples were run for ten minutes with a flow rate of 1 mL/min of 150 mM sodium hydroxide buffer. A standard curve for sucrose, glucose, fructose, and lactose was developed each time the buffer was replenished. For each sample, sugar values were corrected based on the ratio of lactose detected/lactose expected. Sucrose, glucose and fructose weights were converted to grams per liter of juice (juice sucrose, juice glucose, and juice fructose traits) and these values were summed for total sugar concentration (juice sugars).

Measurement of Grain Quality Traits: One thousand seeds were counted using a seed counter and weighed to obtain thousand seed weight. Thousand seed volume was then measured in a graduated cylinder. Thousand seed density was calculated as thousand seed weight divided by thousand seed volume. The ratio of corneous to floury grain endosperm (corneous endosperm) was the mean of 10 seeds

that were halved and scored visually on a 1 to 10 scale (not measured for CS06). The percent of seeds retaining glumes after threshing was also estimated by visual inspection.

Approximately 60 g of seed were ground in a cyclone mill (UDY Corporation, Fort Collins, Co) using a 1 mm screen with a stainless steel grinding ring and an aluminum impeller. Ground grain samples were then stored at 4^{o}C for approximately 1 month. Before assaying, samples were acclimated for three weeks in a room housing the analytical instrument, a FOSS Model 5000 Feed and Forage Analyzer (FOSS NIRS Systems, Silver Spring, MD), and analyzed with WinISI II software (Infrasoft International, State College, PA). Near infrared spectroscopy (NIRS), a technique for rapid measurement of most organic and some inorganic compounds in tissue is an accurate, reliable and repeatable method for analyzing components of grains, including sorghum (Williams and Sobering, 1993; Hicks et al., 2002; deAlencarFigueiredo et al., 2006; Hooks et al., 2006). A total of 1006 samples from this population were analyzed by NIRS.

NIRS calibration: To obtain accurate data from NIRS, the system must be calibrated based on values obtained from chemical analyses of a subset of samples. For developing calibration equations, 111 of the most informative grain samples were chosen with the WinISI software (76 from the Rio x BTx623 RILs and 35 from diverse sweet and grain sorghums grown at same time in the same locations as the RILS). Grain samples were then analyzed for starch (YSI, 2000), fat (Padmore, 1990), crude protein (Miller et al., 1998) and moisture content by Ward Laboratories (Kearney, NE).

NIRS equations for each grain trait were developed using WinISI. Trait values from a randomly selected group comprising 74 of the 111 samples were used to produce the calibration equations while values from a second group, the remaining

one-third of the samples, were used to evaluate the derived equations. In all, 28 equations (each with different wavelengths and math treatments, and the inclusion of the repeatability file) were tested for each trait. The equations that maximized the prediction of trait values (based on low standard errors of prediction and high R^2) were retained. This process was repeated three times with different subsets of random samples for deriving and validating equations. The best calibration equations were then evaluated using the full subset of 111 samples, and the best equation from each repetition (a total of three equations) was used to predict composition values from the NIRS spectra of all grain samples (n= 1006).

To investigate the effect of calibration sample size on NIRS calibration, we repeated the above procedure for all traits except grain moisture (data were not available) using an expanded dataset that included raw data from additional samples (Hooks et al., 2006). Sample sizes for each trait are reported in Table 2.1. We, therefore, evaluated predicted values from a total of six different calibration equations, the best three from our samples only and the best three equations from the larger sample set that included both our samples and those from Hooks et al. (2006). From these, one equation was selected for each trait based on low standard error of calibration (SEC), low standard error of cross validation (SECV), high R^2, high heritability and repeatable detection of QTL in the RIL population (see below).

The best calibration equation for each trait is shown in Table 2.1. Starch, fat and ADF trait prediction improved with the inclusion of additional sample data but, overall, little difference between most equations was observed. ADF equations were derived solely from the calibration samples of Hooks et al. (2006). However, a calibration developed exclusively from our data using 100 minus the percent starch, fat, and protein after correcting for moisture resulted in prediction of values very similar to Hooks et al. (2006). The phosphorus calibration equation was also based

solely on data from the samples of Hooks et al. (2006). Although we did not measure phosphorus in our calibration samples, the values predicted showed high heritability. NIRS calibration for grain moisture was based only on our sample data because this trait was not assayed by Hooks et al. (2006).

Table 2.1. Calibration treatment and statistics for grain NIRS

Constituent	Math treatment†	N‡	Mean (SD)§	R^2	SEC ¶	SECV #
Starch	2,6,4,1	625	67.71 (5.09)	0.92	0.146	0.154
Fat	3,5,5,1	617	3.9 (0.88)	0.75	0.044	0.045
Crude Protein	2,6,4,1	617	13.56 (1.96)	0.91	0.058	0.06
Phosphorus	1,4,4,1	516	0.43 (0.07)	0.7	0.004	0.004
ADF	1,4,4,1	516	5.97 (1.49)	0.7	0.082	0.086
Moisture	3,5,5,1	107	8.53(0.34)	0.66	0.02	0.025

† Math treatment reflects the derivative number, gap (the nm wavelength over which derivative is calculated), smooth (number of points used to smooth the data), and a secondary smooth respectively. These were applied to reflectance of wavelengths 1108-2492 nm by increments of 8 nm. Standard normal variate transformation (SNV) and detrend scatter correction options were used.
‡ Total number of calibration samples.
§ Mean percentage of sample value in calibration with standard deviation in parentheses.
¶ Standard error calibration (g/kg^{-1}).
# Standard error cross validation (g/kg^{-1}).

Statistical Analyses

Identifying sources of experimental variation and trait heritability: For the statistical analyses, WE05, CS05, CS06, were treated as different environments. Models were evaluated separately for each trait using SAS PROC MIXED (SAS Institute Inc. 2007) software. Trait variance σ^2_{Trait} was estimated as:

$$\sigma^2_{Trait} = \sigma^2_G + \sigma^2_E + \sigma^2_{G*E} + \sum(\sigma^2_X)_{ij} + \sigma^2_{error} \qquad (Eq.1).$$

where σ^2_G is variance due to genotype, σ^2_E is the variance due to environment (*i.e.*, location), σ^2_{G*E} is variance due to interaction of genotype and environment, $\sum(\sigma^2_X)_{ij}$ is the sum of variances due to a number of predicting effects, X, ranging *i* to

j, and σ^2_{error} is the variance due to error. Here, the predicting effects (X) constitute non-genetic sources of experimental error, such as harvest date (see top of Table 2.3 for a complete list of predicting effects). Most of these effects were nested in environment. All effects were treated as random except genotype. Genotype was treated as fixed to allow inferences on specific RILs in later data correction and analyses. Only effects deemed significant (p=0.05) by type III sums of squares and the main effects, in the case of significant interactions, were retained in the reduced model (Table 2.3). Type III sums of squares were also used to estimate variance components from the reduced model. Variance components for genotype (G), environment *(i.e.*, location) (E), genetics by environment interaction (G*E) were used to calculate broad sense heritability

$$H^2 = \sigma^2_G \Big/ \sigma^2_G + \frac{\sigma^2_{G*E}}{E} + \frac{\sigma^2_{error}}{ER} \qquad (Eq.\ 2)$$

where E is the number of environments and R is the number of replicates.

Data correction model, QTL and trait correlation analyses: Input for QTL and trait correlation analyses was obtained from residual values using a mixed model that corrected for the sources of non-genetic experimental error identified above (Table 2.3: i.e. harvest date).

$$Trait^*_{Line,RL} = Trait_{Line,RL} - \overline{x_E} - \sum(\bar{x})_{ij} \qquad (Eq.3)$$

where $Trait^*_{Line,RL}$ is a residual trait value of a RIL replicate, $Trait_{Line,RL}$ is the observed trait value in a RIL replicate, $\overline{x_E}$ is the mean effect in each environment and $\sum(\bar{x})_{ij}$ is the sum of all other the predicting effects (X) identified from the reduced model of *Eq.1* except the G and G*E interactions. The estimates of residuals ($Trait^*_{Line,RL}$) from this model, therefore, contained only the G, G*E and unexplained error variances. These residuals appeared to be near normally distributed in all traits. This approach was similar to statistical correction methodology used in microarray experiments (Wolfinger et al., 2001).

The corrected residuals ($Trait^{*}_{Line,RL}$) were used as input for QTL analysis, after averaging across replicates within a location, and separately for fitting a second model that partitioned variance explained by genetic effects, G and G*E, from the unexplained error variance.

$$Trait^{*}_{Line,RL} = error_{Line,RL} + (\overline{x_G} + \overline{x_{GE}}) \qquad (Eq.4)$$

To identify genetic correlations across environments, the predicted, trait values containing only G and G*E were used as input for SAS PROC CORR (SAS Institute Inc., 2007) software.

Single marker analysis and QTL interval and composite interval mapping were performed with Windows QTL cartographer version 2.5 (WINQTL) (Wang et al., 2007). We used WINQTL settings 'RI1' for the cross type and 2cM for the walk speed. A conservative permutation threshold at the 0.01 significance level was obtained for each trait using 300 permutations; all other settings were left to default (five control markers, 10cM window size forward regression method). QTL Figure 1 was created using R (R Development Core Team, 2005). We should note that the difference in degrees of freedom after fitting the *Eq.3* model was trivial but may result in minor over estimation of QTL and genetic correlation effects.

Genetic Map Construction: Leaf tissue was collected from all 176 RILs and the parents at the CS05 location. DNA was extracted from pooled tissue from four or more plants per line using a standard CTAB extraction protocol (Doyle and Doyle, 1987). A total of 289 markers were scored in this population including 67 SSRs and 222 AFLP markers (AFLP is a registered trademark of Keygene N.V., Wageningen, the Netherlands). Genotyping was performed according to Menz et al. (2002). Missing, non-parental, and heterozyogous alleles were treated as missing data for map construction and QTL mapping. Mapmaker/Exp version 3.0b (Whitehead Institute, Cambridge, MA), Kosambi centiMorgan function, was used to create the genetic map

and JoinMap 4 (Van Ooijen and Voorrips., 2001) was used to place previously unmapped AFLP markers. Linkage groups were assigned to chromosomes using the designations of Kim et al. (2005).

Results

Phenotypic Data: Phenotypes of the RIL parents were as expected based on their selection history. Rio, the sweet sorghum parent, was a tall, late flowering, high sugar content and high juice producer that exhibited more secondary growth (increased tillering) than BTx623 (Table 2.2). BTx623, the grain sorghum parent, was shorter, earlier, and produced larger panicles with more seed than Rio. Overall, trait values for the RILs tended to be intermediate but transgressive segregation was observed for all traits (Table 2.2).

Sugar and grain yields varied between locations due to both experimental and biotic factors. For example, experimental variation in extraction efficiency of different cane presses (on average, 46% of the stem juice was extracted in WE05, 20% in CS05, and 51% in CS06) resulted in variation in first press juice yield across locations (Table 2.2). Grain yield and its components were affected primarily by environment, which was optimal in WE05, and poor in CS05 and CS06 (*i.e.*, damage from rain in CS05, and sorghum midge in both CS05 and CS06). Finally, latitudinal differences (Weslaco latitude = 26.2° N, College Station latitude = 30.6°N) led to longer flowering times in CS05 and CS06 for the slightly photoperiod sensitive Rio parent (Table 2.2).

Variance Components and Error Sources: The relative contributions of genetic, environmental, and experimental causes of variation from calculated variance components are presented in Table 2.3. Genetic effects (G), explained a highly significant, but small, percentage of total phenotypic variation for all traits (median was 14% with values ranging from 0.01% for juice fructose to greater than 50% for

Table 2.2. Trait values for Rio X BTx623 recombinant inbred and parental lines at three locations

	Weslaco, TX 2005				College Station, TX 2005				College Station, TX 2006			
Trait	**Rio**	**BTx623**	**RILs Mean (SD)†**	**RILs Range**	**Rio**	**BTx623**	**RILs Mean (SD)†**	**RILs Range**	**Rio**	**BTx623**	**RILs Mean (SD)†**	**RILs Range**
Brix, °brix	19.4	14.4	17.2 (1.7)	12.3 - 22.5	19.7	9.7	15.5 (1.9)	9.8 - 20	21.7	14.5	16.3 (2.4)	7.9 - 21.5
Juice sugars, g L^{-1}	195	98	148 (25)	54 - 288	178	60	142 (23)	75 - 209	199	87	135 (30)	22 - 202
Juice glucose, g L^{-1}	17	12	11 (3)	0 - 25	15	13	13 (6)	5 - 40	12	19	12 (5)	4 - 43
Juice fructose, g L^{-1}	14	23	14 (10)	4 - 121	13	12	11 (6)	3 - 47	7	5	8 (4)	0 - 34
Juice sucrose, g L^{-1}	164	63	123 (28)	8 - 193	151	36	117 (25)	43 - 182	180	64	114 (31)	9 - 172
Total sugar yield, t ha^{-1} §	2.9	0.8	2.2 (.8)	0.5 - 6.5	7.3	0.6	3.6 (1.4)	0.6 - 9.4	6.6	1.1	3.1 (1.2)	.3 - 7.9
Sugar in dry stem, g kg^{-1}¶	560	310	500 (80)	230 - 790	460	260	530 (100)	310 - 1640	530	330	500 (100)	120 - 730
Juice yield 1st press, t ha^{-1}	7.9	3.5	7.8 (3.5)	0.5 - 24	9.7	1.6	5.7 (2.3)	1 - 17.7	19.8	5.7	13.1 (4.5)	2.6 - 29.8
Total stem water, t ha^{-1}#	14.8	9.5	14.9 (3.9)	4.5 - 34.1	41.5	9.8	25.3 (8.4)	4.6 - 55.8	33.4	12.3	23.1 (7)	6 - 51
Stem fresh yield, t ha^{-1}	23.3	10.3	21.9 (6)	6.2 - 48.1	65.5	13.2	36 (12)	6.5 - 79.3	53	16.8	32.5 (10)	8.5 - 69.3
Stem juiciness, g kg^{-1}††	740	790	770 (20)	680 - 830	720	810	790 (20)	720 - 920	730	790	790 (20)	710 - 850
Panicle fresh yield, t ha^{-1}	4.2	7.3	5.5 (1.1)	2.6 - 9.8	3.6	4.3	3.8 (1.7)	0.4 - 11.4	2.3	6.5	4.3 (1.7)	.9 - 11.3
Grain dry yield, t ha^{-1}‡‡	2.6	4.6	3.5 (0.8)	0 - 6.2	1.1	2	1.4 (1.2)	0 - 6.3	0	3.8	2.3 (1.2)	0 - 6.7
Thousand seed weight, g	18.5	23.3	22 (2.9)	15.3 - 30.8	14.9	20.9	16.9 (3.2)	7.9 - 28.8	14.6	27.4	24.6 (3.3)	11.8 - 32.9
Thousand seed density, g mL^{-1}	0.73	0.75	0.76 (0.03)	0.63 – 1.00	0.64	0.65	0.64 (0.05)	0.45 – 0.75	0.77	0.87	0.85 (0.04)	0.64 – 0.92
Corneous endosperm	7.2	7.3	6.7 (1.8)	1.5 - 9.6	7.3	2.8	4.8 (2.1)	0 - 9.2	na ‡	na ‡	na ‡	na ‡
Grain starch, g kg^{-1}	630	670	650 (10)	600 - 690	600	650	610 (30)	450 - 660	570	680	640 (20)	550 - 680
Grain fat, g kg^{-1}	40	40	40 (0)	30 - 50	30	30	30 (0)	10 - 40	30	30	40 (0)	30 - 50
Grain crude protein, g kg^{-1}	140	110	130 (10)	90 -170	160	130	150 (20)	120 - 190	180	120	140 (10)	110 - 190
Grain moisture NIRS, g kg^{-1}	88	87	87 (1)	70 - 82	88	87	86 (20)	80 - 90	84	84	86 (30)	79 - 94
Grain phosphorus, g kg^{-1}	2.1	1.1	16 (0.4)	0.5 - 2.7	2.9	1.2	2 (0.6)	0.5 - 3.5	3.0	1.0	2 (0.5)	0.6 - 3.4
Grain ADF, g kg^{-1}	130	90	100 (10)	80 - 140	140	90	110 (20)	80 - 170	140	90	110 (10)	90 - 170
Glumes retained after threshing, %	2	10	9 (6)	0 - 30	6	13	14 (11)	0 - 75	1	1	1 (3)	0 - 40
Stand density	7.5	6.5	7.7 (0.6)	4 - 9	6	4.5	5.3 (1.4)	1 - 8	7	7	6.6 (1.3)	2 - 9
Tillering	6.5	2.5	6.1 (1.5)	2 - 9	4.5	1.5	4.5 (1.7)	0 - 8	8	4	5.5 (1.7)	1 - 9
Mean stem thickness	2	6	3.4 (0.9)	1.5 - 7	5.8	5.8	4.3 (1)	2 - 7	4	5	3.5 (1)	1.5 - 7
Plant height, cm	210	130	200 (26)	130 - 274	273	119	227 (29)	119 - 297	227	123	204 (26)	109 - 259
Flowering time, days	116	109	111 (5)	104 - 123	180	161	168 (5)	157 - 185	na ‡	na ‡	na ‡	na ‡

† Standard deviation in parentheses.
‡ Not assayed.
§ Total sugar yield = (juice sugars * (pressed juice + (pressed stem weight wet – pressed stem weight dry))).
¶ Sugar concentration in dry stem = total sugar yield / (dry stem biomass yield – ((pressed stem weight wet – pressed stem weight dry) *juice sugars)).
# Total stem water weight = juice yield 1st press corrected for juices sugars + (pressed stem weight wet – pressed stem weight dry) + (pressed stem dry weight* (1 – NIRS stem dry matter content)).
†† percent water of fresh stem by weight = total stem water weight / (total stem water weight + dry stem wt. + total sugar yield)
‡‡ Grain dry matter yield = panicle fresh yield*(dry grain subsample / fresh panicle subsample)*(NIRS grain dry matter content).

grain ADF and thousand seed weight). Environmental (E) and G*E effects were highly significant for most traits including sugar yield, grain yield and grain composition. Environmental effects on sugar yield were primarily due to locational differences in fresh stem biomass and sugar concentration. Environmental effects on grain yield and composition, on the other hand, were caused by variation in weather and midge damage between locations. Notably, E and G*E effects were markedly reduced for acid detergent fiber (ADF) suggesting that this was the only grain composition trait unaffected by midge or weather. For some traits (*e.g.*, stem juiciness or the percent water in fresh stem biomass, Table 2.3), location was not significant although it accounted for a large amount of variation. This result was due to other significant experimental effects in the model nested within location, which if removed, made the location effect highly significant but did not much alter the percentage of variation explained.

Of the other significant effects, within location harvest date influenced almost half of the traits including both sugar and grain composition (Table 2.3). Strip date, or the number of days that elapsed between harvesting and stripping plants and subsequent processing, nested within harvest date and location had a significant effect on sugar composition because of increased degradation of sucrose to fructose and glucose over time. More importantly, total sugar yield, brix, sugar concentration, and juice volume were not affected by strip date. Within location storage effects (sub-samples were stored in the same box as they were removed from the drying oven) influenced many traits because of differences in sample residual moisture resulting from variation between oven drying cycles. As might be expected, sample processing dates had significant impacts on experimental values of both sugar (HPLC dilution dates) and grain (grain grind and NIRS dates) composition traits. Although grain samples were assayed over a one week period, the NIRS date was highly significant

Table 2. 3. Trait heritability and variance component percentage attributable to genetic (G), environmental (E), genetic/environmental interaction (G*E) and other effects

Traits	Heritability	G	E	G * E	(E)‡ Rep.	(E)‡ Har. Date	(E)‡ Har. Dt.* Strip Dt.	(E)‡ North Border	(E)‡ East Border	(E)‡ South Border	(E)‡ West Border	(E)‡ Storage Box	HPLC Dilution Date	Grain Grind Date	NIRS Date	Residual Error
Brix, °brix	0.65	0.21***	0.17***	0.15***		0.07***										0.40
Juice sugars, g L^{-1}	0.56	0.16***	0.05^{ns}	0.15***		0.03**							0.14***			0.47
Juice glucose, g L^{-1}	0.55	0.09***	0^{ns}	0^{ns}			0.34***						0.11***			0.45
Juice fructose, g L^{-1}	0.15	0.01ns	0.14^{ns}	0^{ns}		0.27*	0.15***						0.02*			0.4
Juice sucrose, g L^{-1}	0.58	0.17***	0^{ns}	0.1***									0.09***			0.53
Total sugar yield, t ha^{-1}	0.62	0.15***	0.38***	0.11***								0.02**	0.02**			0.32
Sugar in dry stem, g kg^{-1}	0.43	0.09***	0^{ns}	0.08**		0.11***							0.11***			0.59
Juice yield 1st press, t ha^{-1}	0.69	0.1***	0.63***	0.02*		0.02***										0.22
Total stem water, t ha^{-1}	0.67	0.16***	0.42***	0.05**												0.37
Stem fresh yield, t ha^{-1}	0.71	0.2***	0.39***	0.07***												0.34
Stem juiciness, g kg^{-1}	0.49	0.11***	0.13^{ns}	0.09***	0.01*	0.08***						0.06*			0.05*	0.47
Panicle fresh yield, t ha^{-1}	0.65	0.18***	0.2^{ns}	0.07***					0.11***							0.44
Grain dry yield, t ha^{-1}	0.63	0.08***	0.5*	0.07***					0.05**	0.05**	0.04**	0.01**		0.03***		0.17
Thousand seed weight, g	0.79	0.16***	0.64***	0.07***												0.13
Thousand seed density, g mL^{-1}	0.53	0.03***	0.83***	0.03***		0.02***						0.01**				0.08
Corneous endosperm	0.72	0.25***	0.19*	0.09**	0.02**							0.07***				0.38
Grain starch, g kg^{-1}	0.70	0.16***	0.47***	0.08***		0.02***									0.03***	0.25
Grain fat, g kg^{-1}	0.80	0.26***	0.37***	0.09***								0.06***			0.03***	0.19
Grain crude protein, g kg^{-1}	0.80	0.14***	0.67***	0.05***		0.01**								0.01***	0.01**	0.12
Grain moisture NIRS, g kg^{-1}	0.70	0.13***	0.11^{ns}	0^{ns}		0.02*						0.37***		0.03**		0.34
Grain phosphorus, g kg^{-1}	0.78	0.27***	0.27***	0.12***								0.04***		0.04***	0.05***	0.22
Grain ADF, g kg^{-1}	0.89	0.54***	0.10***	0.07***		0.01*									0.03***	0.25
Glumes retained after threshing, %	0.43	0.04***	0.60***	0.05***			0.02**	0.03**				0.04***				0.18
Stand density	0.43	0.06***	0.47*	0.09***						0.04**	0.06***					0.27
Tillering	0.54	0.14***	0.17^{ns}	0.11*	0.05***											0.52
Mean stem thickness	0.59	0.17***	0.16*	0.09***	0.02***	0.03**										0.52
Plant height, cm	0.83	0.36***	0.07^{ns}	0.08***		0.04***				0.15***						0.29
Flowering time, days	0.68	0.21***	0.53***	0.12***												0.14

† Variance component of each significant effect divided by total variance components.
Genetic (G), environment (E), genetic by environment (G*E) and main effects were retained regardless of significance.
May not sum to one due to rounding.
‡ Effect is nested in environment (E)
* p=0.05, **p=0.01, ***p=0.001, ^{ns} = not significant

for all composition traits, moisture, starch, fat, protein, ADF and phosphorus. This variation is likely the result of fluctuations in room temperature and humidity affecting the instrument rather than physiological changes in the sample. Over all traits, the within location border effect was fairly minimal. Accounting for significant experimental error effects reduced the error variance for all traits and correcting the raw data for significant error sources increased trait normality, heritability, and QTL detection (see below).

Trait Heritability: In general, the broad sense heritability of measured traits was fairly high (Table 2.3). As has been reported by other authors, height, flowering time, and thousand seed weight had very high heritability values (Brown et al., 2006; Ritter, 2007). For many of the calculated traits (*e.g.*, stem juiciness, sugar concentration in dry stems, and total sugar) multiplicative error led to a lower heritability than for traits that were measured directly. Also, lower heritability was observed for HPLC-measured sugar composition traits in general (juice glucose, fructose, and sucrose) than for brix, although HPLC results were highly repeatable. This apparent inconsistency was due to degradation of the HPLC samples prior to analysis. Brix measurements were collected from juice samples soon after pressing. Samples were then placed on ice and later frozen at -20°C for transport to the HPLC laboratory. Equal heritabilities of sugars measured by HPLC and brix would have been expected had juice samples been frozen in liquid nitrogen immediately after pressing (Ritter 2007).

Trait Correlations: Overall, correlations among sugar composition and yield traits were highly significant and appropriate with respect to sign (Table 2.4). For example, brix was correlated with increased juice sucrose, total juice sugars and sugar yield traits (correlation coefficients were positive). Total sugar yield was only moderately correlated with brix but had very high positive correlations with stem

water weight and stem sugar yield. Therefore, juice yield had a larger influence than sugar concentration in determining total sugar yield. Sugar traits, in general, exhibited low to moderate negative correlations with grain yield and grain starch content.

Grain production measurements are presented both as panicle fresh yield (harvest weight of panicles and associated stems) and dry grain yield (weight of grain after panicles were dried and threshed corrected for moisture content estimated by NIRS). Our results showed that the minor grain moisture variation was correlated with both grain composition and yield (Table 2.4), suggesting that the chemical composition of grain affects moisture retention. Grain composition was measured as percent grain starch, protein, fat, acid detergent fiber (ADF), and phosphorus. These combined traits explained nearly 100% of grain sample weight, with a small remainder consisting of ash and/or error. Starch, the most important feedstock for ethanol, was the major component of grain weight. Starch had a low positive correlation with thousand seed weight and density and a strong positive correlation with grain yield. Because grain composition measurements are reported on a percentage basis, it was not surprising that starch had negative correlations with all other composition traits (Table 2.4).

The ratio of corneous to floury endosperm and percentage of glumes retained after threshing were investigated for relationships with composition. Corneous endosperm was moderately negatively correlated with ADF and positively correlated with the amount of protein. Glumes, being mostly fiber, were expected to be positively correlated with ADF but our results showed a small negative genetic correlation, likely because glumes comprise a very small component of the seed sample weight. In general, improved grain composition for biofuel (high starch, low protein) had small negative correlations with improved stem sugar concentration and yield. Grain yield had low negative correlations with stem sugar yield and moderately negative

Table 2.4. Pearson correlation coefficients of genetic and genetic by environmental effects

TRAITS	Juice sugars	Juice glucose	Juice fructose	Juice sucrose	Total sugar yield	Sugar % of total dry stem	Juice yield 1st press	Stem water wt.	Stem fresh weight	Stem juiciness	Panicle fresh weight	Grain dry yield	Thousand seed weight	Thousand seed density
Brix, °brix	0.91***	-0.12***	-0.14***	0.88***	0.58***	0.59***	0.27***	0.28***	0.34***	-0.57***	-0.37***	-0.38***	-0.2***	-0.05
Juice sugars, g L^{-1}	-	-0.05	-0.1**	0.95***	0.65***	0.65***	0.33***	0.32***	0.38***	-0.57***	-0.36***	-0.37***	-0.2***	-0.05
Juice glucose, g L^{-1}		-	0.53***	-0.25***	0.04	-0.07*	0.04	0.08*	0.07*	0.0-	-0.29***	-0.3***	-0.21***	-0.15***
Juice fructose, g L^{-1}			-	-0.32***	-0.06	-0.03	-0.0-	-0.05	-0.04	0.07*	-0.08*	-0.11***	-0.13***	0
Juice sucrose, g L^{-1}				-	0.61***	0.62***	0.31***	0.3***	0.36***	-0.55***	-0.29***	-0.29***	-0.15***	-0.03
Total sugar yield, t ha^{-1} §					-	0.51***	0.83***	0.9***	0.92***	-0.38***	-0.21***	-0.28***	-0.12***	-0.08*
Sugar in dry stem, g kg^{-1}¶						-	0.37***	0.28***	0.3***	0.15***	-0.16***	-0.24***	-0.1**	-0.04
Juice yield 1st press, t ha^{-1}							-	0.92***	0.93***	-0.13***	-0.05	-0.14***	-0.02	-0.07*
Total stem water, t ha^{-1}#								-	0.98***	-0.2***	-0.08*	-0.16***	-0.06*	-0.07*
Stem fresh yield, t ha^{-1}									-	-0.27***	-0.11***	-0.19***	-0.08**	-0.07*
Stem juiciness, g kg^{-1}††										-	0.33***	0.25***	0.15***	0.02
Panicle fresh yield, t ha^{-1}											-	0.93***	0.34***	0.1**
Grain dry yield, t ha^{-1}‡‡												-	0.39***	0.1**
Thousand seed weight, g													-	0.04
Thousand seed density, g mL^{-1}														-

TRAITS	Corneous endosperm	Grain starch	Grain fat	Grain crude protein	Grain moisture NIRS	Grain phosphorus	Grain ADF	Glumes retained after threshing	Stand density	Tillering	Mean stem thickness	Plant Height	Flowering Time
Brix, °brix	-0.04	-0.25***	-0.04	0.13***	0.0	0.11***	0.29***	-0.21***	0.0-	0.04	-0.03	0.37***	0.39***
Juice sugars, g L^{-1}	-0.05	-0.24***	-0.06*	0.12***	0.0	0.09**	0.26***	-0.17***	0.0-	0.05	-0.04	0.43***	0.4***
Juice glucose, g L^{-1}	-0.03	-0.25***	0.16***	0.25***	-0.18***	0.21***	0.07*	-0.08*	0.09**	0.07*	0.04	-0.02	0.16***
Juice fructose, g L^{-1}	0.14***	-0.0-	0.0-	0.05	-0.12***	-0.02	-0.17***	0.07*	-0.03	0	0.05	-0.06*	0.0-
Juice sucrose, g L^{-1}	-0.07*	-0.19***	-0.08*	0.06*	0.05	0.06	0.26***	-0.15***	0.02	0.05	-0.06*	0.42***	0.36***
Total sugar yield, t ha^{-1} §	0	-0.32***	-0.03	0.2***	-0.05	0.15***	0.24***	-0.02	0.12***	0.16***	0.04	0.75***	0.43***
Sugar in dry stem, g kg^{-1}¶	0.04	-0.12***	-0.03	0.04	-0.05	-0.09**	0.04	-0.07*	-0.12***	-0.09**	0.11***	0.25***	0.34***
Juice yield 1st press, t ha^{-1}	0	-0.24***	0.0-	0.17***	-0.04	0.11***	0.15***	0.05	0.13***	0.18***	0.05	0.77***	0.34***
Total stem water, t ha^{-1}#	-0.0-	-0.29***	0	0.19***	-0.03	0.16***	0.21***	0.06	0.19***	0.23***	0.0-	0.77***	0.33***
Stem fresh yield, t ha^{-1}	0	-0.3***	0	0.21***	-0.03	0.19***	0.22***	0.04	0.19***	0.23***	0.02	0.8***	0.35***
Stem juiciness, g kg^{-1}††	0.11***	0.2***	0.03	-0.11***	-0.05	-0.19***	-0.3***	0.1***	-0.13***	-0.18***	0.19***	-0.39***	-0.14***
Panicle fresh yield, t ha^{-1}	-0.11***	0.57***	-0.05	-0.6***	0.36***	-0.5***	-0.19***	0.14***	0.12***	0.12***	-0.16***	-0.2***	-0.41***
Grain dry yield, t ha^{-1}‡‡	-0.13***	0.67***	0	-0.66***	0.36***	-0.54***	-0.23***	0.1**	0.11***	0.11***	-0.21***	-0.2***	-0.46***
Thousand seed weight, g	-0.05	0.19***	0.09**	-0.12***	-0.02	-0.24***	-0.14***	0.17***	-0.14***	-0.22***	0.16***	-0.1**	-0.28***
Thousand seed density, g mL^{-1}	0.11***	0.04	0.07*	-0.06*	0.07*	0.06	0.02	0.02	0.23***	0.15***	-0.2***	-0.15***	-0.15***
Corneous endosperm	-	0.0-	0.13***	0.34***	-0.41***	0.32***	-0.49***	0.15***	-0.26***	-0.29***	0.25***	0.03	0.14***
Grain starch, g kg^{-1}		-	-0.07*	-0.79***	0.35***	-0.69***	-0.47***	-0.1***	0	0.05	-0.16***	-0.14***	-0.27***
Grain fat, g kg^{-1}			-	0.26***	-0.38***	0.34***	-0.05	0.05	0.11***	0.15***	-0.18***	-0.07*	-0.05
Grain crude protein, g kg^{-1}				-	-0.65***	0.84***	0.08**	0.0-	-0.13***	-0.18***	0.24***	0.07*	0.35***
Grain moisture NIRS, g kg^{-1}					-	-0.35***	0.45***	-0.14***	0.22***	0.24***	-0.21***	0.04	-0.27***
Grain phosphorus, g kg^{-1}						-	0.31***	-0.08**	0.05	0	0.06	0.05	0.25***
Grain ADF, g kg^{-1}							-	-0.2***	0.36***	0.35***	-0.21***	0.13***	0.04
Glumes retained after threshing, %								-	-0.0-	0.0-	-0.09**	0.1**	-0.32***
Stand density									-	0.72***	-0.58***	0.05	-0.09**
Tillering										-	-0.75***	0.13***	-0.15***
Mean stem thickness											-	-0.02	0.33***
Plant height, cm												-	0.29***
Flowering time, days													-

* p=0.05, **p=0.01, ***p=0.001

correlations with stem sugar concentration. Therefore, improved grain and stem sugar yield and composition appear to have only small physiological tradeoffs.

Height and flowering time were highly correlated, as taller plants tended to flower later (Table 2.4). The effects of these two traits, most quantifiably height, were significantly correlated with most of the traits of interest for biofuel production. Taller plants had more stem biomass, more juice, higher stem sugar concentrations and ultimately more sugar yield per hectare. Conversely, later flowering, and therefore often taller plants, had moderate correlations with lower grain yield, lower grain starch, higher protein and phosphorus, though these grain traits must be carefully evaluated for bias due to midge grain damage in some locations. Flowering time, in particular, played an important role in reduced grain yield in midge infested plots (CS05 and CS06) where establishment of sorghum midge on early-flowering lines led to large populations that were then able to overwhelm the lines that flowered later.

Increasing mean stem thickness, which was not correlated with height, had negative effects on grain yield and composition similar to height, but without concurrent improvements in sugar yield and concentration. Mean stem thickness also had very low correlations with stem juiciness and no correlation with first press juice yield. The lack of correlation suggests larger stems do not hold more moisture and are not easier to press. Because stand density was tightly correlated with tillering, we could not adequately separate the two traits. In general, the effect of increasing stand density/tillering was minor but significant for improving sugar yield, grain yield and overall grain composition for biofuel.

Genetic Mapping: The genetic map derived for our RIL population contained a total of 259 SSR and AFLP markers that were assembled into ten linkage groups with good co-linearity with a previously published map (BTx623 x IS3620C, Menz et al., 2002) (Figure 1). The total genetic distance represented on the map was 1836cM.

In total, eight markers were not included in calculating the map distances, one marker was unlinked, and 32 AFLPs could be placed on to their respective chromosomes but the exact positions of these markers could not be determined.

Low residual heterozygosity was observed in the RILs and heterozygous alleles were treated as missing data for both genetic map construction and QTL mapping. There was very little marker segregation distortion (BTx623 parent average = 48.3%, min = 32% max = 64%) the only exception being a single marker tightly linked with a major height QTL, *dw3* (Multani et al., 2003; Patrick J. Brown, personal communication, 2007) on chromosome 7 where only 14% of lines had the dwarf BTx623 allele.

QTL Mapping: Both interval mapping (IM) and composite interval mapping (CIM) consistently identified and mapped QTL at the same positions. Single marker analysis supported many of these QTL and also identified additional QTL which were not significant under the stringent permutation thresholds used for IM and CIM (Table 2.5). Compared to IM, CIM detected more significant QTL effects with smaller genetic distance (1 and 2 LOD intervals). Approximate QTL map positions (CIM, 2 LOD intervals) are presented for all locations in Figure 2 and exact positions are shown in Table 2.5. The majority of favorable alleles for QTL were derived from the expected parent. For example, BTx623 had positive QTL for grain yield while Rio had positive QTL for sugar traits. However, for most traits at least one positive QTL across locations was contributed by the unexpected parent. For example, on chromosome 1 (WE05), Rio had a positive allele for thousand seed weight and BTx623 had a positive allele for sugar yield. For many traits, QTL co-localized either within or between locations, especially for the major height and/or flowering time genes. Such co-localization of QTL across traits suggested a single gene pleiotropic effect, especially

when the traits have obvious biological relationships. Genetic linkage of multiple genes, however, cannot be ruled out.

QTL co-localization clusters were observed on chromosomes 4, 6, 7, and 9. These corresponded to height, flowering time, or stand density/tillering QTL. The height and flowering time QTL at the proximal end of chromosome 9 co-localized with low grain yield and high stem sugar yield (WE05). In CS05, there was a much larger effect flowering time QTL on chromosome 6 that also co-localized with QTLs having highly opposing effects between grain yield and stem sugar.

In all environments, the QTL for brix and stem sugar concentration (total juice sugars) mapped to near identical locations on chromosome 3. As would be expected based on heritability estimates (Table 2.3), the brix QTL had higher significance than total juice sugars and explained 28% of the variance for WE05 and 13% for CS05. A peak for this brix QTL also was found on chromosome 3 in CS06 but its effect was not significant ($p > .01$). Additional brix QTL were present in CS05 and the strongest of these, accounting for 15% of the variance, co-localized with the chromosome 7 height QTL, likely *dw3*. Subsequent analysis for the brix QTL on chromosome 3 and *dw3* showed the highest significance for the interaction term, suggesting an epistatic interaction (data not shown). However, because only 14% percent of the population had the *dw3* allele, we lack power for a rigorous test of epistasis at this locus.

For grain yield in the midge free environment, WE05, we observed multiple positive QTL from both parents. In the stressed environments, CS05 and CS06, only two QTL for grain yield were identified. Here, the major QTL allele on chromosome 6 for increased grain yield under stress originated from BTx623 (the grain sorghum parent) and co-localized with both increased grain starch and decreased flowering time QTL.

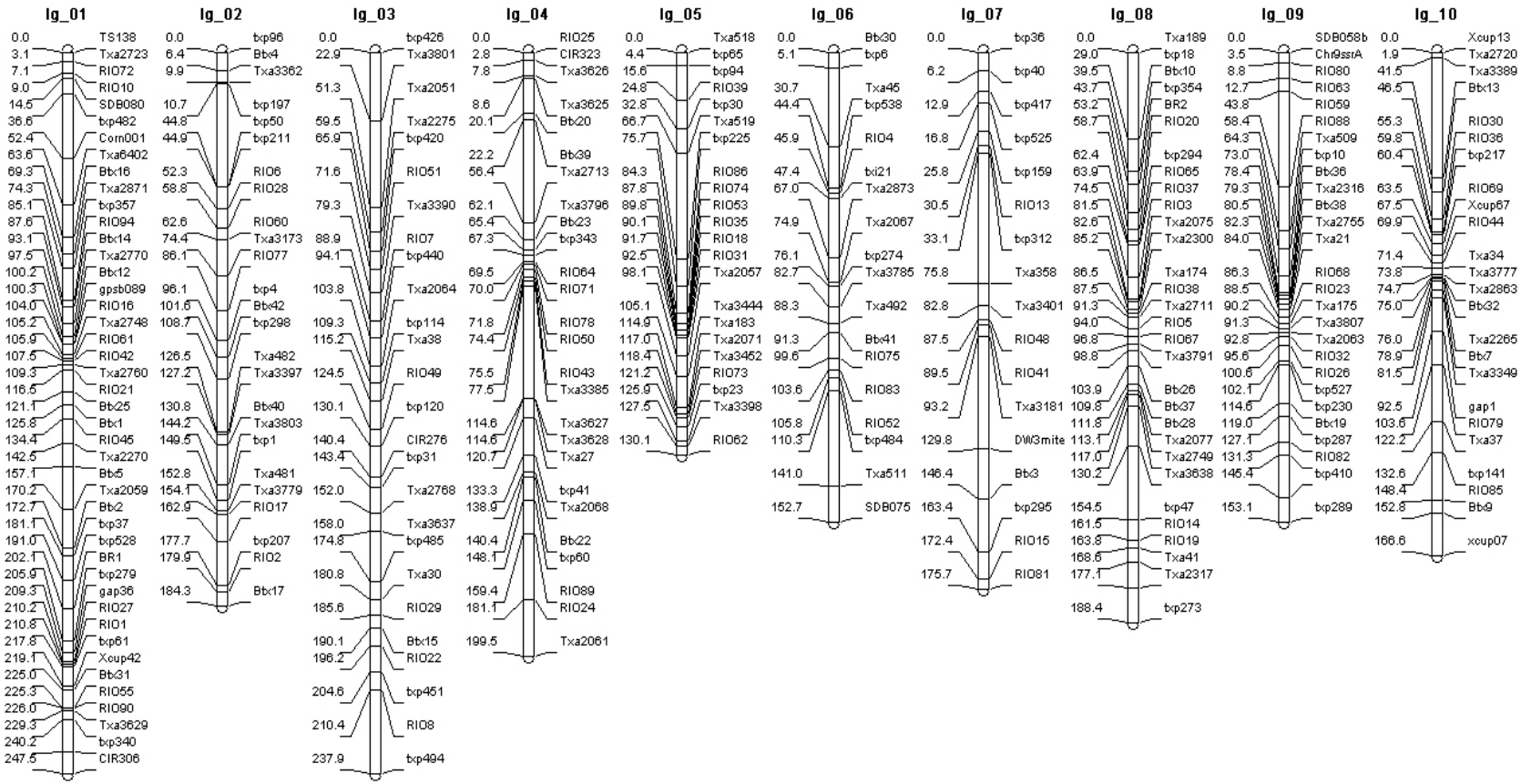


Figure 2.1. Genetic map derived from the Rio X BTx623 RIL population

The ten sorghum chromosomes are named by convention of Kim et al. (2005) and displayed in the orientation of Menz et al. (2002). Marker names beginning with txp, CIR, gap, and xcup denote SSRs. The Txa markers are AFLPs which were previously mapped in RIL population (BTx623 x IS3620c). Marker names beginning with BTx and Rio denote AFLPs that are the unique to our RIL population (Rio x BTx623).

Figure 2.2. Map positions and strength of QTL effects from data collected at three locations.

Colored bars represent QTL detected by CIM. Alleles that increased trait values inherited from BTx623 (grain parent) are shown in blue (WE05), green (CS05) and orange (CS06). Alleles that increased trait values inherited from Rio (sweet parent) are shown in red (WE05), purple (CS05) and pink (CS06). Traits names are shown on the left side of the figure. Marker positions are shown as vertical lines on the bottom above chromosome numbers. The length of each bar represents the 2-LOD QTL interval, and height represents the variance (R^2) explained by the QTL. No QTL were detected for juice fructose and QTL for other traits were not always detected in all locations. Exact QTL map positions are presented in Table 2.5.

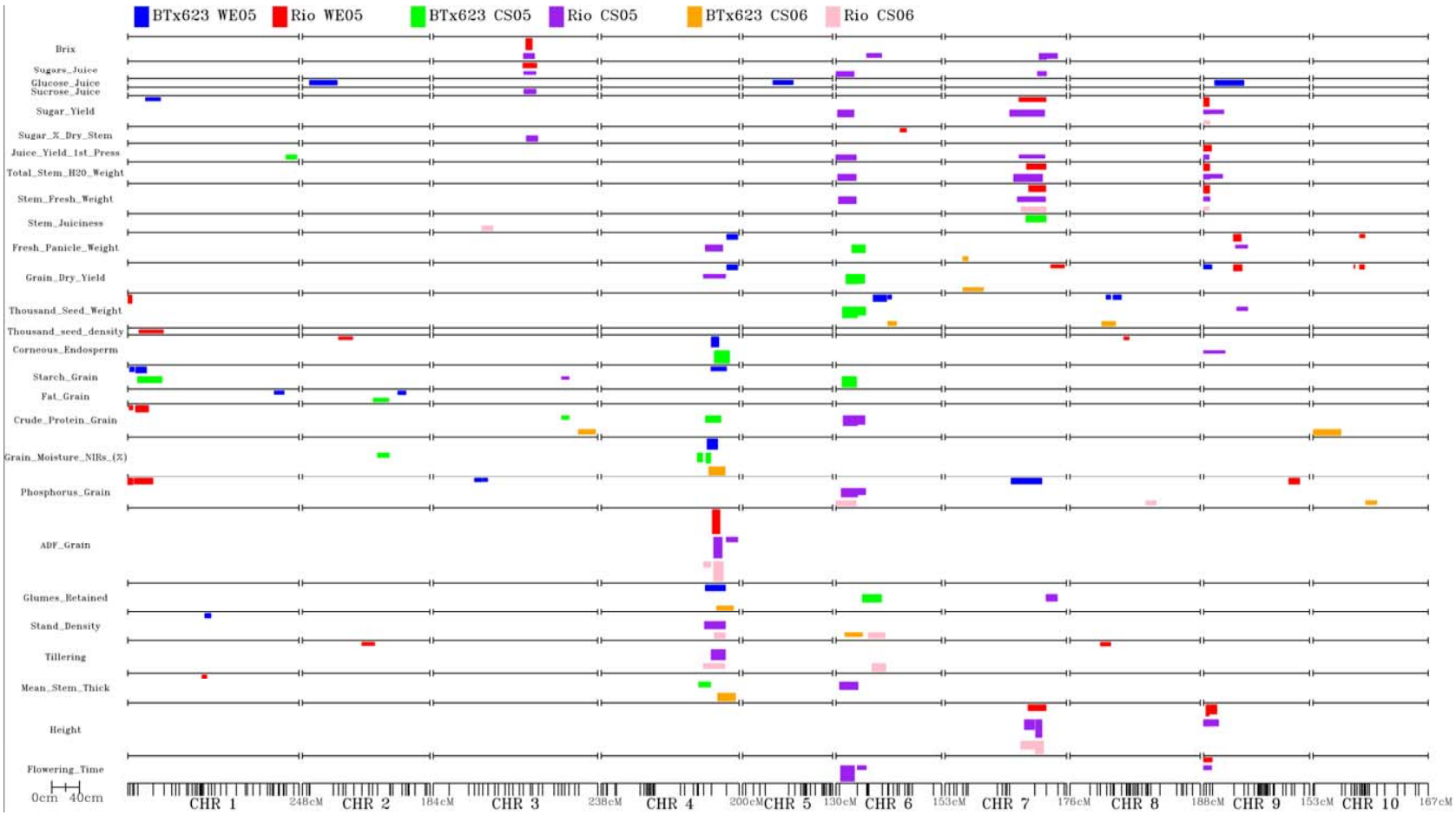

BTx623 WE05
Rio WE05
BTx623 CS05
Rio CS05
BTx623 CS06
Rio CS06
Brix
Sugars_Juice
Glucose_Juice
Sucrose_Juice
Sugar_Yield
Sugar_%_Dry_Stem
Juice_Yield_1st_Press
Total_Stem_H20_Weight
Stem_Fresh_Weight
Stem_Juiciness
Fresh_Panicle_Weight
Grain_Dry_Yield
Thousand_Seed_Weight
Thousand_seed_density
Corneous_Endosperm
Starch_Grain
Fat_Grain
Crude_Protein_Grain
Grain_Moisture_NIRs_(%)
Phosphorus_Grain
ADF_Grain
Glumes_Retained
Stand_Density
Tillering
Mean_Stem_Thick
Height
Flowering_Time
0cm
40cm
CHR 1
248cM
CHR 2
184cM
CHR 3
238cM
CHR 4
200cM
CHR 5
130cM
CHR 6
153cM
CHR 7
176cM
CHR 8
188cM
CHR 9
153cM
CHR 10
167cM

Table 2.5. Positions of QTLs identified by CIM and single marker analyses.
A. QTLs identified by CIM analyses

QTL information				CIM QTL R^2 and peak position with 2LOD interval				Single marker analysis peaks and corresponding LOD score	
Location	Trait	Effect	Chr.	R^2	Center cM	2 LOD left cM	2 LOD right cM	cM (LOD)	cM (LOD)
WE05	Flowering Time	R	9	0.10	2	0	12	3.5 (3.8)	
WE05	Height	R	9	0.28	5.5	3.5	7.5	3.5 (8.6)	
WE05	Height	R	9	0.23	10.8	8.8	18.9		
WE05	Height	R	7	0.14	133.8	120.1	145.9	129.8 (4.1)	
WE05	Mean Stem Thick	R	1	0.10	109.4	107.5	114		
WE05	Tillering	R	8	0.08	53.2	44.2	58.7		
WE05	Tillering	R	2	0.07	96.1	86.1	104.3		
WE05	Stand Density	B	1	0.11	116.5	111.5	120		
WE05	Glumes Retained	B	4	0.14	169.4	150.8	179.4	159.4 (3.4)	
WE05	ADF Grain	R	4	0.60	165.4	160.9	171.4	138.9 (15.8)	159.4 (20)
WE05	Phosphorus Grain	R	1	0.16	2	0	7.1	0 (3.6)	14.5 (3.3)
WE05	Phosphorus Grain	R	9	0.14	131.4	123.3	138.6	131.3 (3.3)	
WE05	Phosphorus Grain	R	1	0.14	13.1	9.1	36.1		
WE05	Phosphorus Grain	B	3	0.08	77.6	71.6	78.5	79.3 (3.4)	
WE05	Phosphorus Grain	B	3	0.08	65.9	59.8	69.9	65.9 (4.1)	
WE05	Phosphorus Grain	B	7	0.14	123.2	95.7	139.7	129.8 (2.5)	
WE05	Grain Moisture NIRS (%)	B	4	0.25	159.4	153.3	168.1	140.4 (9.3)	159.4 (10)
WE05	Crude Protein Grain	R	1	0.17	16.5	11.3	29.8		
WE05	Crude Protein Grain	R	1	0.12	3.1	2	7.1	3.1 (4.6)	14.5 (5.1)
WE05	Fat Grain	B	2	0.09	140.8	138.1	149.5	144.2 (2.6)	152.8 (2.7)
WE05	Fat Grain	B	1	0.08	217.8	212	225.8	181.1 (4.1)	209.3 (3.3)
WE05	Starch Grain	B	4	0.09	173.4	158.8	181.1		
WE05	Starch Grain	B	1	0.12	5.1	2.8	9.1	0 (5.9)	
WE05	Starch Grain	B	1	0.15	14.5	11.6	26.9	14.5 (9)	
WE05	Corneous Endosperm	R	8	0.07	82.7	78	85.2		
WE05	Corneous Endosperm	R	2	0.06	60.8	52.5	72.6		
WE05	Corneous Endosperm	B	4	0.27	161.4	159.4	169.8	140.4 (10.6)	159.4 (12.3)
WE05	Thousand seed density	R	1	0.07	36.7	16.3	51.2	36.6 (7.6)	
WE05	Thousand Seed Weight	R	1	0.20	3.1	0.4	5.7	3.1 (9.8)	
WE05	Thousand Seed Weight	B	8	0.10	55.2	52.6	58.7	53.2 (3.2)	
WE05	Thousand Seed Weight	B	6	0.10	76.1	75	80.1	76.1 (4.1)	
WE05	Thousand Seed Weight	B	8	0.11	67.9	62.4	74	63.9 (3.5)	
WE05	Thousand Seed Weight	B	6	0.15	63.4	53.9	73	67 (4.4)	
WE05	Grain Dry Yield	R	9	0.15	45.8	43.2	55.3	43.8 (2.6)	
WE05	Grain Dry Yield	R	10	0.11	69.9	67.5	74.1		
WE05	Grain Dry Yield	R	10	0.09	59.8	59.3	60.4		
WE05	Grain Dry Yield	R	7	0.07	169.4	153.2	172.4		
WE05	Grain Dry Yield	B	9	0.10	5.5	0	11.6		
WE05	Grain Dry Yield	B	4	0.12	195.1	181.9	197.1		
WE05	Panicle Fresh Weight	R	9	0.16	45.8	43.1	54	43.8 (3.5)	
WE05	Panicle Fresh Weight	R	10	0.08	69.9	67.5	74.8		
WE05	Panicle Fresh Weight	B	4	0.12	195.1	181.7	197.1	181.1 (2.5)	
WE05	Stem Fresh Weight	R	7	0.15	137.8	121	145.4	129.8 (3)	
WE05	Stem Fresh Weight	R	9	0.19	3.5	0	8.3	3.5 (7.1)	
WE05	Total Stem H20 Weight	R	7	0.13	135.8	118	145.8	129.8 (2.8)	
WE05	Total Stem H20 Weight	R	9	0.17	3.5	0	8.4	3.5 (7.2)	
WE05	Juice Yield 1st Press	R	9	0.15	5.5	0	11	3.5 (6.2)	
WE05	Sugar % Dry Biomass	R	6	0.09	99.6	93	101.6	99.6 (2.7)	
WE05	Sugar Yield	R	9	0.21	3.5	0.2	7.6	3.5 (6.7)	
WE05	Sugar Yield	R	7	0.10	131.8	107.1	145.7		
WE05	Sugar Yield	B	1	0.07	36.7	25.7	47.1		
WE05	Glucose Juice	B	5	0.10	66.8	44	73	66.7 (3.3)	
WE05	Glucose Juice	B	2	0.11	28.7	10.4	50.1		
WE05	Glucose Juice	B	9	0.12	36.8	16	58		
WE05	Sugars Juice	R	3	0.11	138.1	129.8	149.5	140.4 (2.8)	
WE05	Brix	R	3	0.28	140.4	134.1	142.8	140.4 (8)	158 (4.5)
CS05	Flowering Time	R	6	0.36	15.1	6.8	26.5	5.1 (6.4)	
CS05	Flowering Time	R	6	0.09	34.8	30.8	43.6		
CS05	Flowering Time	R	9	0.09	2	0	11.2		
CS05	Height	R	7	0.44	133.8	131.3	139.8	129.8 (9.3)	
CS05	Height	R	7	0.24	125.2	115	129.2		
CS05	Height	R	9	0.15	2	0	8.8	3.5 (4.9)	
CS05	Height	R	9	0.15	12.8	9.6	21.2	12.7 (4.8)	
CS05	Mean Stem Thick	R	6	0.18	17.1	5.2	31.5		
CS05	Mean Stem Thick	B	4	0.12	148.1	141.2	158	148.1 (4.2)	
CS05	Tillering	R	4	0.25	169.4	159.3	179.4	159.4 (5.2)	
CS05	Stand Density	R	4	0.17	169.4	149.6	179.4	159.4 (3.9)	

Table 2.5. (Continued)

CS05	Glumes Retained	R	7	0.16	152.4	146.4	162.1	146.4 (3.5)	
CS05	Glumes Retained	B	6	0.18	55.4	38.3	65.7	47.4 (4.6)	76.1 (2.8)
CS05	ADF Grain	R	4	0.53	167.4	162.9	174.7	138.9 (9.5)	159.4 (12.8)
CS05	ADF Grain	R	4	0.11	187.1	181.1	197.1		
CS05	Phosphorus Grain	R	6	0.21	21.1	7.8	30.8	30.7 (6.4)	
CS05	Phosphorus Grain	R	6	0.15	34.8	30.8	42.5	45.9 (4.2)	
CS05	Grain moisture NIRS (%)	B	2	0.10	114.7	108.7	125.4		
CS05	Grain moisture NIRS (%)	B	4	0.22	142.4	139.1	146.4	140.4 (6.9)	
CS05	Grain moisture NIRS (%)	B	4	0.24	154.1	151.5	158.1	159.4 (6.8)	
CS05	Crude Protein Grain	R	6	0.24	23.1	10.6	30.8		
CS05	Crude Protein Grain	R	6	0.21	34.8	30.8	41.5	30.7 (6.2)	
CS05	Crude Protein Grain	B	3	0.08	192.1	185.7	196.2	190.1 (2.6)	
CS05	Crude Protein Grain	B	4	0.16	163.4	151	173	140.4 (4.3)	159.4 (4.9)
CS05	Fat Grain	B	2	0.08	110.7	102.6	124.7	108.7 (2.7)	
CS05	Starch Grain	R	3	0.06	190.1	185.7	196.2		
CS05	Starch Grain	B	1	0.15	26.5	14.2	49.2	36.6 (5.1)	
CS05	Starch Grain	B	6	0.26	21.1	8.9	29.4	5.1 (6.6)	
CS05	Corneous Endosperm	R	9	0.06	7.5	0	30.7		
CS05	Corneous Endosperm	B	4	0.31	173.4	163.7	185.5	140.4 (5.2)	181.1 (8.8)
CS05	Thousand Seed Weight	R	9	0.08	58.4	48.1	63.4	58.4 (3.6)	
CS05	Thousand Seed Weight	B	6	0.20	34.8	30.8	42.8	67 (3.3)	
CS05	Thousand Seed Weight	B	6	0.26	21.1	9.6	30.8	30.7 (4.6)	
CS05	Grain Dry Yield	R	4	0.10	163.4	148	179.6		
CS05	Grain Dry Yield	B	6	0.23	34.8	30.8	41.1		
CS05	Grain Dry Yield	B	6	0.24	27.1	14.6	30.8	30.7 (6.5)	
CS05	Panicle Fresh Weight	R	4	0.15	163.4	150.7	175.5	159.4 (2.5)	
CS05	Panicle Fresh Weight	R	9	0.07	58.4	46.2	63.2		
CS05	Panicle Fresh Weight	B	6	0.19	34.8	23.1	42.1	30.7 (4.4)	
CS05	Stem Juicy N	B	7	0.16	135.8	117.1	145.9	129.8 (3)	
CS05	Stem Fresh Weight	R	7	0.12	133.8	104.7	145.1	129.8 (4.3)	163.4 (4.6)
CS05	Stem Fresh Weight	R	6	0.17	15.1	3.7	28.9		
CS05	Stem Fresh Weight	R	9	0.10	2	0	8.8	3.5 (3.1)	
CS05	Total Stem H20 Weight	R	7	0.18	121.2	99.5	140.7	129.8 (3.4)	
CS05	Total Stem H20 Weight	R	6	0.15	13.1	2.6	28.9		
CS05	Total Stem H20 Weight	R	9	0.11	2	0	8.8	3.5 (3.2)	
CS05	Total Stem H20 Weight	R	9	0.09	10.8	8.8	27.1		
CS05	Juice Yield 1st Press	R	7	0.08	129.8	107.3	144.2	129.8 (2.8)	163.4 (3.4)
CS05	Juice Yield 1st Press	R	6	0.12	13.1	0	29.1		
CS05	Juice Yield 1st Press	R	9	0.11	2	0	7.5	3.5 (3.3)	
CS05	Juice Yield 1st Press	B	1	0.10	235.3	228.9	244.2		
CS05	Sugar in dry stem (%)	R	3	0.15	140.4	134.9	150.9	140.4 (5.1)	
CS05	Sugar Yield	R	7	0.15	117.2	93.8	143.7	129.8 (3.2)	
CS05	Sugar Yield	R	6	0.16	13.1	2.4	25.8	5.1 (3)	
CS05	Sugar Yield	R	9	0.09	5.5	0	8.8	3.5 (3.2)	
CS05	Sugar Yield	R	9	0.08	10.8	8.8	28.9		
CS05	Sucrose Juice	R	3	0.11	140.1	131.3	148.5	140.4 (3.7)	
CS05	Sugars Juice	R	7	0.12	137.8	133.9	146.4		
CS05	Sugars Juice	R	6	0.14	11.1	0.3	25.9	5.1 (3.3)	
CS05	Sugars Juice	R	3	0.09	140.1	131	148.2	140.4 (2.9)	
CS05	Brix	R	7	0.14	137.8	136.3	146.4		
CS05	Brix	R	7	0.12	152.4	146.4	162.4		
CS05	Brix	R	3	0.12	138.1	130.7	146.1	140.4 (3.4)	
CS05	Brix	R	6	0.10	51.4	44.4	65.8	47.4 (4.7)	
CS06	Height	R	7	0.33	133.8	130.5	142.5		
CS06	Height	R	7	0.21	125.2	109.7	129.2	129.8 (7.1)	
CS06	Mean Stem Thick	B	4	0.19	183.1	168.5	194	181.1 (5.2)	
CS06	Tillering	R	6	0.19	61.4	52.2	71.8		
CS06	Tillering	R	4	0.12	156.1	147.7	178.5	138.9 (4.1)	159.4 (4.3)
CS06	Stand Density	R	6	0.13	57.4	47.1	70.7		
CS06	Stand Density	R	4	0.14	165.4	163.7	179.4	140.4 (4.5)	159.4 (4.7)
CS06	Stand Density	B	6	0.09	30.8	12.9	38.2		
CS06	Glumes Retained	B	4	0.10	181.1	167.1	190.9	181.1 (2.8)	
CS06	ADF Grain	R	4	0.48	169.4	162.7	176.2	159.4 (12.7)	
CS06	ADF Grain	R	4	0.13	156.1	148.2	158.1	138.9 (11.5)	
CS06	Phosphorus Grain	R	6	0.12	13.1	0.2	29.1	5.1 (2.9)	
CS06	Phosphorus Grain	R	8	0.10	111.8	109.8	124.5	86.5 (2.9)	96.8 (3.1)
CS06	Phosphorus Grain	B	10	0.09	85.5	76.3	91.9		
CS06	Grain moisture NIRS (%)	B	4	0.20	165.4	155.8	178.9	140.4 (5.5)	159.4 (6)
CS06	Crude Protein Grain	B	3	0.11	230.4	210	234.4		
CS06	Crude Protein Grain	B	10	0.14	13.9	0.6	40	46.5 (3.4)	
CS06	Thousand Seed Weight	B	6	0.10	78.1	75	87.1	76.1 (3.1)	
CS06	Thousand Seed Weight	B	8	0.11	57.2	46.2	65.5	58.7 (3.2)	
CS06	Grain Dry Yield	B	7	0.08	30.6	25.8	55.3	30.5 (2.9)	
CS06	Panicle Fresh Weight	B	7	0.10	29.8	25.8	33.1	30.5 (3.1)	
CS06	Stem Juicy N	R	3	0.11	73.6	70.4	86		
CS06	Stem Fresh Weight	R	7	0.12	135.8	110.2	146.1		
CS06	Stem FreshWeight	R	9	0.08	3.5	0	7.5	3.5 (2.6)	

Table 2.5. (Continued)

CS06	Sugar Yield	R	9	0.09	3.5	0	8.8

B. QTLs identified by single marker analysis only

QTL information				CIM QTL R^2 and peak position with 2LOD interval				Single marker analysis peaks and corresponding LOD score	
Location	Trait	Effect	Chr.	R^2	Center cM	2 LOD left cM	2 LOD right cM	cM (LOD)	cM (LOD)
WE05	ADF Grain	R	1					14.5 (3.1)	
WE05	Crude Protein Grain	B	10					41.5 (3.7)	67.5 (3.8)
WE05	Fat Grain	R	1					3.1 (2.5)	14.5 (2.8)
WE05	Flowering Time	R	2					0 (2.8)	
WE05	Glumes Retained	R	7					146.4 (3.9)	
WE05	Glumes Retained	B	10					63.5 (2.9)	92.5 (2.7)
WE05	Juice Yield 1st Press	R	7					129.8 (2.6)	
WE05	Phosphorus Grain	B	3					22.9 (2.9)	
WE05	Phosphorus Grain	R	4					67.3 (2.6)	
WE05	Starch Grain	B	1					63.6 (3.1)	
WE05	Starch Grain	R	10					67.5 (2.6)	
WE05	Sugar % Dry Stem	R	3					130.1 (2.7)	
WE05	Thousand Seed Weight	R	1					74.3 (2.8)	
WE05	Thousand seed density	B	4					140.4 (2.7)	159.4 (2.7)
WE05	Thousand Seed Weight	B	8					39.5 (2.5)	
CS05	Crude Protein Grain	B	4					0 (2.8)	
CS05	Crude Protein Grain	B	10					41.5 (2.7)	55.3 (2.8)
CS05	Fat Grain	B	1					87.6 (3.7)	104 (4.3)
CS05	Grain moisture NIRS (%)	B	4					65.4 (2.7)	77.5 (3.2)
CS05	Grain moisture NIRS (%)	B	10					55.3 (3.2)	
CS05	Panicle Fresh Weight	R	2					74.4 (2.7)	
CS05	Stem Juicy N	B	5					90.1 (2.9)	
CS05	Sucrose Juice	R	6					30.7 (2.7)	45.9 (3.1)
CS05	Sugar Yield	R	6					45.9 (3.1)	
CS05	Sugars Juice	R	6					45.9 (3.1)	
CS05	Thousand seed density	R	4					138.9 (3.1)	
CS05	Thousand Seed Weight	B	1					172.7 (2.7)	
CS05	Tillering	R	4					138.9 (3.4)	
CS06	ADF Grain	R	7					75.8 (2.8)	
CS06	Crude Protein Grain	R	1					14.5 (2.9)	
CS06	Crude Protein Grain	R	6					5.1 (3)	
CS06	Crude Protein Grain	B	10					67.5 (2.8)	78.9 (2.7)
CS06	Fat Grain	B	1					225.3 (2.5)	
CS06	Grain moisture NIRS (%)	R	6					5.1 (2.5)	
CS06	Phosphorus Grain	R	8					29 (2.6)	58.7 (3.1)
CS06	Starch Grain	B	1					14.5 (3.6)	63.6 (2.7)
CS06	Sugar % Dry Stem	R	3					94.1 (3.8)	109.3 (3.6)
CS06	Thousand Seed Weight	B	8					103.9 (2.6)	
CS06	Thousand Seed Weight	R	1					0 (3)	
CS06	Total Stem H20 Weight	R	7					129.8 (2.7)	

A minor QTL for increased grain yield under stress on chromosome 4 originated from Rio (the sweet sorghum parent) and co-localized with increased stem density / tillering. Another QTL on chromosome 1 for increased grain starch from BTx623 did not co-localize with grain yield but, oddly, did co-localize with a positive sugar yield QTL, also from the grain parent. This QTL on chromosome 1, in addition to the QTL on chromosome 3 for sugar concentration, would be good breeding targets for improved energy content without physiological tradeoffs.

Discussion

Breeding Sorghum for Increased Sugar Yield: For BTx623, the "grain" parent of our RIL population, grain starch was the primary sink for nonstructural carbohydrates. Phenotypes of grain-related traits generally had higher heritability than those related to stem sugar composition or yield. Variation of total energy per hectare in grain was primarily associated with grain yield. Composition, primarily from increased starch content in grain, slightly increases ethanol yield and importantly, increases ethanol fermentation efficiency (Wu et al., 2007). Thus, breeding to increase ethanol yields from grain should focus primarily on increasing grain yield, with increasing the proportion of starch in grain as an important secondary goal.

Total stem sugar yield per hectare is dependent on two traits, sugar concentration in the stem and stem juice yield per hectare. Increasing sugar concentration would be very valuable to increase energy density and reduce processing and transportation costs. However, sugar yield from sugarcane has increased almost exclusively by increasing crop biomass and stem juice yield rather than sugar concentration, perhaps because the concentration has been maximized at 62% of dry weight, or 25% of fresh weight (Jackson, 2005; Moore and Maretzki, 1996). Juice sugar concentrations in many elite sweet sorghum cultivars already reaches 20 - 25 brix, or 66-70% of stem dry weight (SCM, unpublished data). Furthermore, sorghum stem sugar concentration appears to be a primarily additive trait across genetic backgrounds with no noticeable increases in hybrids (Clark, 1981; WLR, unpublished data). Therefore, stem sugar concentration in sweet sorghum is unlikely to be significantly increased by breeding practices. However, stem sugar concentration in grain sorghum may be increased by introducing QTL alleles from sweet sorghum.

In our population, stem juice yield accounted for almost twice as much variation in stem sugar yield than sugar concentration and, therefore, may be a better

initial target for improvement. This supposition is also supported by the fact that sugar yield QTL co-localized with juice yield and stem fresh weight but not with sugar concentration. Juice yield is a function of both stem juiciness (total stem water content/stem fresh weight) and stem fresh weight. Stem juiciness in our population differed little between the parents and had little genetic variance but a major QTL for low stem juiciness, or "dry stalk" (Bennetzen et al., 2001), suggests that sorghum harbors additional genetic variation that could be exploited. The other component of juice yield, stem fresh weight, was highly correlated with height and slightly correlated with stand density/tillering. Stem fresh yield has high genetic variation and heterosis potential in sorghum (WLR, unpublished). Therefore, stem sugar yield per hectare may be best improved in sweet sorghums by increasing stem fresh weight while maintaining maximum sugar concentration and stem juiciness. For grain sorghums, increasing stem fresh weight by increasing height may be undesirable. Because grain sorghums have not been selected for sugar traits, however, stem sugar concentration could be easily improved.

Tradeoffs Between Grain and Stem Sugar Yield: In the environment that experienced no major biotic stress (WE05), only the proximal QTL on chromosome 9 exhibited a genetic basis for tradeoffs between stem sugar yield (~50% of the variation identified by all QTL; see Table 2.5) and grain yield (~16% of the QTL variation). This tradeoff was offset by a closely linked locus contributed by the sweet sorghum parent that increased grain yield (representing ~23% of QTL variation). Under midge and rain stress (CS05), only the QTL on chromosome 6 exhibited a genetic tradeoff between stem sugar (representing 40% of QTL variation) and grain yield QTL (100% of QTL variation). These results suggest that stress created negative relationships between grain and stem sugar yields under standard, non-limited agronomic practices. Consequently, breeders should be able to improve grain starch

and stem sugar simultaneously in both grain and sweet sorghum types, but tradeoffs will increase with stress. The feasibility of concurrent improvement of grain and sugar yields is supported by other studies. Lingle (1987) found sink (in this case, energy stored in grain starch) but not source (photosynthetic) limitations in crops grown in non-limited environments. This study concluded that the developing grain is not a significant sink for whole plant carbohydrates. Other studies (Wu and Birch, 2007) have found that transforming sugarcane to produce a second sugar (an additional sink), isomaltulose, in addition to sucrose, nearly doubled the total sugar concentrations in harvested juice. Results from both studies imply that sinks operate independently and it is possible to increase a plant's ability to store photosynthates. In the past cheap energy and a lack of infrastructure allowed growers to focus on harvesting a single product (*i.e*., grain, stem sugar, or forage). Plants may be most efficient at producing energy, however, if there are a number of different sinks for storing nonstructural carbohydrates throughout the growing season, especially under ideal agronomic conditions.

Co-localization with QTL from Other Studies: In QTL analyses, only loci with alleles that differ between the parents of the study population can be identified and mapped. Here, we evaluated a population derived from a low sugar accumulating grain sorghum (~12.6 brix and low juice volume) crossed to a very high sugar producing sweet sorghum (~20 brix and high juice volume). Therefore, we were likely to find major QTL for high sugar accumulation. Three other studies have identified QTL for sugar concentration in sorghum. Natoli et al. (2002) used an F_2 population derived from two sweet sorghum parents (brix of 15.4 and 15.9) to map QTL that differed between high sugar types. Ritter (2007) used a RIL population derived from a very low sugar grain sorghum parent (~6.6 brix) crossed to a low sugar (~12.1 brix) photoperiod insensitive dwarf grain variety of Rio (not selected for high sugar) to

identify QTL for stem sugar in grain sorghum. Finally, Bian et al. (2006) identified brix QTL in an F_3 population derived from a sweet sorghum and a grain inbred line but did not report sugar values.

Given differences in populations, locations, and measurements we did not expect high QTL co-localization between these studies, but found a number of strong similarities. As in our study, Natoli et al. (2002) identified and mapped a major QTL for brix to the middle of chromosome 3. In Ritter (2007) it is likely that an unassigned linkage group containing the largest height, sugar, and flowering time QTL corresponds to QTL near the telomere of the long arm of chromosome 9 in our study. No brix QTL were shared with Bian et al. (2006). Both Natoli et al. (2002) and Ritter (2007) detected height and sugar yield QTL on chromosome 5 that were not identified in this study, possibly because BTx623 carried the same alleles as Rio in this region.

To date there have been few published molecular genetic studies of grain yield or composition in sorghum. Rami et al. (1998) evaluated two sorghum mapping populations for grain yield and quality traits, and found little co-localization between QTL except for a few major height genes which affected many traits, findings similar to our results. There were also several QTL identified by Rami et al. (1998) that co-localized with our study, specifically a grain protein QTL on chromosome 1, a fat and thousand seed weight QTL on chromosome 1, and a corneous grain starch QTL on chromosome 2. There may also be a common grain yield QTL on chromosome 10 with Ritter (2007), but in our study the allele from the Rio parent had an opposite effect. The QTL on chromosome 10 reported by Ritter (2007) also co-localized with increased dry matter and stem sugar (though it slightly decreased sugar concentration). Thus, Ritter (2007) found no QTL tradeoffs between grain yield and stem sugar yield indicating that differences in height caused tradeoffs under non-stress conditions.

The height QTL identified on chromosome 7 of this study has been detected in other studies (Rami et al., 1998; Brown et al., 2006) and is likely *dw3* (Multani et al., 2003; Pereira and Lee, 1995). The height and flowering time QTL identified on chromosome 9 was also detected by Pereira and Lee (1995) and by Lin et al. (1995). The QTL effects identified on chromosome 1 are consistent with flowering time QTL from Crasta et al. (1999), Ritter et al. (2007) and Natoli et al. (2002). The major flowering time QTL on chromosome 6 in CS05 was reported by Lin et al.(1995) as *ma1*, and was also detected in Rami et al. (1998), and Brown et al. (2006). *Ma1* is known to be regulated by photoperiod and the fact that we do not detect this large QTL in WE05 can be explained by latitudinal differences between locations (Quinby and Karper, 1945; Lin et al., 1995).

Limits of QTL Studies: Population size, trait heritability, and recombination all affect the ability to accurately detect QTL (Beavis, 1994; Kearsey and Farquhar, 1998; Collard et al., 2005). Given the large amount of sample processing and the cost of phenotyping, the population size evaluated for this experiment was as large as feasible. There are two other major limitations to identifying QTL in bi-parental RIL populations. First, only two alleles, at most, can be evaluated. We assume the parental lines adequately represent respective grain and stem sugar sorghum types, and judging from the results of Natoli et al. (2002) and Ritter (2007), this assumption seems reasonable. However, more sweet and grain sorghums need to be investigated. Second, elite varieties in the developed world are usually grown in hybrid combinations which rely on significant dominance effects. Dominance effects can not be evaluated in a study of homozygous RILs, but additive effects can be identified and are of more universal value for crop improvement. Although stem sugar concentration (as indicated by brix and total juice sugar) appears to be additive, the dominance heterosis in hybrids for biomass, juice volume, and grain yield can be up to 150%, raising total

sugar yields and grain yields significantly (WLR, unpublished). It is possible that heterosis could affect the relationships between nonstructural carbohydrates identified in this study.

Conclusions

To our knowledge, this study is the first to investigate the genetics of tradeoffs between grain starch and stem sugar production in sorghum. By measuring many traits concurrently, we indentified QTL clusters where co-localization of sugar and grain QTL are likely due to changes in plant architecture (height, flowering time, stand density/tillering). Results suggested that increases in plant sinks may increase total energy production potential, especially in non-stressed growing environments. This work represents only a first step in understanding the genetics of carbohydrate accumulation and partitioning in sorghum. Future studies should: (1) determine whether these findings are common in sorghum by surveying a larger number of parental alleles from other grain and sweet sorghum cultivars; (2) investigate trait heterosis by evaluating populations as hybrids useful to growers; and (3) use phenotypic assays that are less time consuming and costly by evaluating whole plants with NIRS calibrated for stem sugar and grain starch. These studies will ultimately allow genetic improvement of sorghum to maximize energy capture and storage for sustainable biofuel production.

REFERENCES

Beavis, W.D., O.S. Smith, D. Grant, and F. Fincher. 1994. Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Sci. 34:882–896.

Bennetzen, J.L., V. Subramanian, J. Xu, S.S. Salimath, S. Subramanian, D. Bhattramakki, and G.E. Hart. 2001. A framework genetic map of sorghum containing RFLP, SSR and morphological markers. P. 347-356 In: R.L. Phillips and I.K. Vasil (ed.) DNA-based markers in plants. 2nd ed. Kluwer Academic Publishers, Boston.

Bian, Y.-l., Y. Seiji, I. Maiko, and H.-W. Cai. 2006. QTLs for sugar content of stalk in sweet sorghum (Sorghum bicolor L. Moench). Agricultural Sciences in China 5:736-744.

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

Brown, P.J., P. Klein, E. Bortiri, C. Acharya, W. Rooney, and S. Kresovich. 2006. Inheritance of inflorescence architecture in sorghum. Theor. Appl. Genet. 113:931–942.

Clark, J.W. 1981. The inheritance of fermentable carbohydrates in stems of Sorghum bicolor (L.) Moench. PhD Thesis, College Station, Texas A&M University.

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

Crasta, O.R., W.W. Xu, D.T. Rosenow, J. Mullet, and H.T. Nguyen. 1999. Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol. Gen. Genet. 262:579–588.

deAlencarFigueiredo, L.F., F. Davrieux, G. Fliedel, J.F. Rami, J. Chantereau, M. Deu, B. Courtois, and C. Mestres. 2006. Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. J. Agric. Food Chem. 54:8501-8509.

Dionex Corporation. 2006. The determination of sugars in molasses by high-performance anion exchange with pulsed amperometric detection. Application Note 92. [Online]. Available at http://www1.dionex.com/en-us/webdocs/4666_an92.pdf [verified 8 Aug. 2007]. Dionex Corporation, Sunnyvale, CA

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15.

Farrell, A.E., R.J. Plevin, B.T. Turner, A.D. Jones, M. O'Hare, and D.M. Kammen. 2006. Ethanol can contribute to energy and environmental goals. Science 27:506-508.

Frederiksen, R.A., and F. Miller. 1972. Proposal for release and increase ATx622, BTx622, ATx623, BTx623, ATx624, BTx624. TAES Form 96–72: Seed Release Committee of the Texas Agricultural Experiment Station College Station.

Hallam, A., I.C. Anderson, and D.R. Buxton. 2001. Comparative economic analysis of perennial, annual, and intercrops for biomass production. Biomass Bioenerg. 21:407-424.

Hicks, C., M.R. Tuinstra, J.F. Pedersen, F.E. Dowell, and K.D. Kofoid. 2002. Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica 127:31-40.

Hooks, T., J.F. Pedersen, D.B. Marx, and K.P.Vogel. 2006. Variation in the US photoperiod insensitive sorghum collection for chemical and nutritional traits. Crop Sci. 46:751-757.

Jackson, D.R., M.F. Arthur, M. Davis, S. Kresovich, W.T. Lawhon, E.S. Lipinsky, M. Price, and A. Rudolph. 1980. Research report on development of sweet sorghum as an energy crop volume 1: Agricultural task. USDOE, Battele, Columbus. U.S. Department of Commerce.

Jackson, P.A. 2005. Breeding for improved sugar content in sugarcane. Field Crops Res. 92:277–290.

Kearsey, M.J., and A.G.L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity 80:137-142.

Kim, J.S., P.E. Klein, R.R. Klein, H.J. Price, J.E. Mullet, and D.M. Stelly. 2005. Chromosome identification and nomenclature of Sorghum bicolor. Genetics 169:1169-1173.

Kresovich, S., and P.R. Henderlong. 1984. Agronomic potential of sorghum as a raw material for ethanol production in central Ohio. Energ. Agr. 3:145-153.

Lichtenthaler, F.W., and S. Peters. 2004. Carbohydrates as green raw materials for the chemical industry. C. R. Chim. 7:65–90.

Lin, Y.-R., K.F. Schertz, and A.H. Paterson. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411.

Lingle, S.E. 1987. Sucrose metabolism in the primary culm of sweet sorghum during development. Crop Sci. 27:1214-1219.

Mazur, B., E. Krebbers, and S. Tingey. 1999. Gene discovery and product development for grain quality traits. Science 285:372-375.

Menz, M.A., R.R. Klein, J.E. Mullet, J.A. Obert, N.C. Unruh, and P.E. Klein. 2002. A high-density genetic map of Sorghum bicolor (L.) Moench based on 2926 AFLP® , RFLP, and SSR markers. Plant Mol. Biol. 48:483–499.

Miller, R.O., J. Kotuby-Amacher, and J. Rodriguez. 1998. Total nitrogen in botanical materials- automated combustion method. p. 116–117. In Western states laboratory proficiency testing program. Soil and plant analytical methods. Version 4.10. Colorado State University, Ft. Collins. CO.

Ming, R., S.C. Liu, P.H. Moore, J.E. Irvine, and A.H. Paterson. 2001. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res. 11:2075-2084.

Mohanty, A.K., M. Misra, and L.T. Drzal. 2002. Sustainable bio-composites from renewable resources: opportunities and challenges in the green materials world. J. Polym. Environ. 10:19-26.

Moore, P.H., and A.Maretzki. 1996. Sugarcane. p. 643–669. In E. Zamski and A.A. Schaffer (ed.) Photoassimilate Distribution in Plants and Crops: Source-Sink Relationships. Marcel Dekker, New York-Basel-Hong Kong.

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

National Sorghum Producers. 2007. [Online]. Available at http://www.sorghumgrowers.com (verified 12 Nov. 2007).

Natoli, A., C. Gorni, F. Chegdani, P. Ajmone Marsan, C. Colombi, C. Lorenzoni, and A. Marocco. 2002. Identification of QTLs associated with sweet sorghum quality. Maydica 47:311-322.

Padmore, J.M. 1990. Fat (crude) or ether extract in animal feed, method 920.39. In K. Helrich (ed.) Official Methods of Analysis of the Association of Official Analytical Chemists, 15th ed. AOAC, Inc., Arlington, VA.

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

Quinby, J.R., and R.E. KARPER. 1945. The inheritance of three genes that influence time of floral initiation and maturity date in milo. J. Am. Soc. Agron. 37: 916-936.

R Development Core Team. 2005. R: A language and environment for statistical computing. Vienna, Austria. [Online]. Available at http://www.R-project.org (verified 12 Nov. 2007).

Rami, J.F., P. Dufour, G. Trouche, G. Fliedel, C. Mestres, F. Davrieux, P. Blanchard, and P. Hamon. 1998. Quantitative trait loci for pain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor l. Moench). Theor. Appl. Genet. 97:605-616.

Ritter, K.B. 2007. An investigation into the genetics and physiology of sugar accumulation in sweet sorghum as a potential model for sugarcane. Ph.D. diss. University of Queensland, School of Land, Crop and Food Sciences. St Lucia, Australia.

SAS Institute Inc.. 2007. SAS 9.1.3 Help and Documentation. 2000-2007. SAS Institute Inc. Cary, NC.

Schluhuber, A.M. (1945) Inheritance of stem charecters: In certain sorghum varietes and their hybrids. J. Hered. 1945 36: 219-222

Somerville, C. 2007. Biofuels. Curr. Biol. 17: R115-R119.

Tarpley, L., D.M.Vietor, F.R. Miller. (1994) Internodal compartmentation of stem-infused [14C]sucrose in sweet and grain sorghum. Crop Science. 1994;34:1116–1120.

Tarpley, L., D.M. Vietor. (2007) Compartmentation of sucrose during radial transfer in mature sorghum culm. BMC Plant Biol. 7:33.

U.S. DOE. 2006. Breaking the biological barriers to cellulosic ethanol: A joint research agenda. DOE/SC 0095. USDOE Office of Science and Office of Energy Efficiency and Renewable Energy, Washington, DC. [Online]. Available at http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml (verified 14 Nov. 2007).

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.

Vietor, D.M., and F.R. Miller. 1990. Assimilation, partitioning, and nonstructural carbohydrate in sweet compared with grain sorghum. Crop Sci. 30:1109-1115.

Wang, S., C.J. Basten, and Z.B. Zeng. 2007. Windows QTL Cartographer 2.5. Available at Department of Statistics, North Carolina State University, Raleigh, NC. [Online]. Available at http://statgen.ncsu.edu/qtlcart/WQTLCart.htm (verified 14 Nov. 2007).

Williams, P.C., and D.C. Sobering. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. J. Near Infrared Spectrosc. 1:25-33.

Wolfinger, R.D., G. Gibson, E. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R.S. Paules. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Compu. Biol. 8: 625-637.

Wu, L., and R.G. Birch. 2007. Doubled sugar content in sugarcane plants modified to produce a sucrose isomer. Plant Biotechnol. J. 5:109-117.

Wu, X., R. Zhao, S.R. Bean, P.A. Seib, J.S. McLaren, R.L. Madl, M. Tuinstra, M.C. Lenz, and D. Wang. 2007. Factors impacting ethanol production from grain sorghum in the dry-grind process. Cereal Chem. 84:130-136.

YSI incorporated. 2000. Determination of % cook in extruded cereal products using chemical solubilization. Application # 322. Yellow Springs, OH.

CHAPTER 3

GENETIC IMPROVEMENT OF SORGHUM AS A BIOFUEL FEEDSTOCK: II. QTL FOR STEM AND LEAF STRUCTURAL CARBOHYDRATES[2]

Abstract

Digestion and fermentation of lignocellulosic biomass (*i.e.*, structural carbohydrates) are predicted to deliver higher yields of energy per hectare than sugar and starch (non-structural carbohydrates), yet little research on genetic variation in crop feedstock biomass traits has been conducted. We investigated the genetic basis of leaf and stem biomass yield and composition in a population derived from a high biomass sweet sorghum, 'Rio', and a grain sorghum inbred line, 'BTx623', and compared these results with those from analyses of grain and stem sugar traits that we reported previously. Thirty two traits were evaluated and a total of 152 QTL were identified across three locations. Many QTL for structural and non-structural carbohydrate yields co-localized with loci for height, flowering time and stand density/tillering. QTL for composition had little co-localization across tissues and environments. Separate genetic control for leaf and stem structural carbohydrate composition was identified as well as separate genetic control of protein accumulation in leaf, stem, and grain. To maximize energy yields from grain and dedicated biomass sorghums, results suggest yield traits should be targeted for improvement before composition traits.

[2] Murray, S.C., W.L. Rooney, , S.E. Mitchell, A. Sharma, P.E. Klein, J.E. Mullet and S. Kresovich (accepted) Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Science.

Introduction

Production of biofuels from plant structural carbohydrates (the cellulose, hemi-cellulose and lignin-containing portion of stem, leaf and root tissue) is predicted to yield five times more net energy per unit land area than using grain starch and sugar while producing only a quarter of the greenhouse gases (U.S. DOE, 2006; Farrell et al., 2006; Somerville, 2007). These predictions focus on the C4 grasses such as maize, sorghum, sugarcane, miscanthus and switchgrass that can efficiently produce high yields of structural carbohydrates in biomass. Structural carbohydrates, specifically cellulose, may derive either from crop residue, the byproduct of crops bred and harvested primarily for grain or stem sugar, or from dedicated biomass crops bred primarily for production of structural carbohydrates. Competition with food streams can, therefore, be minimized if crop residue is used as an ethanol feedstock or if dedicated biomass crops are grown on marginal land. These approaches may still, however, conflict with goals for wisely managed soil and wildlife conservation and the need for cautious development of sustainable technologies for dedicated biomass production remains (Lal, 2005; Bies, 2006).

Currently, progress in improving crops for structural carbohydrate production lags behind advancement in biofuel production technologies. Thus, the first ligno-cellulose processing plants will likely use both crop residues and feedstocks that have not been improved for biomass yield or composition traits. To attain maximal efficiency, however, the system will eventually require feedstocks that have been selected for various compositional characteristics (*i.e.*, high cellulose and low lignin content). Competing feedstock conversion technologies such as acid hydrolysis, enzymatic hydrolosis, thermochemical methods (syngas) and direct combustion will require both different feedstock traits and economic considerations for optimal system performance (Hamelinck et al., 2005). For example, the degree of feedstock cellulose

polymerization, crystallinity, surface area, lignin content, and protein content will affect end sugar yield, although these effects have yet to be quantified (Hamelinck et al., 2005; U.S. DOE, 2006; Somerville, 2007). Genetic variation for improving these traits likely exists but the tremendous expense of directly measuring cellulose quality and quantity within a crop species makes assessment economically impractical.

As a starting point for advancing biomass quality, we should first consider forage and silage crops where structural carbohydrates have already been the focus of improvement, and economical analysis methods are currently available. Goals shared between forage and biomass feedstock improvement include yield, resistance to lodging, re-growth potential/perennial habit, high cellulose content, and nutrient use efficiency. An important difference, however, is that in forage crops, high protein and mineral content is necessary for animal feed but undesirable in a biofuel feedstock (Casler and Vogel, 1999; Jenkins et al., 1998; Wu et al., 2007). Proteins reduce cellulose digestibility and fermentation efficiency during ethanol production and create air pollution in systems that use direct combustion. Minerals, on the other hand, can foul processing equipment (Jenkins et al., 1998).

Because of its drought tolerance, nutrient use efficiency and ability to adapt to a variety of environments, the C4 grass, sorghum [*Sorghum bicolor* (L.) Moench], is a promising crop for biofuel production (Rooney et al. 2007). In the only published common garden studies for dry biomass yield, sweet and forage sorghums out-yielded maize, switchgrass, reed canary grass, big bluestem, and alfalfa for dry biomass, especially under low input regimes (Anderson et al., 1995; Hallam et al., 2001). Grain, sweet, and forage type sorghums are all compatible with current agricultural and ethanol production systems, allowing sorghum to function as an improved feedstock crop residue or as a dedicated biomass crop.

We have previously identified quantitative trait loci (QTL) for grain and stem sugar composition and yield and results indicated that overall energy yields could be increased by concurrent improvement for both sorghum grain and sugar traits (Murray et.al., 2008). In this study, we identified ligno-cellulosic leaf and stem structural biomass yield, composition and re-growth QTL that could be used to improve sorghum as a biomass feedstock. We were also interested in exploring the relationships between 1) whole plant non-structural and structural carbohydrate yields, 2) leaf and stem structural carbohydrate composition, and 3) protein levels in stems, leaves, and grain. Finally, we address the question of whether genetic improvement of sorghum should be focused on residue or on dedicated structural biomass production.

Materials and Methods

Plant Material: A recombinant inbred line (RIL) population consisting of 176 $F_{4:5}$ lines was developed from a cross between 'BTx623' (a grain sorghum inbred line, Frederiksen and Miller, 1972) and 'Rio' (a high biomass inbred sweet sorghum cultivar, Broadhead, 1972). Two replicates of RILs were planted during the normal summer growing season in 2005 in Weslaco, TX (WE05), and College Station, TX (CS05). In 2006, two replicates of 167 $F_{5:6}$ RILs were planted in College Station (CS06) from seed harvested in CS05. Two replicates of 3.05 meter rows were planted in a randomized complete block design in each location. Environmental conditions of photoperiod, wind, and moisture, between Weslaco and College Station locations were very different (Murray et al., 2008). Through the growing season, rainfall was two centimeters in WE05, 43 cm in CS05 primarily through flowering, and 36cm in CS06.

Field Measurements: Plant height was measured either in the field (WE05) or, due to lodging, at the time of stripping (*i.e.*, division of plants into panicles, stem and leaf components) (CS05, CS06). Stand density and tillering were visually assessed

using a scale that ranged from 0 (no plants or tillers) to 10 (very dense main stalks, very dense tillering) in the harvested area of each row. Average stem thickness was estimated by visual assessment of the base node using a scale from 0 (thin) to 10 (very thick). Flowering time was measured as time from planting to 50% anthesis (WE05, CS05).

Biomass Measurements: Plant harvest was staggered across 16 days (WE05), 14 days (CS05), and 11 days (CS06) due to the volume of work and the logistics of labor and equipment. Harvest date, therefore, was recorded and used as a cofactor in later statistical analyses. For each row, plants were harvested from a randomly selected area (one meter in length) by cutting within one inch of the plant base. Plants were bundled in clear plastic sheeting and taken to a central processing facility within two hours of harvest. The bundled plants were stripped into panicles, stems and leaves and weighed (panicle fresh yield, stem fresh yield, leaf fresh yield, respectively). Although most plants were stripped on the same day, some were not processed for up to 3 days after harvest. Strip date, therefore, was also noted and used as a cofactor in subsequent analyses. Wet stem tissue was pressed to remove the juice on press date, either the same or next day as stripping. Brix, a measure of soluble solids that in sweet sorghums is comprised mostly of sucrose, was measured with two different handheld refractometers (Atago U.S.A. Inc., Bellevue, WA) and averaged.

For each plot, random subsamples of panicles, leaves, and pressed stems were weighed. These were then dried for a few days in a greenhouse (WE05) or a grain drier set at 38°C (CS05, CS06). Dry stem and leaf subsamples were re-weighed and stem dry yield and leaf dry yield were calculated by dividing these values by wet sub-sample weight and multiplying by full wet sample weight. Stem dry harvest index was then calculated by dividing the dry stem yield by the sum of dry stem sugar, stem,

leaf, and panicle yields. Leaf, panicle and grain harvest indices were calculated in the same manner as above, substituting the appropriate tissues.

Regrowth Measurements: After harvest in CS06, full plots were subjected to uniform mowing and allowed to re-grow to maturity. A second harvest of both replications was conducted in a single day. One meter of plants from each row was harvested, weighed (total biomass including grain), and pressed in a sorghum press to extract juice. Measurements on juice volume, brix, height and maturity stage at harvest were recorded, in addition to wet biomass. This material was not analyzed further.

Stem and Leaf NIRS: For each sample, at least 40 g of dry stem tissue was cut and ground in a no.8 Christy mill with 2mm screen. Approximately 15 g of dry leaf tissue, as well as the previously ground stem tissue, were processed separately in a UDY cyclone mill (UDY Corporation, Fort Collins, Co) using a 1mm screen with a stainless steel grinding ring and an aluminum impeller. The ground tissue was stored in a redline zipper storage bag for 1-3 months, then moved to the near infrared spectroscopy (NIRS) laboratory and acclimated for 3 weeks prior to analysis. Scanning was done on a FOSS Model 5000 Feed and Forage Analyzer with ¼ cup cells (NIRS Systems, Silver Spring, MD) and analyzed with WinISI II software (Infrasoft International, State College, PA). A total of 1051 leaf samples and 1050 stem samples from this population were analyzed by NIRS.

To obtain accurate data from NIRS, the system must be calibrated based on values obtained from chemical analyses in a subset of samples. Therefore, leaf and stem samples from each location were selected for chemical analysis to maximize the information content using the WinISI software. In all, 107 leaf samples (72 from the RIL population and 35 from diverse sweet and grain sorghum accessions grown in the same environment as the RILs) and 168 stem samples (82 RILs and 86 diverse sorghums) were analyzed. Amounts of acid detergent fiber (ADF), neutral detergent

fiber (NDF), lignin, crude protein, and dry matter in stems and leaves were measured by Dairy One (Ithaca, NY) using the ANKOM A200 Filter Bag Technique (ANKOM Technology, Macedon, NY) (Vogel et al., 1999). Stems and leaves were dried in a 135ºC oven for two hours and amounts of crude protein (AOAC, 1990b) and dry matter (AOAC, 1990a) in each tissue were also measured by Dairy One. NDF is a measure the three structural carbohydrates cellulose, hemi-cellulose, and lignin. ADF is a measure of cellulose and lignin only (Theander and Westerlund, 1993). Thus, cellulose was calculated as ADF minus lignin, and hemi-cellulose was calculated as NDF minus ADF.

For each trait, the selected samples (107 leaf or 168 stem) were randomly assigned to one of two groups. Two-thirds of the samples comprised the "calibration set" and experimental values from these were used to establish the best calibration equations. The second group (validation set) consisted of the remaining one-third of the samples. Equations developed from the calibration set data were evaluated for their ability to correctly predict trait values in the validation set.

Twenty eight different equation treatments, each testing different wavelengths, math treatments, and the use of the repeatability file, were applied to data from the calibration set. For each trait, the treatment that maximized the prediction of appropriate values in the validation set, as evidenced by low standard error and high R^2, was retained. This process was repeated a total of three times, each with different randomized calibration and validation sets. The best three equations (one from each repetition) were then evaluated in the full sample subset (107 leaf or 168 stem samples) and the best of these equations was used to predict trait values of all samples (1051 leaf or 1050 stem samples).

The best calibration equations are reported for each trait in Table 3.1. Criteria for selecting the best equation included low calibration standard error, low cross

validation standard error, high R^2, high heritability, and adequate ability to predict QTL (see below). The equations derived from each repetition were quite similar.

Statistical Analyses: For each trait we identified significant experimental effects (error) and their variance components, corrected the data for any confounding non-genetic effects and performed correlation and QTL analyses (Murray et al. 2008). We first analyzed the impact of genotype, environment, genotype by environment interaction and non-genetic effects such as harvest date and others (see Table 3.3 for a complete list of non-genetic effects) on trait data in a mixed model. WE05, CS05 and CS06 were considered as different environments and all variables were treated as random except genotype. Genotype was fixed so that inferences on effects would be appropriate for later data correction and analyses where values were obtained for individual RILs. SAS PROC MIXED (SAS Institute Inc. 2007) software was used to test all experimental effects at the top of Table 3.2 for significance. Only effects deemed significant (p=0.05) by type III sums of squares were retained in the reduced model (main effects of significant interactions were also retained irrespective of significance). From the reduced model, variance components were obtained using type III sums of squares, with genotype considered a random effect for appropriate population inferences. Variance components were then used to calculate broad sense heritability (Murray et al., 2008). We used type III sums of squares so that the order of effects would not influence the results.

Next, we corrected the trait data for non-genetic sources of experimental error. To accomplish this, the reduced model, determined above, was further reduced by removing the variable terms for genetic and genetic by environment interaction from the model. The residuals from this new model (containing only genetic, genetic by environment and unexplainable error) appeared to be fairly normally distributed for all traits (data not shown). These residuals were then used as input for correlation and

Table 3.1. Calibration treatment and statistics for leaf and stem NIRS

	NIRS wavelengths used	Math treatment†	N‡	Mean (SD)§	R2	SEC ¶	SECV #
Stem NDF	1108-2492	2,7,7,1	164	66.21 (6.94)	0.95	0.16	0.18
Stem cellulose	1108-2492	1,2,2,1	153	36.18 (4.59)	0.82	0.19	0.21
Stem hemi-cellulose	1108-2492	1,2,2,1	152	23.53 (2.94)	0.46	0.22	0.23
Stem lignin	1104-2496	2,6,1	165	6.09 (1.33)	0.77	0.06	0.08
Stem crude protein	1108-2492	2,7,7,1	165	4.77 (1.1)	0.94	0.03	0.04
Stem ADF	1104-2496	2,6,4,1	152	42.11 (5.26)	0.94	0.13	0.16
Stem moisture content	1104-2496	4,10,10,1	164	93.23 (1.09)	0.22	0.10	0.10
Leaf NDF	1108-2492	4,5,5,1	106	60.68 (3.99)	0.90	0.13	0.18
Leaf cellulose	1104-2496	4,10,10,1	113	33.61 (3.6)	0.84	0.14	0.18
Leaf hemi-cellulose	1108-2492	4,5,5,1	104	22.96 (3.65)	0.83	0.15	0.18
Leaf lignin	1108-2492	2,6,4,1	93	3.97 (0.82)	0.46	0.06	0.07
Leaf crude protein	1108-2492	2,6,4,1	103	11.48 (1.64)	0.98	0.02	0.04
Leaf ADF	1108-2492	3,5,5,1	109	37.5 (3.66)	0.93	0.10	0.13
Leaf moisture content	1108-2492	2,6,4,1	104	92.58 (0.65)	0.68	0.04	0.05

† Math treatment reflects the derivative number, gap (the nm wavelength over which derivative is calculated), smooth (number of points used to smooth the data), and a secondary smooth respectively. These were applied to reflectance of wavelengths by increments of 8 nm. Standard normal variate transformation (SNV) and detrend scatter correction options were used.

‡ Total number of samples used in calibration.

§ Mean percentage of sample value in calibration with standard deviation in parentheses.

¶ SEC = standard error calibration (g/kg-1).

SECV = standard error cross validation (g/kg-1).

QTL analysis. This statistical approach was similar to the 'two step model' correction methodology used in microarray experiments (Wolfinger et al., 2001).

For correlation analysis, the data (residuals from above) were used to fit a simple model with genetic and genetic by environment terms only. The predicted trait values separated genetic effects (including genetic by environment interactions) from error and were used as input for SAS PROC CORR (SAS Institute Inc., 2007) software. These correlations should be interpreted as genetic correlations within environments averaged over all environments. These measurements were determined to be most useful for comparing genetic tradeoffs across locations (environments) with results from the QTL analysis.

Single marker analysis and QTL interval mapping (IM) and composite interval mapping (CIM) were performed with Windows QTL cartographer version 2.5 (WINQTL) (Wang et al., 2007). Genetic data and the genetic map were identical to those presented in Murray et al. (2008). We used WINQTL settings 'RI1' for the cross type and 2cM for the walk speed. A conservative permutation threshold at the 0.01 significance level was obtained for each trait using 300 permutations; all other settings were the default (five control markers, 10cM window size and forward regression method). Resulting QTL maps were created using R (R Development Core Team, 2005).

Results

Phenotypic Data: A total of 32 structural carbohydrate and related traits were measured or calculated (Table 3.2). Across all locations, Rio was tall, late flowering, had high leaf/stem biomass, tillered and produced more secondary growth than BTx623 (Table 3.2). BTx623 was dwarf, early flowering, had much higher panicle and grain harvest indices and slightly higher leaf harvest indices. Transgressive

segregation was observed in the RILs for all traits (Table 3.2). Correcting raw values of all traits for significant non-genetic sources of error modestly increased observed trait normality and heritability.

Biomass Yield: High dry matter content at harvest increases energy density which, in turn, results in decreased harvest, transport, and drying costs. Total biomass dry matter content averaged 62% of fresh yield in WE05, 68% in CS05 and 66% in CS06 (data not shown). Considering the arid conditions at WE05 and the abundance of rain in the CS05 and CS06 locations, the similarity of dry matter content between the Weslaco and College Station locations was surprising. In all locations, dry matter content was lowest in the juicy stem for all lines. Heritability values for fresh stem, fresh panicle and total fresh biomass yields were higher than the corresponding values for dry material (Table 3.3). This result was likely due to multiplicative error associated with calculating dry component heritabilities and differences in sample residual moisture. Heritabilities for stem, leaf and panicle harvest indices were even higher than the dry yields and roughly equivalent to the grain harvest index heritability. This result suggests that the harvest indices of structural components may be as useful as grain harvest index for targeted selection.

Dry structural stem yields (not including stem sugar) were higher than leaf yield by an average of 10% in WE05, 12% in CS05 and 17% in CS06 RILs (Table 3.2). Stem composition, therefore, contributed more to total plant biomass than did leaf composition and this effect was more pronounced in taller plants having much greater stem harvest indices. Stem yield also had a higher proportion of genetic variance than leaf yield (fresh stem 20% vs. fresh leaf yield 8%; dry stem 21% vs. dry leaf yield 12%; 19% stem harvest index vs. 15% leaf harvest index).

Table 3.2. Trait values for Rio X BTx623 recombinant inbred and parental lines at three locations

	WE05				CS05				CS06			
Traits			RILs	RILs			RILs	RILs			RILs	RILs
Fresh biomass yield	**Rio**	**BTx623**	**MEAN (SD)[†]**	**Min, Max**	**Rio**	**BTx623**	**MEAN (SD)[†]**	**Min, Max**	**Rio**	**BTx623**	**MEAN (SD)[†]**	**Min, Max**
Fresh total biomass yield, t ha^{-1}	36.8	28	37.2 (7.5)	16.4 - 70.9	90.2	23.5	55.8 (16)	18.7 - 103	70.7	33.7	48.5 (13.3)	10.9 - 96.9
Fresh stem yield, t ha^{-1}	23.3	10.3	21.9 (6)	6.2 - 48.1	65.5	13.2	36 (12)	6.5 - 79.3	53.1	16.8	32.5 (10)	8.5 - 69.3
Fresh leaf yield, t ha^{-1}	9.3	10.4	9.8 (2.2)	4.1 - 18.6	21.1	6	15.9 (4.9)	4 - 32.5	12.8	9.8	10.6 (3.1)	1.6 - 20.1
Fresh panicle yield, t ha^{-1}	4.2	7.3	5.5 (1.1)	2.6 - 9.8	3.6	4.3	3.8 (1.7)	0.4 - 11.4	2.3	6.5	4.3 (1.7)	0.9 - 11.3
Dry biomass yield												
Dry total biomass yield, t ha^{-1}	14.1	13.4	14.1 (2.5)	7 - 24.2	31.4	6.5	17.6 (4.9)	5.5 - 31.9	24.2	12.4	16.3 (4.4)	3.9 - 31.4
Dry stem structural yield, t ha^{-1}	5.1	1.9	4.5 (1.3)	1.2 - 9.4	15.7	2.3	6.9 (2.4)	1.3 - 14.7	12.4	3.1	6.3 (2.1)	1.6 - 13.9
Dry leaf yield, t ha^{-1}	2.7	2.6	2.9 (0.7)	1.3 - 5.2	5.6	1.5	4.5 (1.4)	1.1 - 9.6	4.6	2.7	3.4 (1)	0.7 - 6.4
Dry panicle yield, t ha^{-1}	3.4	6	4.5 (1)	0 - 7.9	2.9	2.5	2.6 (1.5)	0 - 7.9	0.5	5.4	3.4 (1.4)	0 - 8.8
Dry harvest indices												
Stem dry harvest index[#]	36	18	31 (5)	15 - 47	49	34	38 (5)	12 - 51	51	25	38 (5)	24 - 61
Leaf dry harvest index[#]	19	24	21 (3)	12 35	18	19	26 (5)	13 - 46	19	22	21 (4)	14 - 46
Panicle dry harvest index[#]	24	52	32 (8)	0 - 51	9	39	15 (9)	0 - 43	2	44	21 (8)	0 - 51
Grain dry harvest index [#]	18	40	25 (6)	0 - 41	5	19	9 (7)	0 - 30	0	31	14 (7)	0 - 38
Stem composition												
Stem NDF, g kg^{-1}	67.3	68.7	69.6 (3.8)	58.3 - 80.1	55	67.6	60.4 (3.4)	50.8 - 69.7	62.7	63.8	65.4 (3.6)	55.7 - 77.3
Stem cellulose, g kg^{-1}[§]	36.9	39.3	38.7 (2.3)	31 - 45.7	29.1	37.9	32.4 (2.1)	26.8 - 38.5	34.3	35.7	36.1 (2.2)	30.5 - 43.3
Stem hemi-cellulose, g kg^{-1}[¶]	22.6	23.5	23.5 (1.4)	19.5 - 28	20.8	23.5	22.3 (1.3)	16.5 - 26.8	23.1	22.8	23.5 (1.1)	20.2 - 27.1
Stem lignin, g kg^{-1}	6.1	5.8	6.5 (0.7)	4.7 - 8.7	4.8	6.1	5.2 (0.7)	3.5 - 7.5	6.4	5.6	6.3 (0.6)	4.9 - 9.4
Stem crude protein, g kg^{-1}	3.7	4.7	4.2 (0.5)	3 - 6.4	4.2	5.7	4.5 (0.7)	2.6 - 7.3	3.7	4.4	3.6 (0.6)	2.4 - 5.9
Leaf composition												
Leaf NDF, g kg^{-1}	56.4	57.9	58.9 (2)	52.6 - 64.9	58	57.8	59.8 (2.6)	51.7 - 67	59	57.8	60.9 (2.3)	55.1 - 69.6
Leaf cellulose, g kg^{-1}[§]	33	36.1	34.3 (1.9)	29.9 - 41.9	31.9	32.8	32.7 (2.5)	26.4 - 38.5	28.8	28.5	31.9 (2)	27.5 - 37.7
Leaf hemi-cellulose, g kg^{-1}[¶]	21.7	18.6	20.8 (1.9)	14.9 - 25.9	24.1	18.4	22.8 (1.9)	16.5 - 27.3	26.1	24	24.4 (1.4)	19.1 - 27.7
Leaf lignin, g kg^{-1}	3.1	3.9	3.7 (0.4)	2.4 - 4.8	3.2	4.4	3.7 (0.5)	2.3 - 5.4	3.7	4.3	4.2 (0.5)	2.9 - 5.3
Leaf crude protein, g kg^{-1}	11.2	12.7	12 (1)	8.9 - 16.3	12.2	12.9	12.6 (1.2)	9.9 - 17.3	10.6	12.1	11.1 (1.1)	7.8 - 14.6
Other traits												
Stand density	7.5	6.5	7.7 (0.6)	4 - 9	6	4.5	5.3 (1.4)	1- 8	7	7	6.6 (1.3)	2 - 9
Tillering	6.5	2.5	6.1 (1.5)	2 - 9	4.5	1.5	4.5 (1.7)	0 - 8	8	4	5.5 (1.7)	1- 9
Mean stem thickness	2	6	3.4 (0.9)	1.5 - 7	5.8	5.8	4.3 (1)	2 7	4	5	3.5 (1)	1.5 - 7
Plant height, cm	210	130	200 (26)	130 - 274	273	119	227 (29)	119 - 297	227	123	204 (26)	109 - 259
Flowering time, days	116	109	111 (5)	104 - 123	180	161	168 (5)	157 - 185	na‡	na‡	na‡	na‡
Lodging, %	20	10	25 (15)	10 - 80	73	50	83 (16)	0 - 100	35	0	60 (34)	0 - 100

[†] Standard deviation in parentheses.

[‡] Not assayed.

[§] Cellulose = ADF - lignin.

[¶] Hemi-cellulose = NDF - ADF.

[#] Harvest index (stem, leaf, grain, or stem sugar) = yield of (stem, leaf, grain, or stem sugar) / total biomass yield.

Stem and Leaf Composition: NIRS calibration equations generally performed well (Table 3.1) and resulted in high heritabilities across stem/leaf composition traits in this population (Table 3.3). Protein had the best fitting NIRS prediction equations while moisture and, as might be expected, lignin (a complex heterogeneous biopolymer) had the lowest. Poor performance of NIRS equations for moisture was unexpected, though this outcome could have been at least partially due to low moisture variation between samples. It also seems plausible that the high temperature used to evaluate dried sample residual moisture (135^{o}C) may have affected molecules other than water (*i.e.*, residual sugars). Because levels of moisture and moisture variation were low, the effect of poor calibration did not substantially impact other NIRs measurements performed in the study.

Transgressive segregation of progeny (Table 3.2) and genetic variance (Table 3.3) were lower for biomass composition traits than for yield. In leaf tissue, BTx623 had higher levels of lignin, cellulose, and protein, while Rio had much higher amounts of hemi-cellulose. Combined, the measurements consistently explained ~70% of the total leaf composition across all lines and locations; the remaining 30% was expected to be comprised of non-structural carbohydrates (starch and sugar), and ash. Stem composition had higher heritability than leaf composition. In absolute values, stems had more lignin and cellulose and much less protein than leaves. In the stem, BTx623 had composition levels of NDF structural components equal to or higher than Rio due to lower residual non-structural carbohydrates (stem sugar). Differences in residual sugar complicated comparisons of structural carbohydrate composition.

As shown in Table 3.3, sources of experimental variance influenced biomass composition more than yield traits. For example, harvest date affected leaf and stem composition to a greater extent than biomass yield. This was due to the fact that whole plants were harvested within a three-week window, which contained the

Table 3.3. Trait heritability and variance component percentage attributable to genetic (G), environmental (E), genetic/environmental interaction (G*E), and other significant experimental effects

Traits					(E)[‡]	(E)[§]	(E)	(E)	(E)	(E)	(E)	(E)	(E)	Sample		
						Har.	Har. date*	Har. date*	North	East	South	West	Storage	grind	NIRS	Residual
Fresh yield	Heritability	G	E	G * E	Rep.	date	Strip date	Press date	border	border	border	border	box	date	date	error
Fresh total biomass yield, t ha^{-1}	0.65	0.16***	0.36***	0.05*												0.43
Fresh stem yield, t ha^{-1}	0.71	0.2***	0.39***	0.07***												0.34
Fresh leaf yield, t ha^{-1}	0.48	0.08***	0.42***	0.08***		0.05***										0.38
Fresh panicle yield, t ha^{-1}	0.65	0.18***	0.2ns	0.07**						0.11***						0.44
Dry yield																
Dry total biomass yield, t ha^{-1}	0.52	0.14***	0.16*	0.08**	0.01*											0.61
Dry stem structural yield, t ha^{-1}	0.68	0.21***	0.3***	0.12***												0.37
Dry leaf yield, t ha^{-1}	0.53	0.12***	0.26***	0.1***		0.07***										0.44
Dry panicle yield, t ha^{-1}	0.60	0.12***	0.38***	0.08***									0.04***	0.07***	0.01*	0.3
Dry harvest indices																
Stem dry harvest index	0.72	0.19***	0.35*	0.11***		0^{ns}		0.07***		0.05*			0.02*			0.22
Leaf dry harvest index	0.65	0.15***	0.34***	0.02^{ns}		0^{ns}	0.07***									0.44
Panicle dry harvest index	0.73	0.2***	0.43***	0.11***		0^{ns}	0.03**									0.22
Grain dry harvest index	0.70	0.16***	0.52***	0.1***		0^{ns}	0.03**									0.2
Stem composition																
Stem NDF, g kg^{-1}	0.61	0.06***	0.61***	0.03**		0.06*		0.03**					0.02**			0.18
Stem cellulose, g kg^{-1}	0.60	0.06***	0.66***	0.03***		0.06*		0.03***								0.17
Stem hemi-cellulose, g kg^{-1}	0.50	0.09***	0.2*	0.05*		0.14*		0.05*					0.05**			0.42
Stem lignin, g kg^{-1}	0.75	0.11***	0.61***	0.03**		0.02**							0.03***		0.05***	0.16
Stem crude protein, g kg^{-1}	0.59	0.12***	0.34***	0.11***		0^{ns}		0.07***					0.06***			0.29
Leaf composition																
Leaf NDF, g kg^{-1}	0.55	0.11***	0.16^{ns}	0.07**	0.04***	0.07^{ns}	0.15***									0.41
Leaf cellulose, g kg^{-1}	0.65	0.13***	0.1^{ns}	0.03^{ns}		0.19*	0.16***				0.04*					0.35
Leaf hemi-cellulose, g kg^{-1}	0.45	0.08***	0.24*	0.06*	0.03***	0.12***									0.01*	0.47
Leaf lignin, g kg^{-1}	0.74	0.18***	0.41***	0.06***		0.05^{ns}	0.04***									0.27
Leaf crude protein, g kg^{-1}	0.66	0.1***	0.39**	0.04**	0.03***	0.12*	0.07***									0.24*
Other traits																
Stand density	0.43	0.06***	0.47*	0.09***							0.04**	0.06***				0.27
Tillering	0.54	0.14***	0.17^{ns}	0.11*	0.05***											0.52
Mean stem thickness	0.59	0.17***	0.16*	0.09***	0.02***	0.03**										0.52
Plant height, cm	0.83	0.36***	0.07^{ns}	0.08***		0.04***					0.15***					0.29
Flowering time, days	0.68	0.21***	0.53***	0.12***												0.14
Lodging, %	0.56	0.05***	0.58^{ns}	0.04***	0**						0.14***					0.18

[†] Variance component of each significant effect divided by total variance components.
Genetic (G), environment (E), genetic by environment (G*E) and main effects were retained regardless of significance.
May not sum to one due to rounding.
[‡] Effect was nested within environment (E).
[§] Retained when non-significant due to higher level interaction term(s).
* p=0.05, **p=0.01, ***p=0.001, ns = not significant

separate optimum harvest times for leaf and stem biomass, stem sugar, and grain. During this harvesting period, plant biomass reached and passed maturity. As plants matured, older leaves and stems began to show the effects of physiological age (senescence), weathering and disease while grain matured and new tillers and axillary branches were produced. The date the plant was harvested and stripped also accounted for as much or more variation than genetics for leaf ADF, NDF, cellulose, and crude protein. The date NIRS was performed was less important for leaf and stem composition than for grain composition traits (Murray et al., 2008).

We tested for genetic correlations between composition traits in leaf and stem (Table 3.4). Correlations were low to moderate for cellulose and lignin and there was no correlation for either hemi-cellulose or protein, suggesting separate genetic controls for leaf and stem composition. Leaf and stem protein were both significantly negatively correlated with fresh and dry biomass production. However, leaf protein was positively correlated with measures of grain production (data not shown).

Other Traits: Fresh and dry biomass yield was highly correlated with plant height, and to a lesser extent, flowering time, and stand density/tillering (Table 3.4). Specifically, height was highly correlated with increased stem biomass (and thus total biomass), moderately correlated with an increase in leaf biomass and slightly correlated with a decrease in grain biomass, which highly altered harvest indices. These results contrast with those of Quinby and Karper (1954), who reported that the genes controlling height were brachytic, only affecting stem node elongation, and did not affect other traits (Morgan and Finlayson, 2001). Flowering time was correlated with height and had similar but less dramatic effects than height on leaf and stem yield. Stand density and tillering had slight positive correlations with biomass from all tissue with almost no change in harvest indices. Increasing mean stem thickness showed very low negative correlations with leaf and panicle yield with no significant

correlation with stem yield. Unlike the yield traits, leaf and stem structural composition generally showed little correlation with height, flowering time, stand density/tillering or mean stem thickness.

The conditions for lodging were extremely dependent on environment and were best evaluated in CS06. Because of this environmental dependence, trait correlation values, although highly significant, were fairly low. Increased lodging had some of the highest positive correlations with increased height, leaf cellulose, and surprisingly increased stem lignin; and the highest negative correlations with stem and leaf crude protein. Also, both lodging and height showed a significant south facing border effect (Table 3.3), even though additional border rows had been planted with dwarf grain sorghums. RILs on the south-facing border had average lodging reduced by 12% in WE05, 2% in CS05, and 50% in CS06 (data not shown). RILs on the south-facing border also showed an average reduction in height of 22cm in WE05, 3cm CS06, and 8cm in CS06 (data not shown).

Regrowth Potential: Sorghum, a weak perennial, regrows from the stalk base after each harvest, thus generating more biomass and protecting the soil. Regrowth on mown plots was measured only for the CS06 location. Regrowth constituted a large source of additional biomass, with RILs yielding an additional 7.7 – 39 t ha^{-1} fresh or, assuming the same dry matter content of the first cutting, 3 to 16.7 t ha^{-1} dry biomass (includes leaf, stem, grain, and sugar; see Table 3.5). No lines produced as much structural biomass, sugar or grain as the first harvest and there was high variation in regrowth ability, yielding between 20 to 80% of the first cutting (data not shown). The lines with the highest biomass on the first cutting had a greater variance in regrowth biomass than lower lines, but still outperformed lower biomass types (data not shown).

Table 3.4. Pearson correlation coefficients for corrected trait data

Traits	Fresh stem yield	Fresh leaf yield	Fresh panicle yield	Dry total biomass yield	Dry stem structural yield	Dry leaf yield	Dry panicle yield	Stem harvest index	Leaf harvest index	Panicle harvest index	Grain harvest index	Stem NDF	Stem cellulose	Stem hemi-cell
Fresh total biomass yield, t ha^{-1}	0.97***	0.77***	0na	0.92***	0.91***	0.82***	-0.02na	0.58***	-0.09**	-0.53***	-0.49***	0.04na	0.04na	-0.21***
Fresh stem yield, t ha^{-1}	-	0.64***	-0.11***	0.89***	0.95***	0.73***	-0.13***	0.7***	-0.2***	-0.61***	-0.57***	-0.0-	-0.02na	-0.23***
Fresh leaf yield, t ha^{-1}		-	-0.06na	0.69***	0.61***	0.93***	-0.08**	0.29***	0.42***	-0.44***	-0.41***	0.12***	0.11***	-0.0-
Fresh panicle yield, t ha^{-1}			-	0.14***	-0.18***	-0.14***	0.97***	-0.59***	-0.39***	0.76***	0.78***	0.05na	0.12***	-0.19***
Dry total biomass yield, t ha^{-1}				-	0.9***	0.78***	0.12***	0.5***	-0.22***	-0.41***	-0.35***	-0.0-	0.0-	-0.28***
Dry stem structural yield, t ha^{-1}					-	0.73***	-0.19***	0.78***	-0.21***	-0.66***	-0.61***	0na	-0.02na	-0.2***
Dry leaf yield, t ha^{-1}						-	-0.16***	0.42***	0.36***	-0.54***	-0.49***	0.1**	0.09**	-0.05na
Dry panicle yield, t ha^{-1}							-	-0.58***	-0.4***	0.76***	0.79***	0.08*	0.15***	-0.17***
Stem dry harvest index								-	-0.15***	-0.88***	-0.84***	0.02na	-0.02na	-0.06na
Leaf dry harvest index									-	-0.21***	-0.23***	0.16***	0.12***	0.34***
Panicle dry harvest index										-	0.98***	0.05na	0.11***	0.0-
Grain dry harvest index											-	0.04na	0.11***	-0.02na
Stem NDF, g kg^{-1}												-	0.97***	0.72***
Stem cellulose, g kg^{-1}													-	0.64***

Traits	Stem lignin	Stem crude protein	Leaf NDF	Leaf cellulose	Leaf hemi-cell	Leaf lignin	Leaf crude protein	Stand density	Tillering	Mean stem thickness	Plant height	Flowering time	Lodging	Brix
Fresh total biomass yield, t ha^{-1}	0.29***	-0.36***	0.14***	0.06na	0.14***	0.2***	-0.1**	0.27***	0.29***	-0.03na	0.7***	0.29***	0.11***	0.24***
Fresh stem yield, t ha^{-1}	0.31***	-0.38***	0.13***	0.03na	0.16***	0.19***	-0.1**	0.19***	0.23***	0.02na	0.8***	0.35***	0.15***	0.34***
Fresh leaf yield, t ha^{-1}	0.06na	-0.13***	0.2***	0.11***	0.2***	0.11***	-0.13***	0.4***	0.36***	-0.13***	0.3***	0.13***	0.02na	0.06na
Fresh panicle yield, t ha^{-1}	0.0-	-0.3***	-0.18***	-0.07*	-0.31***	0.09**	0.16***	0.12***	0.12***	-0.16***	-0.2***	-0.41***	0.0-	-0.37***
Dry total biomass yield, t ha^{-1}	0.29***	-0.47***	0.05	0.02	0.05	0.18***	-0.14***	0.34***	0.36***	-0.15***	0.68***	0.2***	0.08**	0.33***
Dry stem structural yield, t ha^{-1}	0.34***	-0.4***	0.11***	0.03	0.15***	0.18***	-0.13***	0.21***	0.26***	-0.02	0.79***	0.35***	0.1**	0.44***
Dry leaf yield, t ha^{-1}	0.15***	-0.19***	0.21***	0.14***	0.19***	0.09**	-0.22***	0.42***	0.4***	-0.18***	0.46***	0.17***	0.07*	0.21***
Dry panicle yield, t ha^{-1}	0.04na	-0.32***	-0.17***	-0.05	-0.31***	0.09**	0.17***	0.11***	0.11***	-0.17***	-0.18***	-0.4***	0.02	-0.38***
Stem dry harvest index	0.35***	-0.2***	0.19***	0.06*	0.21***	0.17***	-0.12***	0.0-	0.05	0.1**	0.73***	0.46***	0.1**	0.46***
Leaf dry harvest index	-0.2***	0.46***	0.18***	0.18***	0.17***	-0.16***	-0.15***	0.17***	0.1**	-0.04	-0.35***	-0.02	-0.12***	-0.17***
Panicle dry harvest index	-0.18***	0na	-0.18***	-0.05	-0.31***	-0.04	0.2***	-0.05	-0.05	-0.11***	-0.57***	-0.51***	-0.04	-0.5***
Grain dry harvest index	-0.16***	-0.06na	-0.19***	-0.03	-0.33***	-0.05	0.17***	-0.03	-0.02	-0.14***	-0.52***	-0.53***	-0.03	-0.48***
Stem NDF, g kg^{-1}	0.61***	-0.02na	0.33***	0.35***	-0.11***	0.26***	0.13***	0.13***	0.16***	-0.1**	0.05	-0.11***	-0.02	-0.53***
Stem cellulose, g kg^{-1}	0.62***	-0.11***	0.33***	0.34***	-0.11***	0.27***	0.11***	0.13***	0.15***	-0.08**	0.04	-0.13***	0.0-	-0.56***
Stem hemi-cellulose, g kg^{-1}	0.22***	0.44***	0.19***	0.16***	0.04	0.02	0.08**	-0.1**	-0.07*	0.08*	-0.23***	-0.07*	-0.17***	-0.36***
Stem lignin, g kg^{-1}	-	-0.46***	0.21***	0.25***	-0.15***	0.31***	0.02	0.12***	0.18***	-0.12***	0.5***	0.06	0.12***	-0.15***
Stem crude protein, g kg^{-1}		-	0.06*	-0.02	0.19***	-0.12***	-0.0-	-0.23***	-0.26***	0.22***	-0.51***	0.08**	-0.23***	-0.1**
Leaf NDF, g kg^{-1}			-	0.64***	0.42***	0.46***	-0.29***	0.09**	0.07*	-0.02	0.11***	0.09**	0.1**	-0.24***
Leaf cellulose, g kg^{-1}				-	-0.27***	0.14***	-0.43***	0.02	0.06*	-0.07*	0.1**	-0.16***	0.27***	-0.32***
Leaf hemi-cellulose, g kg^{-1}					-	-0.0-	-0.12***	0.05	0.03	0.08*	0.04	0.36***	-0.16***	0.2***
Leaf lignin, g kg^{-1}						-	0.2***	0.15***	0.12***	-0.03	0.15***	0.15***	-0.1**	-0.17***
Leaf crude protein, g kg^{-1}							-	-0.16***	-0.16***	0.14***	-0.1**	-0.03	-0.17***	-0.19***
Stand density								-	0.72***	-0.58***	0.05	-0.09**	-0.1**	0.0-

Table 3.4. (Continued)

Traits	Stem lignin	Stem crude protein	Leaf NDF	Leaf cellulose	Leaf hemi-cell	Leaf lignin	Leaf crude protein	Stand density	Tillering	Mean stem thickness	Plant height	Flowering time	Lodging	Brix
Tillering									-	-0.75***	0.13***	-0.15***	-0.07*	0.04
Mean stem thickness										-	-0.02	0.33***	0	-0.03
Plant height, cm											-	0.29***	0.25***	0.37***
Flowering time, days												-	-0.13***	0.39***
Lodging, %													-	-0.11***

* p=0.05, **p=0.01, ***p=0.001, ns = not significant

Regrowth fresh biomass had moderate genetic correlation with fresh biomass (0.44***). Regrowth brix showed an average decrease in raw values from primary growth by 20% (Table 3.5) and had low but significant correlation with primary growth brix (0.21***). The highest correlation was between regrowth fresh biomass and primary height (0.71***). The experimental variance for regrowth in the two randomized complete block replicates was highly significant, suggesting that environmental conditions within the field had large effects on regrowth ability (Table 3.6).

QTL Analysis: Corrections on raw data for identifiable sources of non-genetic error were made before performing QTL analyses. For most traits correcting raw data modestly improved trait normality, heritability, peak LOD scores, and narrowed the marker LOD intervals noticeably but not substantially (data not shown). Using both raw and corrected data, interval mapping (IM) and composite interval mapping (CIM) showed potential QTL likelihood peaks in the same genetic location for all traits (data not shown). CIM using corrected data (Figure 1, Table 3.7) identified more of these peaks as significant and with less genetic distance within the 1 and 2 LOD intervals. Many of these QTL were supported by single marker analysis which also identified additional QTL not significant under the stringent permutation thresholds used for IM and CIM (Table 3.7).

Across the three locations, QTL were identified for all but one of the measured traits (Figure 1), regrowth juice yield. In all environments, the sweet sorghum parent, Rio, provided alleles for taller plants and increased stem and total biomass.

Table 3.5. Regrowth Trait values for Rio X BTx623 recombinant inbred and parental lines in CS06

Traits	Rio	BTx623	RILs MEAN (SD)[‡]	RILs Min, Max
Regrowth fresh biomass, t ha^{-1}	24.2	10.5	19.2 (6.8)	3.5 - 41.3
Regrowth juice weight, t ha^{-1}	7.7	4.2	6 (2)	1 - 12
Regrowth brix, °brix	14.8	12	13.3 (2.1)	6.2 - 18.1R
Regrowth height, cm	232	123	188 (26)	97 - 254

[†] Standard deviation in parentheses.

Table 3.6. Regrowth trait heritability and variance component percentage attributable to genetic (G), environmental (E), genetic/environmental interaction (G*E) and other effects

Traits	Heritability	G	E	G * E	(E)[‡] Rep.	(E)[§] Har. date	(E) Har. date* Strip date	(E) Har. date* Press date	(E) North border	(E) East border	(E) South border	(E) West border	(E) Storage box	Sample grind date	NIRS date	Residual error
Regrowth fresh biomass, t ha^{-1}	**0.58**	**0.41*** **	**na**	**na**												**0.59**
Regrowth juice weight, t ha^{-1}	**0.51**	**0.14*** **	**na**	**na**	**0.59*** **											**0.27**
Regrowth brix	**0.35**	**0.13** **	**na**	**na**	**0.39*** **											**0.48**
Regrowth height, cm	**0.70**	**0.29*** **	**na**	**na**	**0.21*** **				**0.17*** **		**0.09***					**0.24**

[†] Variance component of each significant effect divided by total variance components.
Genetic (G), environment (E), genetic by environment (G*E) and main effects were retained regardless of significance.
May not sum to one due to rounding.
[‡] Effect was nested within environment (E).

Figure 3.1. QTL positions in three locations

Colored bars represent QTL detected by CIM. Alleles that increased trait values inherited from BTx623 (grain parent) are shown in blue (WE05), green (CS05) and orange (CS06). Alleles that increased trait values inherited from Rio (sweet parent) are shown in red (WE05), purple (CS05) and pink (CS06). Traits names are shown on the left side, markers are shown as vertical lines on the bottom above chromosome numbers. The length of each bar represents the 2-LOD QTL interval, and height represents the variance (R^2) explained by the QTL. QTL for traits were not always detected in all three locations. QTL for regrowth and percent structural solids are presented in Table 3.7. Exact QTL locations, LOD scores, and R^2 can be found in Table 3.7.

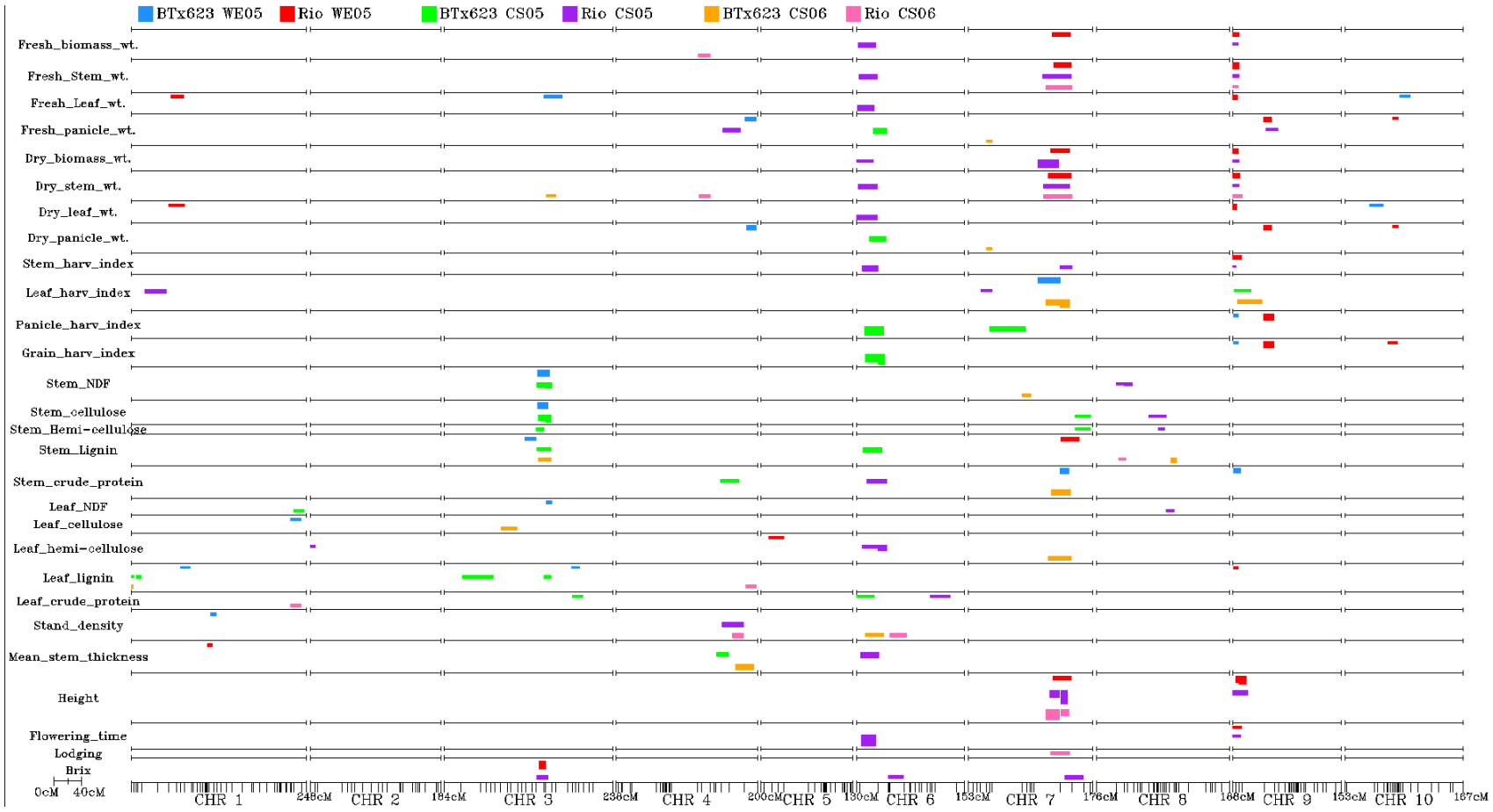

BTx623 WE05
Rio WE05
BTx623 CS05
Rio CS05
BTx623 CS06
Rio CS06
Fresh_biomass_wt.
Fresh_Stem_wt.
Fresh_Leaf_wt.
Fresh_panicle_wt.
Dry_biomass_wt.
Dry_stem_wt.
Dry_leaf_wt.
Dry_panicle_wt.
Stem_harv_index
Leaf_harv_index
Panicle_harv_index
Grain_harv_index
Stem_NDF
Stem_cellulose
Stem_Hemi-cellulose
Stem_Lignin
Stem_crude_protein
Leaf_NDF
Leaf_cellulose
Leaf_hemi-cellulose
Leaf_lignin
Leaf_crude_protein
Stand_density
Mean_stem_thickness
Height
Flowering_time
Lodging
Brix
0cM
40cM
CHR 1
248cM
CHR 2
184cM
CHR 3
236cM
CHR 4
200cM
CHR 5
130cM
CHR 6
153cM
CHR 7
176cM
CHR 8
166cM
CHR 9
153cM
CHR 10
167cM

Table 3.7. Positions of QTLs identified by CIM and single marker analyses in three locations.

A. QTLs identified by CIM analyses

QTL information				CIM QTL R^2 and peak position with 2LOD interval					Single marker analysis peaks and LOD score	
Location	Trait	Effect	Chr.	LOD score	R^2	Center cM	2 LOD left cM	2 LOD right cM	cM (LOD)	cM (LOD)
WE05	Fresh total biomass yield, t ha^{-1}	R	7	6	0.15	133.8	118.3	144	129.8 (3.1)	
WE05	Fresh total biomass yield, t ha^{-1}	R	9	7.9	0.15	3.5	0	8.3	3.5 (6.1)	
WE05	Fresh stem yield, t ha^{-1}	R	7	5.6	0.15	137.8	121	145.4	129.8 (3)	
WE05	Fresh stem yield, t ha^{-1}	R	9	9.5	0.19	3.5	0	8.3	3.5 (7.1)	
WE05	Fresh leaf yield, t ha^{-1}	R	9	7.2	0.14	2	0	6.2	0 (4)	
WE05	Fresh leaf yield, t ha^{-1}	R	1	4.2	0.08	65.6	55.7	69.3	63.6 (3.1)	
WE05	Fresh leaf yield, t ha^{-1}	R	1	4.6	0.09	71.3	69.3	73.3		
WE05	Fresh leaf yield, t ha^{-1}	B	10	3.5	0.07	83.5	77.6	91.9	46.5 (2.9)	
WE05	Fresh leaf yield, t ha^{-1}	B	3	4.8	0.10	147.4	140.3	166.7	152 (3.8)	190.1 (2.7)
WE05	Fresh panicle yield, t ha^{-1}	R	9	7.1	0.16	45.8	43.1	54	43.8 (3.5)	
WE05	Fresh panicle yield, t ha^{-1}	R	10	4	0.08	69.9	67.5	74.8		
WE05	Fresh panicle yield, t ha^{-1}	B	4	5	0.12	195.1	181.7	197.1	181.1 (2.5)	
WE05	Dry total biomass yield, t ha^{-1}	R	7	5.6	0.13	131.8	116.6	142.9	129.8 (4.4)	
WE05	Dry total biomass yield, t ha^{-1}	R	9	8.1	0.16	3.5	0	7.7	3.5 (5.2)	
WE05	Dry stem structural yield, t ha^{-1}	R	7	5.3	0.14	135.8	113.2	145.1	129.8 (3.3)	
WE05	Dry stem structural yield, t ha^{-1}	R	9	7.8	0.15	3.5	0	9.5	3.5 (6.1)	
WE05	Dry leaf yield, t ha^{-1}	R	9	7.8	0.17	2	0	5.6	0 (5)	
WE05	Dry leaf yield, t ha^{-1}	R	1	3.8	0.08	65.6	52	74.4		
WE05	Dry leaf yield, t ha^{-1}	B	10	3.9	0.07	46.5	34.7	54		
WE05	Dry panicle yield, t ha^{-1}	R	9	6.7	0.15	45.8	43.1	54.7	43.8 (3)	
WE05	Dry panicle yield, t ha^{-1}	R	10	4.3	0.08	69.9	67.5	74.6		
WE05	Dry panicle yield, t ha^{-1}	B	4	6.4	0.15	195.1	184.4	197.1		
WE05	Stem dry harvest index	R	9	5.4	0.12	3.5	0	11.6	3.5 (4.3)	
WE05	Leaf dry harvest index	B	7	3.8	0.21	119.2	99	129.8		
WE05	Panicle dry harvest index	R	9	10.3	0.22	45.8	43.2	57.1		
WE05	Panicle dry harvest index	B	9	4.6	0.10	3.5	0.3	7.8	3.5 (5.6)	
WE05	Grain dry harvest index	R	9	10.9	0.22	45.8	43.1	58		
WE05	Grain dry harvest index	R	10	4.2	0.08	65.5	60.5	74.3		
WE05	Grain dry harvest index	B	9	3.9	0.08	3.5	0.5	7.7	3.5 (6.1)	
WE05	Stem NDF, g kg^{-1}	B	3	9.4	0.20	140.1	131.9	148.4	140.4 (8.7)	
WE05	Stem cellulose, g kg^{-1}§	B	3	10.7	0.23	140.1	132.4	146.3	140.4 (9.2)	
WE05	Stem lignin, g kg^{-1}	R	7	6.4	0.15	141.8	131.4	156.8	146.4 (5.4)	
WE05	Stem lignin, g kg^{-1}	B	3	5.2	0.11	121.2	113.9	130.1	124.5 (5.4)	
WE05	Stem crude protein, g kg^{-1}	B	9	8.1	0.16	5.5	1	10.7	3.5 (5.4)	
WE05	Stem crude protein, g kg^{-1}	B	7	7	0.17	131.8	129.8	141.8	129.8 (5.2)	
WE05	Leaf NDF, g kg^{-1}	B	3	3.7	0.08	145.4	144.3	151.7	140.4 (3.7)	
WE05	Leaf cellulose, g kg^{-1}	B	1	4.5	0.09	226	225.3	239.3	217.8 (3.6)	226 (4.5)
WE05	Leaf hemi-cellulose, g kg^{-1}	R	5	4.1	0.09	24.8	11.2	32.6	24.8 (3.2)	
WE05	Leaf lignin, g kg^{-1}	R	9	4.9	0.09	3.5	0.3	7.5	3.5 (3.4)	
WE05	Leaf lignin, g kg^{-1}	B	1	3.6	0.06	74.4	69.3	83	74.3 (3.5)	
WE05	Leaf lignin, g kg^{-1}	B	3	3.7	0.06	185.7	180	191.1	185.6 (5.6)	
WE05	Stand density	B	1	4.9	0.11	116.5	111.5	120		
WE05	Mean stem thickness	R	1	4.7	0.10	109.4	107.5	114		
WE05	Plant height, cm	R	9	11.2	0.23	5.5	3.5	7.5	3.5 (8.6)	

Table 3.7. (Continued)

WE05	Plant height, cm	R	9	13.9	0.28	10.8	8.8	18.9		
WE05	Plant height, cm	R	7	6.5	0.14	133.8	120.1	145.9	129.8 (4.1)	
WE05	Flowering time, days	R	9	4.6	0.10	2	0	12	3.5 (3.8)	
WE05	Brix, °brix	R	3	11.9	0.28	140.4	134.1	142.8	140.4 (8)	158 (4.5)
WE05	Brix, °brix	R	4	3.4	0.09	175.4	159.2	195.8		
WE05	Stem cellulose (% structural solids)	R	9	4.6	0.09	3.5	0	7.5	3.5 (4.8)	
WE05	Stem cellulose (% structural solids)	B	3	4.7	0.10	143.4	132.7	148.9	143.4 (3.6)	
WE05	Stem hemi-cellulose (% structural solids)	B	9	8.7	0.17	18.8	12.8	35.1	3.5 (3.2)	
WE05	Stem hemi-cellulose (% structural solids)	B	9	4.8	0.15	3.5	0.1	6.9	3.5 (6.1)	
WE05	Stem lignin (% structural solids)	R	7	6	0.16	137.8	129.7	156.4	146.4 (4)	
WE05	Stem lignin (% structural solids)	R	9	4.4	0.10	7.5	3.5	12	3.5 (3)	
WE05	Stem crude protein (% structural solids)	R	3	5.7	0.12	136.1	133.7	140.1	124.5 (4.7)	143.4 (5.3)
WE05	Stem crude protein (% structural solids)	B	9	5	0.09	3.5	0	27.4	3.5 (3.8)	
WE05	Stem crude protein (% structural solids)	B	7	8.4	0.38	133.8	129.8	139.3	129.8 (5.4)	
WE05	Stem crude protein (% structural solids)	B	7	9.2	0.26	123.2	116	129.8		
CS05	Fresh total biomass yield, t ha^{-1}	R	6	5.6	0.17	13.1	2.6	27.3		
CS05	Fresh total biomass yield, t ha^{-1}	R	9	3.8	0.08	2	0	7.5		
CS05	Fresh stem yield, t ha^{-1}	R	7	4.5	0.12	133.8	104.7	145.1	129.8 (4.3)	
CS05	Fresh stem yield, t ha^{-1}	R	6	5.9	0.17	15.1	3.7	28.9		
CS05	Fresh stem yield, t ha^{-1}	R	9	5.3	0.10	2	0	8.8	3.5 (3.1)	
CS05	Fresh leaf yield, t ha^{-1}	R	6	6.3	0.18	11.1	1	25	5.1 (4)	
CS05	Fresh panicle yield, t ha^{-1}	R	4	5.8	0.15	163.4	150.7	175.5	159.4 (2.5)	
CS05	Fresh panicle yield, t ha^{-1}	R	9	3.9	0.07	58.4	46.2	63.2		
CS05	Fresh panicle yield, t ha^{-1}	B	6	7.6	0.19	34.8	23.1	42.1	30.7 (4.4)	
CS05	Dry total biomass yield, t ha^{-1}	R	7	5.9	0.27	113.2	98.6	127.3	129.8 (3.2)	
CS05	Dry total biomass yield, t ha^{-1}	R	6	4.9	0.10	5.1	0.6	24.1		
CS05	Dry total biomass yield, t ha^{-1}	R	9	4.1	0.08	2	0	8.8		
CS05	Dry stem structural yield, t ha^{-1}	R	7	5.8	0.13	131.8	106.1	142.9	129.8 (4.5)	
CS05	Dry stem structural yield, t ha^{-1}	R	6	5.7	0.15	13.1	2.1	29	5.1 (3)	
CS05	Dry stem structural yield, t ha^{-1}	R	9	4.4	0.08	5.5	0	8.8		
CS05	Dry leaf yield, t ha^{-1}	R	6	4.5	0.14	13.1	0.1	29.1	5.1 (3.2)	
CS05	Dry panicle yield, t ha^{-1}	B	6	7.2	0.16	32.8	17.8	41.6	30.7 (5.1)	
CS05	Stem dry harvest index	R	6	6	0.19	21.1	7.6	30.8	30.7 (3.1)	
CS05	Stem dry harvest index	R	7	3.4	0.12	137.8	129.8	146.4		
CS05	Stem dry harvest index	R	9	3.4	0.07	0	0	3.5		
CS05	Leaf dry harvest index	R	1	6.1	0.13	34.5	18.5	49.3	36.6 (5.6)	
CS05	Leaf dry harvest index	R	7	3.7	0.07	30.6	17.8	33.1	30.5 (2.9)	
CS05	Leaf dry harvest index	B	9	4.6	0.08	8.8	1.8	24.9	8.8 (2.9)	
CS05	Panicle dry harvest index	B	7	3.7	0.16	55.1	30.6	80.7		
CS05	Panicle dry harvest index	B	6	9.8	0.29	23.1	11.6	38.5	30.7 (5.8)	
CS05	Grain dry harvest index	B	6	11.7	0.33	32.8	30.8	38.7	30.7 (6.7)	
CS05	Grain dry harvest index	B	6	10.7	0.25	23.1	12.8	30.8		
CS05	Stem NDF, g kg^{-1}	R	8	5.5	0.10	33	28.7	39.5	29 (4.5)	
CS05	Stem NDF, g kg^{-1}	R	8	5.6	0.13	43.7	39.5	51	43.7 (4)	
CS05	Stem NDF, g kg^{-1}	B	3	8.2	0.18	145.4	143.4	151.5	140.4 (6.1)	
CS05	Stem NDF, g kg^{-1}	B	3	6.9	0.16	138.1	131.2	143.4		

Table 3.7. (Continued)

CS05	Stem cellulose, g kg^{-1}§	R	8	5.2	0.09	91.3	88.3	98.8	117 (3.1)	
CS05	Stem cellulose, g kg^{-1}§	R	8	4.3	0.08	76.5	74.3	86.5	43.7 (3.5)	
CS05	Stem cellulose, g kg^{-1}§	B	7	3.9	0.09	167.4	150.6	172.4		
CS05	Stem cellulose, g kg^{-1}§	B	3	10.7	0.23	145.4	143.4	151.4		
CS05	Stem cellulose, g kg^{-1}§	B	3	9.3	0.20	138.1	132.7	142.4	140.4 (6)	
CS05	Stem hemi-cellulose, g kg^{-1}	R	8	4.1	0.08	91.3	88	96.2	43.7 (2.5)	91.3 (2.7)
CS05	Stem hemi-cellulose, g kg^{-1}	B	7	3.9	0.08	163.4	151	172.4	163.4 (3.5)	
CS05	Stem hemi-cellulose, g kg^{-1}	B	3	4.2	0.10	138.1	129.9	140.4	103.8 (4.2)	130.1 (3.8)
CS05	Stem lignin, g kg^{-1}	B	3	6	0.12	140.1	130.8	151.1	124.5 (5.2)	140.4 (5.2)
CS05	Stem lignin, g kg^{-1}	B	6	5.9	0.15	25.1	8.8	35.5	30.7 (4.7)	
CS05	Stem crude protein, g kg^{-1}	R	6	5.4	0.13	34.8	14.5	42.8	30.7 (3.4)	
CS05	Stem crude protein, g kg^{-1}	B	4	4.5	0.09	158.1	147.7	173.3	148.1 (2.6)	
CS05	Leaf NDF, g kg^{-1}	R	8	3.7	0.07	104	99.5	109.8		
CS05	Leaf NDF, g kg^{-1}	B	1	4	0.08	240.2	229.5	244.2		
CS05	Leaf hemi-cellulose, g kg^{-1}	R	6	3.8	0.11	21.1	7.5	30.8		
CS05	Leaf hemi-cellulose, g kg^{-1}	R	6	4.6	0.18	36.8	30.8	42.8	30.7 (2.9)	
CS05	Leaf hemi-cellulose, g kg^{-1}	R	2	4	0.09	2	0	6.4		
CS05	Leaf lignin, g kg^{-1}	B	1	6.5	0.11	13.1	7.1	13.8		
CS05	Leaf lignin, g kg^{-1}	B	3	5.9	0.11	59.3	51.3	69.5	59.5 (4.6)	
CS05	Leaf lignin, g kg^{-1}	B	3	4.7	0.13	142.4	140.4	151.4	140.4 (5)	158 (3.6)
CS05	Leaf lignin, g kg^{-1}	B	1	4.1	0.08	0	0	2.8	0 (5.4)	
CS05	Leaf lignin, g kg^{-1}	B	3	6.2	0.10	42.9	26.1	51.3		
CS05	Leaf crude protein, g kg^{-1}	R	6	4	0.08	109.8	103.6	131.6		
CS05	Leaf crude protein, g kg^{-1}	B	3	3.7	0.07	184.8	180.8	196.1		
CS05	Leaf crude protein, g kg^{-1}	B	6	4.4	0.09	5.1	0	24.4		
CS05	Stand density	R	4	5.1	0.17	169.4	149.6	179.4	159.4 (3.9)	
CS05	Tillering	R	4	7.8	0.25	169.4	159.3	179.4	138.9 (3.4)	159.4 (5.2)
CS05	Mean stem thickness	R	6	5.1	0.18	17.1	5.2	31.5		
CS05	Mean stem thickness	B	4	5.5	0.12	148.1	141.2	158	148.1 (4.2)	
CS05	Plant height, cm	R	7	17.5	0.44	133.8	131.3	139.8		
CS05	Plant height, cm	R	7	7.4	0.24	125.2	115	129.2	129.8 (9.3)	
CS05	Plant height, cm	R	9	8.8	0.15	2	0	8.8	3.5 (4.9)	
CS05	Plant height, cm	R	9	7.8	0.15	12.8	9.6	21.2	12.7 (4.8)	
CS05	Flowering time, days	R	6	11.9	0.36	15.1	6.8	26.5	5.1 (6.4)	
CS05	Flowering time, days	R	9	4.5	0.09	2	0	11.2		
CS05	Brix, °brix	R	7	4.5	0.12	137.8	136.3	146.4		
CS05	Brix, °brix	R	7	5	0.14	152.4	146.4	162.4		
CS05	Brix, °brix	R	3	5.5	0.12	138.1	130.7	146.1	140.4 (3.4)	
CS05	Brix, °brix	R	6	4.3	0.10	51.4	44.4	65.8	47.4 (4.7)	
CS05	Stem cellulose (% structural solids)	R	4	3.6	0.07	158.1	147.7	174.7	138.9 (3.4)	159.4 (3.4)
CS05	Stem cellulose (% structural solids)	R	9	4	0.08	145.4	137.5	149.4	145.4 (2.9)	
CS05	Stem cellulose (% structural solids)	B	3	4.2	0.08	103.8	95.3	108.8	103.8 (3.5)	
CS05	Stem cellulose (% structural solids)	B	6	4.3	0.12	23.1	5	41.3		
CS05	Stem lignin (% structural solids)	B	6	6.5	0.18	25.1	11.9	35.7	30.7 (6.4)	
CS05	Stem crude protein (% structural solids)	R	6	6.5	0.16	25.1	11.7	39.5	30.7 (4.5)	
CS05	Stem crude protein (% structural solids)	R	1	4	0.07	240.2	235.3	244.2		
CS05	Stem crude protein (% structural solids)	B	8	4.1	0.08	0	0	19.8	0 (3.4)	
CS06	Fresh total biomass yield, t ha^{-1}	R	4	3.6	0.10	126.7	115.5	133.1		

Table 3.7. (Continued)

CS06	Fresh stem yield, t ha^{-1}	R	7	4.1	0.12	135.8	110.2	146.1		
CS06	Fresh stem yield, t ha^{-1}	R	9	4.1	0.08	3.5	0	7.5	3.5 (2.6)	
CS06	Fresh panicle yield, t ha^{-1}	B	7	4.3	0.10	29.8	25.8	33.1	30.5 (3.1)	
CS06	Dry stem structural yield, t ha^{-1}	R	7	4.1	0.12	135.8	106.8	146.1		
CS06	Dry stem structural yield, t ha^{-1}	R	4	4.3	0.11	126.7	116.7	133		
CS06	Dry stem structural yield, t ha^{-1}	R	9	5.1	0.10	3.5	0	12.5	3.5 (3.3)	
CS06	Dry stem structural yield, t ha^{-1}	B	3	3.5	0.07	152	144.1	158		
CS06	Dry panicle yield, t ha^{-1}	B	7	3.8	0.08	30.6	25.8	33.1	30.5 (3)	
CS06	Leaf dry harvest index	B	9	4.8	0.14	18.8	6.7	40.6	12.7 (3.3)	
CS06	Leaf dry harvest index	B	7	5.6	0.28	133.8	129.8	142.8	129.8 (3)	
CS06	Leaf dry harvest index	B	7	5.8	0.19	123.2	109.5	129.8		
CS06	Stem NDF, g kg^{-1}	B	7	4.6	0.10	82.8	75.8	88.9	82.8 (3.1)	
CS06	Stem lignin, g kg^{-1}	R	8	4.1	0.08	37	31.3	42.4	39.5 (3.5)	
CS06	Stem lignin, g kg^{-1}	B	3	4.5	0.09	140.4	132.8	150.7	115.2 (3.8)	140.4 (4.2)
CS06	Stem lignin, g kg^{-1}	B	8	5.6	0.15	109.8	106.2	113.3		
CS06	Stem crude protein, g kg^{-1}	B	7	6.3	0.18	133.8	117.4	143.8	129.8 (5)	
CS06	Leaf cellulose, g kg^{-1}	B	3	4.5	0.11	90.9	80.5	102.5	65.9 (3.6)	
CS06	Leaf hemi-cellulose, g kg^{-1}	B	7	4	0.13	133.8	113.1	145.9	129.8 (3.4)	
CS06	Leaf lignin, g kg^{-1}	R	4	4.2	0.11	195.1	183.4	197.1		
CS06	Leaf lignin, g kg^{-1}	B	1	6.9	0.14	0	0	2.5	0 (5.9)	14.5 (2.7)
CS06	Leaf crude protein, g kg^{-1}	R	1	3.7	0.09	233.3	225.3	239.7	219.1 (3.1)	229.3 (3.2)
CS06	Stand density	R	6	4	0.13	57.4	47.1	70.7		
CS06	Stand density	R	4	5.3	0.14	165.4	163.7	179.4	140.4 (4.5)	159.4 (4.7)
CS06	Stand density	B	6	4.7	0.09	30.8	12.9	38.2		
CS06	Tillering	R	6	5	0.19	61.4	52.2	71.8		
CS06	Tillering	R	4	5	0.12	156.1	147.7	178.5	138.9 (4.1)	159.4 (4.3)
CS06	Mean stem thickness	B	4	7.6	0.19	183.1	168.5	194	181.1 (5.2)	
CS06	Plant height, cm	R	7	4.8	0.21	133.8	130.5	142.5		
CS06	Plant height, cm	R	7	10.5	0.33	125.2	109.7	129.2	129.8 (7.1)	
CS06	Lodging, %	R	7	4.7	0.10	129.8	116.8	143.4	129.8 (4.4)	163.4 (3.3)
CS06	Regrowth fresh biomass, t ha^{-1}	B	3	4.7	0.11	202.2	196.2	204.2	210.4 (3.7)	
CS06	Regrowth fresh biomass, t ha^{-1}	B	3	4.2	0.11	187.7	185.7	195.9	143.4 (3.1)	190.1 (4)
CS06	Regrowth brix, °brix	R	3	3.9	0.08	140.4	130.3	149.5	140.4 (2.7)	
CS06	Regrowth height, cm	R	7	3.9	0.22	135.8	131.5	143.3	146.4 (5.9)	
CS06	Regrowth height, cm	R	7	10.7	0.34	121.2	101.4	129.2		
CS06	Regrowth height, cm	R	9	4.8	0.11	14.8	8.8	31.9	12.7 (2.9)	
CS06	Regrowth height, cm	R	9	4	0.09	2	0	8.8		
CS06	Regrowth maturity CS06	R	4	4.9	0.12	161.4	148.8	173.6	138.9 (2.9)	159.4 (4.2)
CS06	Stem lignin (% structural solids)	R	7	5.6	0.15	139.8	129.8	151.6	146.4 (3.6)	
CS06	Stem lignin (% structural solids)	R	9	6.3	0.12	7.5	1.3	12.8	8.8 (3.3)	
CS06	Stem crude protein (% structural solids)	B	7	6.7	0.20	133.8	120.7	143.6	129.8 (3.8)	

B. QTLs identified by single marker analysis only

QTL information				CIM QTL R^2 and peak position with 2LOD interval					Single marker analysis peaks and LOD score	
Location	Trait	Effect	Chr.	LOD score	R^2	Center cM	2 LOD left cM	2 LOD right cM	cM (LOD)	cM (LOD)
WE05	Fresh leaf yield, t ha^{-1}	R	1						3.1 (2.7)	
WE05	Fresh leaf yield, t ha^{-1}	B	3						22.9 (2.8)	
WE05	Dry leaf yield, t ha^{-1}	R	1						3.1 (3.1)	

Table 3.7. (Continued)

WE05	Leaf dry harvest index	R	1	3.1 (3.8)	36.6 (4)
WE05	Stem lignin, g kg^{-1}	B	6	30.7 (3.5)	
WE05	Stem crude protein, g kg^{-1}	R	3	124.5 (2.5)	
WE05	Leaf NDF, g kg^{-1}	B	3	65.9 (2.6)	94.1 (2.5)
WE05	Leaf NDF, g kg^{-1}	R	8	111.8 (2.7)	
WE05	Leaf cellulose, g kg^{-1}	B	3	130.1 (2.6)	
WE05	Leaf lignin, g kg^{-1}	B	1	0 (2.8)	
WE05	Leaf lignin, g kg^{-1}	B	1	121.1 (3.2)	
WE05	Leaf lignin, g kg^{-1}	B	3	103.8 (3.1)	143.4 (4.2)
WE05	Leaf crude protein, g kg^{-1}	R	9	114.6 (2.6)	
WE05	Flowering time, days	R	2	0 (2.8)	
WE05	Stem cellulose (% structural solids)	B	3	109.3 (2.6)	
WE05	Stem hemi-cellulose (% structural solids)	R	3	124.5 (2.9)	152 (2.5)
WE05	Stem hemi-cellulose (% structural solids)	R	6	30.7 (2.5)	
WE05	Stem lignin (% structural solids)	B	3	124.5 (2.7)	
WE05	Stem lignin (% structural solids)	B	6	30.7 (3.8)	
WE05	Lodging, %	R	3	88.9 (2.6)	
WE05	Lodging, %	R	7	163.4 (2.6)	
CS05	Fresh total biomass yield, t ha^{-1}	R	7	16.8 (2.6)	
CS05	Fresh total biomass yield, t ha^{-1}	R	7	129.8 (3.1)	163.4 (4.3)
CS05	Fresh leaf yield, t ha^{-1}	R	1	36.6 (3.7)	
CS05	Fresh leaf yield, t ha^{-1}	R	7	16.8 (2.6)	
CS05	Fresh panicle yield, t ha^{-1}	R	2	74.4 (2.7)	
CS05	Dry leaf yield, t ha^{-1}	R	1	69.3 (3.4)	
CS05	Dry panicle yield, t ha^{-1}	R	2	74.4 (3)	
CS05	Stem NDF, g kg^{-1}	R	8	117 (3.4)	
CS05	Stem cellulose, g kg$^{-1\S}$	R	8	29 (3.9)	
CS05	Stem hemi-cellulose, g kg^{-1}	B	9	58.4 (3)	
CS05	Stem lignin, g kg^{-1}	R	8	29 (4.3)	
CS05	Stem lignin, g kg^{-1}	R	8	130.2 (2.6)	
CS05	Stem crude protein, g kg^{-1}	B	8	0 (3.4)	
CS05	Leaf cellulose, g kg^{-1}	B	3	124.5 (2.5)	
CS05	Leaf lignin, g kg^{-1}	B	4	159.4 (2.7)	
CS05	Leaf lignin, g kg^{-1}	B	3	185.6 (3.8)	
CS05	Leaf crude protein, g kg^{-1}	B	1	0 (2.8)	
CS05	Leaf crude protein, g kg^{-1}	R	1	219.1 (2.9)	229.3 (2.9)
CS05	Leaf crude protein, g kg^{-1}	R	9	127.1 (2.7)	
CS05	Stem cellulose (% structural solids)	R	8	0 (2.5)	
CS05	Stem hemi-cellulose (% structural solids)	R	6	30.7 (3)	
CS05	Stem lignin (% structural solids)	B	3	124.5 (3.9)	
CS05	Stem lignin (% structural solids)	R	8	29 (2.8)	
CS05	Stem lignin (% structural solids)	R	8	130.2 (2.5)	
CS05	Stem crude protein (% structural solids)	B	4	148.1 (2.9)	
CS05	Stem crude protein (% structural solids)	R	5	4.4 (2.9)	

Table 3.7. (Continued)

CS05	Stem crude protein (% structural solids)	B	8	74.5 (3.4)	
CS05	Lodging, %	R	7	163.4 (2.7)	
CS06	Dry total biomass yield, t ha^{-1}	B	6	30.7 (2.6)	
CS06	Dry total biomass yield, t ha^{-1}	R	9	3.5 (2.8)	
CS06	Dry leaf yield, t ha^{-1}	R	1	14.5 (3.1)	
CS06	Grain dry harvest index	B	7	30.5 (2.8)	
CS06	Stem NDF, g kg^{-1}	B	3	109.3 (3.2)	140.4 (3.1)
CS06	Stem cellulose, g kg^{-1}§	B	3	109.3 (3.1)	
CS06	Stem cellulose, g kg^{-1}§	B	7	82.8 (2.7)	
CS06	Stem hemi-cellulose, g kg^{-1}	B	3	130.1 (3.1)	
CS06	Stem hemi-cellulose (% structural solids)	B	10	132.6 (2.9)	
CS06	Stem lignin, g kg^{-1}	B	6	30.7 (2.6)	
CS06	Stem crude protein, g kg^{-1}	B	1	100.2 (2.8)	
CS06	Lodging, %	R	1	74.3 (3.3)	
CS06	Regrowth brix, °brix	R	7	82.8 (2.9)	
CS06	Regrowth brix, °brix	R	7	129.8 (3.4)	
CS06	Stem lignin (% structural solids)	R	8	130.2 (2.6)	
CS06	Stem lignin (% structural solids)	R	8	39.5 (3.2)	
CS06	Stem crude protein (% structural solids)	B	1	100.2 (3)	
CS06	Stem crude protein (% structural solids)	R	3	109.3 (2.7)	
CS06	Stem crude protein (% structural solids)	R	6	5.1 (2.7)	
CS06	Stem crude protein (% structural solids)	B	8	29 (2.7)	

Unlike our previous study on nonstructural carbohydrates, the grain parent, BTx623, did not provide alleles that consistently increased any traits.

Structural biomass yield "hotspots" for QTL co-localization appeared in similar locations as non-structural carbohydrate QTL. This result was not surprising, since both types of carbohydrates are strongly correlated with height, flowering time, and stand density/tillering, especially in the midge and rain stressed environment CS05 (Murray et al. 2008). Although linkage of two separate genes can not be ruled out for QTL co-localization between traits, it is likely that co-localization was due to the pleiotropic effects of a single gene (*e.g.*, taller plants produce more stem biomass given consistent stem diameters and stand density/tillering ability).

Across the three locations, five regions were responsible for the majority of QTL co-localizations. Height QTL on chromosomes 7 and 9, were consistent across environments, and co-localized with increased stem and total biomass QTL and decreased stem protein QTL. A delayed flowering time QTL on chromosome 6, likely photoperiod sensitive *ma1*, co-localized with increased stem, leaf and total biomass, decreased grain yield, and changes in leaf and stem composition only in CS05. A stand density/tillering QTL on chromosome 4 co-localized with increased regrowth maturity and decreased mean stem thickness and stem crude protein.

Finally, the major brix/sugar concentration QTL on chromosome 3 co-localized to regrowth brix and altered structural stem composition across all three environments (Figure 3.1). The co-localization of opposite effects between brix and structural stem composition was directly influenced by stem residual sugar concentration. To adjust for residual sugar differences in QTL mapping, the lignin, cellulose, hemi-cellulose and protein were expressed as % structural solids (calculated by dividing by the sum of these components), which controls for residual sugar. After adjustment, stem % structural cellulose and stem % structural protein still co-localized to the chromosome 3 brix QTL suggesting that some change in stem structural composition is pleiotropic with stem sugar concentration (Table 3.7).

QTL for dry measures of leaf, stem, grain and total biomass generally co-localized with fresh measures and explained similar amounts of variation. Given the higher heritability, it was surprising that QTL for harvest indices did not explain more variance compared to fresh or dry yields. For leaf harvest index, many unique QTL were detected. For panicle and grain harvest index, QTL were detected in the same location as fresh or dry measurements but were of larger magnitude.

Discussion

In this study, we evaluated progeny from a cross between an elite grain parent and a high biomass/high stem sugar parent to investigate the genetic basis of traits that might be useful for improving sorghum as a crop residue and/or dedicated biomass feedstock. We also examined genetic correlations between structural (leaf and stem cellulose) and non-structural (stem sugar and grain starch) carbohydrate yield and composition traits. To our knowledge this is the first documented study to analyze the genetic relationships among yield and composition traits of all above-ground products (sorghum grain, stem sugar and biomass). Genetic relationships among traits were identified from both trait correlation and QTL analyses. Results from correlation analyses included the effects of major genetic loci, genetic background (small effect loci and/or epistatic interactions) and genetic by environmental interactions. The QTL analyses, on the other hand, identified genetic tradeoffs in different environments only at major genetic loci (caused either by pleiotropy or genetic linkage). Across all traits, the QTL identified did not fully explain the genetic variation suggested by heritability calculations. QTL co-localization also did not fully explain the correlations between traits suggested by correlation coefficients. Our power to detect QTL was restrained by the use of a stringent statistical significance threshold (p=0.01). However, the stringent threshold increased our confidence that the QTL identified were not false positives.

Relationships Between Non-Structural and Structural Carbohydrate Yields Were Primarily Due to Height, Flowering Time, and Stand Density: In Murray et al. (2008), a few positive correlations between stem sugar and grain starch yields were identified, primarily due to genetic differences for height, flowering time and stem density/tillering. In this study, leaf and stem biomass yields were also found to be strongly correlated with height, flowering time and stand density/tillering.

Increased structural stem yield and biomass yield co-localized to the same height (chromosomes 7 and 9), flowering time (chromosome 6 and 9) and stand density/tillering (chromosome 4) QTL identified for stem sugar (Murray et al., 2008). Although few leaf yield QTL were identified, these also co-localized with flowering time (chromosome 6 and 9). Panicle yield, which is mostly non-structural grain yield, co-localized with flowering time QTL (chromosome 6). Therefore, a simplified general relationship between structural and non-structural yield is that increasing biomass increases the yield of stem sugar and slightly decreases yield of grain starch.

Stem and Leaf Carbohydrate Compositions Were Independent: Although stem, leaf, and grain yield were highly correlated because of height, flowering time, or stand density/tillering variation, this effect was not observed for structural carbohydrate composition traits. Furthermore, we found composition traits had low genetic correlation between leaf and stem tissues. In addition, there was no co-localization between leaf and stem carbohydrate composition QTL except for total structural carbohydrates (NDF) which co-localized with the major brix QTL on chromosome 3. In forage maize, Krakowsky et al. (2005 and 2006) also found that composition of leaf and stem tissues were under separate genetic control. This finding suggests that improvement of whole plant composition for biofuel production would proceed more quickly by selecting on leaf and stem tissue composition separately. Furthermore, because the stem contributes more to total biomass than leaf tissue, selection for composition alone could potentially change harvest indices.

Protein Levels in Leaf, Stem, and Grain are Also Under Separate Genetic Control: Our results showed that in non-stressed environments leaf, stem and grain protein levels were not correlated and QTL for theses traits did not co-localize. Therefore, protein composition across tissues was under separate genetic control. This finding contradicts results of Moyer et. al. (2003) who showed that among sorghum

hay and forage types crude protein levels in stems and leaves were significantly positively correlated. Because hay and forage sorghums are selected for increased total protein the latter result could either be an artifact of selection or it is possible that QTL that affect protein levels in different tissues were not evaluated because they were fixed in our experimental population.

We did find some evidence for negative correlations between protein in all tissues and carbohydrate composition and/or yield. Since there appears to be some tradeoff between protein content and carbohydrate yields, improvement of sorghum as a biofuel crop should focus on lowering protein levels in harvested tissues. As with structural carbohydrate composition, this finding suggests that breeding for lower overall protein may be made most quickly with separate selection for grain, leaf, and stem protein.

Regrowth Protects the Soil and Increases Harvestable Energy: A major concern of using crop residues for bio-feedstocks is that this practice leaves soil bare and prone to erosion and removes organic matter that could be incorporated into the soil. Unlike maize, sorghum continues to produce tillers (ratoon) after it is harvested, given sufficient water and protection from freezing temperatures. Because sorghum does regrow, vegetative material is available both for erosion control and for providing additional soil organic matter. An additional benefit is that, like sugarcane, regrowth sorghum can also be harvested as an additional source of biomass. In the CS06 location, sorghum plants were allowed to regrow after the initial cutting and the first regrowth was harvested around the time of grain maturity. Although there was large variation in regrowth, this second harvest provided a large additional source of harvestable energy and there was still sufficient time before frost for a third regrowth of tillers to protect the soil. For structural biomass feedstocks, experiments to evaluate

the efficacy of harvesting repeated cuttings of regrowth material compared to replanting must be conducted.

Energy Considerations: In the US, grain starch is currently the primary feedstock of ethanol production. Theoretical yield is 0.72 liters of ethanol per kg of starch and actual efficiency (yield) is 85% to 89% of this value for sorghum (Wu et al., 2007). Sugar has a theoretical ethanol yield of 0.68 liter per kg sugar while actual efficiency using raw sugar is about 83% (USDA 2006). Cellulose has a theoretical ethanol yield equal to starch (0.72 liters of ethanol per kg of cellulose). Cellulose conversion efficiency, however, varies widely, although values are improving as the conversion technologies continue to evolve (NREL Theoretical Ethanol Yield Calculator, 2007; Hamelinck et al., 2005). From crop physiology and energy production perspectives, therefore, 1 kg grain starch is approximately equal to 1 kg of stem sugar or 1 kg of cellulose. Figure 3.2 summarizes dry yield data in WE05 for starch, sugar and biomass and theoretical ethanol yields for the parents and selected RILS with extreme phenotypes (high brix, starch, cellulose, etc).

Our data indicated that, on average, starch comprised 63.3% of dry grain yield (53% - 69%), sugar accounted for 12.4% of juice yield (4.6% - 17%), and cellulose comprised 33% (27% - 39%) of dry leaf and 35.6% (29% - 46%) of dry stem yields. Grain is, therefore, more "energy dense" than stem juice or biomass, although there is less total energy produced from grain because of lower yields (Figure 3.2). In dry form, stem, and leaf biomass is also energy dense. Stem juice sugar is not energy dense but, unlike starch and cellulose, is immediately available for fermentation without supplementation with additional water.

Energy density is important because it directly affects the cost of transporting plant material from the field to ethanol production facilities; as energy density increases transportation cost decrease. It seems logical, therefore, that breeding

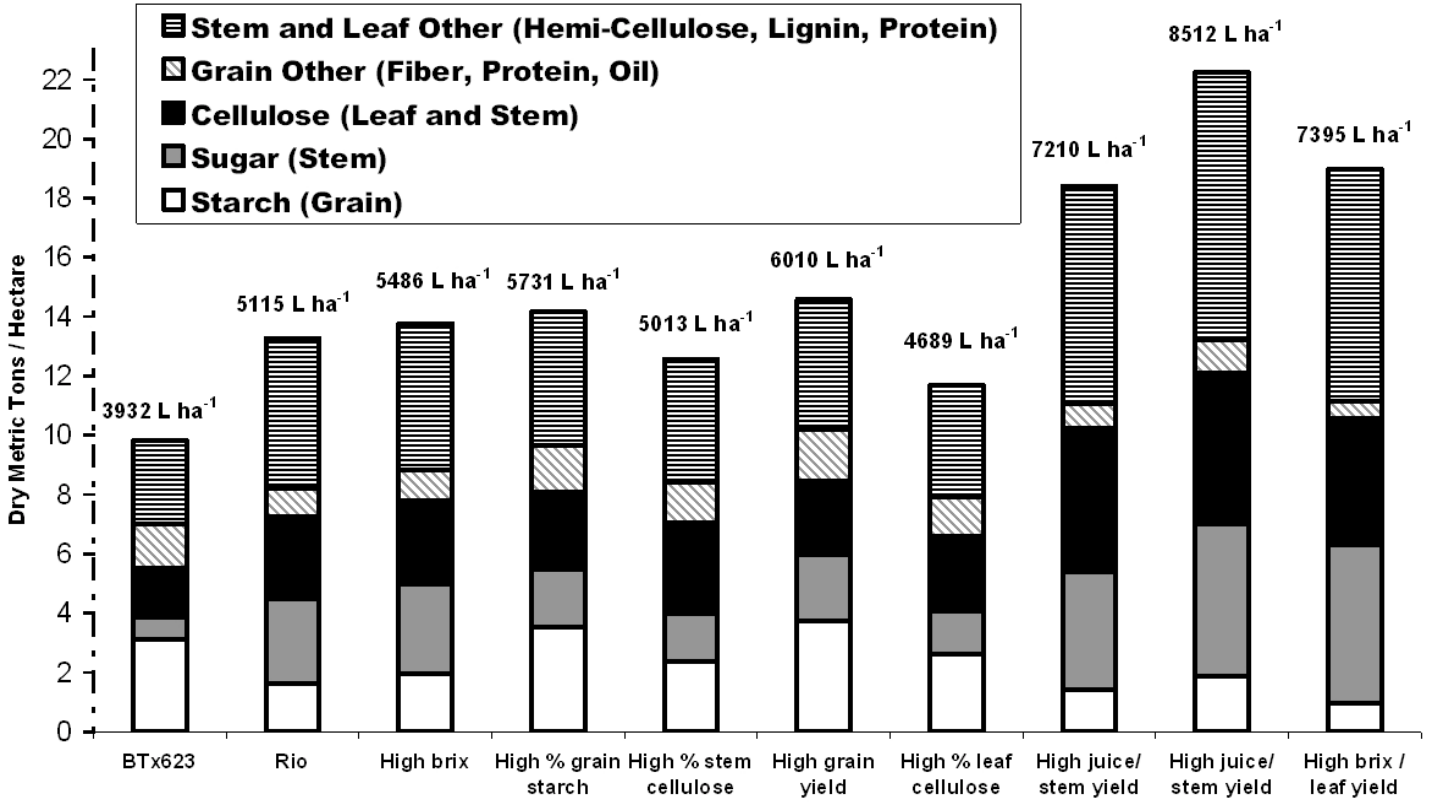


Figure 3.2. Yield, composition and theoretical ethanol for parents and eight selected RILs

Parents and seven extreme RILs for yield and composition (at bottom) are shown for the two replicate average yield of grain starch, stem sugar, leaf / stem cellulose, grain byproducts, and leaf / stem byproducts. Theoretical ethanol yield (shown at top of each bar) was calculated as 1kg starch and cellulose is equal to 0.72 Liters of ethanol (Wu et al., 2007; NREL theoretical ethanol yield calculator, 2007), and 1 kg sugar was equal to 0.68 L ethanol (Shappouri and Salassi, 2006). RILs selected on the basis of yield produce more theoretical ethanol than those selected based on composition. Data is from WE05 only.

strategies should focus on improving composition for high energy density. As shown in Figure 3.2, however, data from RILs that are energy dense (high brix, starch, cellulose) and high yielding (high stem juice, grain, leaf, and stem biomass yields) indicate that increased yields are more important than improved composition for ethanol production in this population. Based on mean yield and composition, improving only the starch, sugar and cellulose composition to the maximum observed levels would raise the amount of theoretical ethanol produced by 17%, improving the yield of grain, juice, and stem and leaf dry biomass would increase theoretical ethanol by 89%, and improving both the maximum composition and yield would increase theoretical ethanol by 124% over the population mean. This finding clearly argues for focusing first on yield increases, and then on composition (assuming that all traits can be improved simultaneously and hybrid heterosis would not affect these relationships).

Strategies for Sorghum Improvement: We suggest two ideotypes as goals for sorghum improvement for energy: a primary grain crop with residue improved for stem sugar and structural biomass composition, and a dedicated biomass crop maximizing cellulose yield. In our RIL population, yields of leaf, stem, and grain biomass contributed more variation than composition. Therefore, yield improvement should be a primary goal for breeding both residue and dedicated biomass feedstocks. With grain sorghum, increases in leaf and stem yields (total biomass) could be achieved by avoiding the height and flowering time QTL on chromosome 9 and the flowering time QTL on chromosome 6. For sweet and dedicated biofuel sorghums, selecting for increasingly tall, late flowering material with greater stand density/tillering ability appears promising, and plentiful genetic variation is available for these traits. Because height and flowering time affect so many traits, development of molecular markers at these QTL would be highly advantageous for rapid selection of desirable phenotypes (Holland, 2004).

Our results suggest that stem and leaf carbohydrate composition, as well as protein composition in non-stress environments, are under separate genetic control. These traits, therefore, should be evaluated separately in future studies to maximize improvement. Since protein is undesirable in a biomass feedstock we conclude that crop improvement should focus on lowering protein separately in each tissue harvested. The cost and labor of this approach, however, can only be justified if major QTL for height and flowering time are fixed in the experimental population.

Conclusions

Demand for agriculture feedstocks coupled with new biofuel processing technologies are creating a major shift from regarding plants as sources of food, feed and fiber to viewing whole plants as a method to capture and store energy. To our knowledge, this is the first study to simultaneously evaluate genetic variation in yield and composition of the whole plant (stem, leaf, and grain) for biofuel applications. We found that yield contributed more variation to theoretical ethanol than composition in all tissues tested. Although correlations between estimates of tissue chemical composition from forage methods (NIRS) and ethanol yield have not been established, direct chemical measurement by current laboratory methods is both impractical and prohibitively expensive for assaying large numbers of samples. Therefore, the expense of performing comprehensive chemical analyses is probably not justified before biomass yield is improved in dedicated biofuel sorghums.

Much of the current work in developing biofuel feedstocks has focused on transgenic technologies, both for improved composition and digestibility of cellulosic components (Sticklen, 2006). Presently, it is not clear what genetic diversity exists for biomass traits and future work should concentrate on evaluating a broader range of germplasm. Discussions of economic viability are premature until improved

feedstocks are available and a consensus is reached on location-dependent, feasible digestion technologies.

REFERENCES

Anderson, I.C., D.R. Buxton, and A. Hallam. 1995. Performance of annual and perennial crops for forage and biomass energy production. U.S. Dairy Forage Research Center Research Summary. [Online]. Available at http://www.dfrc.ars.usda.gov/RS95_pdfs/fp5.pdf (verified 12 Nov. 2007).

Association of Official Analytical Chemists (AOAC). 1990b. Method 930.15. In K. Helrich (ed.) Official methods of plant analysts (15th ed.). AOAC, Arlington, VA.

Association of Official Analytical Chemists (AOAC). 1990a. Method 990.03. In K. Helrich (ed.) Official methods of plant analysts (15th ed.). AOAC, Arlington, VA.

Bies, L. 2006. The biofuels explosion: Is green energy good for wildlife? Wildlife Society Bulletin 34:1203-1205.

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

Casler, M.D., and K.P. Vogel. 1999. Accomplishments and impact from breeding for increased forage nutritional value. Crop Sci. 39:12–20.

Farrell, A.E., R.J. Plevin, B.T. Turner, A.D. Jones, M. O'Hare, and D.M. Kammen. 2006. Ethanol can contribute to energy and environmental goals. Science 27:506-508.

Frederiksen, R.A., and F. Miller. 1972. Proposal for release and increase ATx622, BTx622, ATx623, BTx623, ATx624, BTx624. TAES Form 96–72: Seed Release Committee of the Texas Agricultural Experiment Station College Station.

Hallam, A., I.C. Anderson, and D.R. Buxton. 2001. Comparative economic analysis of perennial, annual, and intercrops for biomass production. Biomass Bioenerg. 21:407-424.

Hamelinck, C.N., G. van Hooijdonk, and A.P.C. Faaij. 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. Biomass Bioenerg. 28:384-410.

Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in breeding programs- challenges and opportunities. In T. Fischer et al. (ed.) New directions for a diverse planet: Proceedings for the 4th International Crop Science Congress, Brisbane, Australia 2004. [Online]. Available at http://www.cropscience.org.au/icsc2004/symposia/3/4/203_hollandjb.htm (verified 11 April 2006).

Jenkins, B.M., L.L. Baxter, T.R. Miles Jr., and T.R. Miles. 1998. Combustion properties of biomass. Fuel Processing Tech. 54:17–46

Krakowsky, M.D., M. Lee, J.G. Coors. 2005. Quantitative trait loci for cell wall components in recombinant inbred lines of maize (Zea mays L.) I: Stalk tissue. Theor. Appl. Gen. 111:337-346.

Krakowsky, M.D., M. Lee, J.G. Coors. 2006. Quantitative trait loci for cell wall components in recombinant inbred lines of maize (Zea mays L.) II: Leaf sheath tissue. Theor. Appl. Gen. 112: 717-726.

Lal, R. 2005. World crop residues production and implications of its use as a biofuel. Environ. Int. 31:575–584.

Morgan, P.W., and S.A. Finlayson. 2001. Techniques for developing new cultivars. p. 329–347. In C.W. Smith and R.A. Frederiksen (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, New York.

Moyer, J.L., J.O. Fritz, and J.J. Higgins. 2003. Relationships among forage yield and quality factors of hay-type sorghums. Crop management journal. [Online]. Available at http://www.plantmanagementnetwork.org/cm/2003.asp (verified 12 Nov. 2007).

Murray, S.C., W.L. Rooney, P.E. Klein, A. Sharma, J.E. Mullet, S.E. Mitchell, and S. Kresovich. 2008. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem and grain nonstructural carbohydrates. Crop Sci. accepted pending revision.

NREL theoretical ethanol yield calculator (2007) national renewable energy laboratory. [Online]. Available at http://www1.eere.energy.gov/biomass/ethanol_yield_calculator.html (verified 12 Nov. 2007).

Quinby JR, Karper RE (1954) Inheritance of height in sorghum. Agron J. 46:211–216

R Development Core Team. 2005. R: A language and environment for statistical computing. Vienna, Austria. [Online]. Available at http://www.R-project.org (verified 12 Nov. 2007).

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

Shappouri, H. and M. Salassi. 2006. The economic feasibility of ethanol production from sugar in the United States. USDA, Washington DC. [Online]. Available at http://www.usda.gov/oce/EthanolSugarFeasibilityReport3.pdf (verified 14 Nov. 2007).

Somerville, C. 2007. Biofuels. Curr. Biol. 17: R115-R119.

Sticklen, M. 2006. Plant genetic engineering to improve biomass characteristics for biofuels. Curr. Opin. Biotech. 17:315-319.

Theander, O., E. Westerlund. 1993. Quantitative analysis of cell wall components. p. 33–71. In H.G. Jung et al. (ed.) Forage cell wall structure and digestibility. ASA,CSSA, and SSSA, Madison, WI.

Wolfinger, R.D., G. Gibson, E. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R.S. Paules. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Compu. Biol. 8: 625-637.

Wu, X., R. Zhao, S.R. Bean, P.A. Seib, J.S. McLaren, R.L. Madl, M. Tuinstra, M.C. Lenz, and D. Wang. 2007. Factors impacting ethanol production from grain sorghum in the dry-grind process. Cereal Chem. 84:130-136.

USDOE. 2006. Breaking the biological barriers to cellulosic ethanol: A joint research agenda. DOE/SC 0095. USDOE Office of Science and Office of Energy Efficiency and Renewable Energy, Washington, DC. [Online]. Available at http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml (verified 14 Nov. 2007).

CHAPTER 4

SWEET SORGHUM GENETIC DIVERSITY AND ASSOCIATION MAPPING FOR BRIX AND HEIGHT

Abstract

Sweet sorghums [*Sorghum bicolor* (L.) Moench] have been selected to accumulate high levels of edible sugars in the stem, similar to a close relative, sugarcane. Sweet sorghums are tall and produce high biomass in addition to sugar. Little has been documented about the genetic relationships and diversity within sweet sorghums and how sweet sorghums relate to grain racial types. In this study, a diversity panel of 125 sorghums, mostly sweet, were genotyped with 48SSRs and 384SNPs. Using PCoA and structure analyses we identified three main populations of sweet sorghum: historical and modern syrup, modern sugar/energy types, and amber types. These populations were consistent with phenotype and historical origins. Using SSR markers shared with an available large grain sorghum germplasm panel we found these three sweet types clustered with kafir/bicolor, caudatum and bicolor types respectively. Using the information on population structure and relatedness, whole genome association mapping was performed for height and stem sugar (brix) traits. Three significant associations for height were detected. Two of these, on chromosomes 9 and 6, support published QTL studies. Only one significant association for brix, on chromosome 1, 12kb from a glucose-6-phosphate isomerase homologue, was detected. Previously published QTL for brix were not detected, likely due to low heritability, epistasis, and potentially different QTL for stem sugar in the three populations.

Introduction

Sweet sorghums belong to the same domesticated species [*Sorghum bicolor* (L.) Moench] as grain, forage and broomcorn sorghums but have been selected to accumulate high levels of sucrose in the parenchyma of juicy stems (Harlan and deWet, 1972; Vietor and Miller, 1990). Sweet sorghum sugar accumulation is similar to that in sugarcane, a close relative, though studies on enzymatic control and carbon transport suggest that the mechanism of accumulation is different (Lingle et al., 1987; Tarpley and Vietor, 2007). The stems of sweet sorghum are desired for food-grade syrup (juice is pressed out and boiled) but also for fresh chewing and alcohol production in Brazil and India (House et al. 2000).

Recent demand for biofuel, in light of perceived Brazilian success with sugarcane, has caused a re-evaluation of sweet sorghums as a source of energy (Rooney et al., 2007; Vermerris et al., 2007). Up to 13.2 t/ha of total sugars, equivalent to 7,682 liters of ethanol per hectare, can be produced by sweet sorghum under favorable conditions (Jackson et. al. 1980). Sweet sorghum and other sugar crops have been researched for biofuel production in the US for over 30 years. Primary research, development, and breeding began in the late 1970's when the high cost of oil spurred interest in alternative energy sources (Lipinsky et al., 1977). These investigations were ended by 1987 when petroleum costs had decreased (DOE-OSTI, 2008).

Sweet sorghum breeding and research: Sweet sorghums, also called sorgos, were originally brought to the US as landraces from China (*cv.* 'Chinese Amber') and Africa (*cvs.* 'Orange', 'Sumac'/ 'Redtop', 'Goosneck' / 'Texas Seeded Ribbon Cane', 'Honey', ' White African', and others) via France in the 1850's for producing syrup (sirup) and forage (Winberry, 1980; Smith and Frederiksen, 2000). Many of these original sweet sorghum landraces continued to be selected by farmers regionally in the US and were renamed. Other cultivars were introduced later: 'Collier' from South Africa, 'McLean' from Australia, and others with unknown origin such as 'Folger', 'Coleman',

'Sugar-Drip', and 'Rex', referenced as early as 1923 (Sherwood, 1923; Smith and Frederikson, 2000).

Almost all sweet sorghum cultivars improved by modern breeding were improved at the USDA-sponsored U.S. Sugar Crops Field Station in Meridian, MS, from the 1940s until it closed in 1983. The Meridian station used landraces for plant improvement and released improved syrup lines. A few lines were also selected for sugar production and energy (biomass tonnage) in collaboration with others across the US, notably Texas and Georgia. Of the syrup lines bred and released by the Meridian station, release notes suggest primary improvement was focused on improving disease resistance. Disease can alter sorghum juice, reducing the desirability of syrup and contributing to lodging. Besides disease resistance, other selected traits include high brix (very few report stem sugar), juicy stalks, high yields, stalk erectness, and good quality syrup.

The Meridian, MS, station additionally curated a 'sweet sorghum world germplasm collection'. When it closed, many of these materials were transferred to the USDA sorghum collection in Griffin, GA (Freeman, 1975; USDA-ARS, 2008). Many accessions from this collection, used in later breeding, were obtained in a 1945 collecting trip by Carl O. Grassl around the African center of sorghum domestication (Freeman, 1975). Six of these African landraces, specifically MN960, MN1048, MN1054, MN1056, MN1060, and MN1500 were used in the pedigrees of many US released improved sweet sorghum lines (Table 1). This suggests that there may be a narrow genetic base for the sweet sorghums resulting in close genetic relationships between cultivars. If the genetic base is too narrow there may be difficulty in selecting from this material to develop energy types.

Although published pedigree information is available for some of the more recent sweet sorghum lines, the relationships with historic sweet lines and grain sorghums are poorly understood. A few studies (Anas and Yoshida 2004, Casa et al. 2008) investigated grain sorghum germplasm panels that included some sweet sorghums with genetic

markers. Further work by Seetharama et al. (1987) and Ritter et al. (2007) suggested that sweet sorghums are of polyphyletic origin, with relatives among kafir, caudatum and other grain sorghum types.

Presently, no quantitative objective criteria exist, such as a molecular marker or sugar concentration level, to differentiate sweet sorghums from grain sorghums. There are multiple generalized phenotypic differences however: sweet sorghums are always tall, have high biomass and juicy stem (as opposed to dry stem controlled by a major gene), and most importantly have high stem sugar concentrations. Stem sugar concentration may be quantitatively measured by HPLC or by brix, a measurement of soluble solids which in sorghums are mostly sucrose. Stem sugar is much easier to measure by brix than by HPLC and was the criterion used for most sweet sorghum breeding. Additionally, Murray et al. (2008a) found heritability of brix to be higher than that of sugar content measured by HPLC. This difference was due to spoilage incurred during the lengthier processing required for HPLC. Ritter (2007) found sugar content heritability to be equal using the two methods.

Stem sugar concentration inheritance is not simple; environment, genetic x environment interaction, and the genetic background (epistasis) all play a role. Within mapping populations, few QTL have been identified and they explain little variation given the moderate heritability of the trait (Schlehuber, 1949; Clark, 1981; Natoli et al., 2002; Bian et al., 2006; Ritter, 2007; Murray et al., 2008a).In two different populations Natoli et al. (2002) and Murray et al. (2008a) both identified the strongest QTL for stem sugar on chromosome 3. We chose to follow up this QTL as a candidate for association mapping in a diverse panel of sorghums.

Association mapping: Association mapping uses diverse material to associate genetic markers with a phenotype of interest, taking advantage of historical recombination not available in linkage populations. Association mapping has been used to identify genes of interest in many plant species with varying degrees of success

(Wilson, 2004; Aranzana, 2005; Breseghello and Sorrells, 2006). In sorghum, a diverse grain sorghum germplasm panel for association mapping was previously reported by Casa et al. (2008). However, only eight of the 356 lines could be considered 'sweet sorghum' types. Though there was likely variation for brix, the panel was mostly dwarf grain sorghum uncharacteristic of tall and high-biomass sorghums of interest. We therefore assembled a panel that represents historically important sweet sorghum lines, important sweet landrace progenitors, and lines that would serve as non-sweet controls.

Questions we were trying to address: In this study we were interested in addressing three questions. 1) What are the genetic relationships among sweet sorghums in the United States? 2) What are the genetic relationships among sweet and grain sorghums across racial classifications? 3) Can we confirm the major QTL for stem sugar (brix), or any of the QTL for height identified in Murray et al. (2008a), using association mapping?

Materials and Methods

Plant material and phenotypic analysis: Two replicates of 125 diverse accessions were planted in College Station, Texas in 2006 (CS06) and 2007 (CS07) and one replicate was planted in Ithaca, NY in 2007 (ITH07). These accessions were primarily historical and modern sweet sorghum cultivars, though grain, and forage sorghums were also included (Table 1). These accessions will be referred to as the 'sweet sorghum panel' from here on. Literature searches and the GRIN database (USDA-ARS 2008) were used to identify lines as amber, historical sweet, modern sweet, modern sugar and energy, MN landraces (brought to Meridian, MS from Africa by C.O. Grassl), or grain types. We use the term 'modern' to denote improved lines that have published pedigree information. Seed was obtained from a variety of sources for CS06 (Table1), and seed bulked from self pollinated plants was planted for CS07 and ITH07. In CS06 and CS07, 3 meter rows with 76 cm spacing (~160 000 plants ha^{-1}) were planted in a

randomized complete block design. In ITH07 30 seeds were hand planted in 1.5 meter rows with 76 cm spacing, and hand weeded.

Plants were harvested when most accessions were in the soft-dough to hard-dough stage but there was wide diversity for maturity. In each location, one meter per row was harvested by cutting within three centimeters of the soil. Stems were separated from panicles and leaf tissue. Stem juice was extracted using a three roller mill. Brix was measured using a handheld refractometer. Measurements were collected on one meter of row in CS06 and CS07. Measurements were collected from 3 random plants in ITH07. CS06 juice samples were shipped to Ithaca, NY for further analysis by HPLC. HPLC was performed according to Murray et al. (2008a). No HPLC analysis was performed for CS07 or ITH07. Plant height was measured in the field for all three locations.

Genetic Analysis: Leaf tissue was collected from plants grown at CS06 location. DNA was extracted from pooled tissue of five or more plants using a standard CTAB protocol (Doyle and Doyle, 1987). Forty-six polymorphic SSRs, used in the diverse association panel of Casa et al. (2008), were evaluated using the same equipment and published methods (Xcup19, Xtxp065, Xtxp287 were not included). 384 SNP genotypes were collected using an Illumina Goldengate assay (Fan et al. 2006) at Cornell's Life Sciences Core Laboratories Center (Ithaca, NY) using recommended procedures. These 384 SNP assays were developed from SNPs discovered in previously published (Hamblin et al. 2004, 2005, 2006, 2007a) and unpublished (Murray, this study; Salas et al., in preparation) resequencing studies.

To identify candidate genes for brix, the major QTL for brix in a cross between a grain sorghum and a sweet sorghum (Murray et al. 2008a) was located on the sorghum genome sequence (Phytozome, 2008) using BLAST analysis with sequenced based markers (Menz et al. 2002; Feltus et al., 2007).

Table 4.1. Sweet panel line names and associated information

Name[†]	Full name	Source[‡]	Source 2[§]	Type[¶]	Parentage or Place of Origin[#]	Reference
7035S	7035S	U	PI 552851	?		
Atlas1	Atlas	T	ASA.61	HS		
Atlas2	Atlas Sorgo	T		HS		
Axtel	Axtel	T		HS		
Bailey	Bailey	K	NSL 187557	MS	Wiley, Tracy	Duncan, et al. 1984
Brandes	Brandes	T	NSL 29336	MS	Collier 706-C, MN1500	Colman and Broadhead,1968
Brawley1	Brawley	U	PI 533998	MS	Rex, White-seeded Collier	USDA, 1958
Brawley2	Brawley	T		MS		
CAmber1	Chinese Amber	U	PI 22913	A		Smith and Frederiksen, 2000
CAmber2	Chinese Amber	U	PI 248298	A		Smith and Frederiksen, 2000
CAmber3	Chinese Amber	T	ASA.45	A		Smith and Frederiksen, 2000
Colier1	Collier	U	PI 19770	HS		Smith and Frederiksen, 2000
Colier2	Collier	T	ASA.64	HS		Smith and Frederiksen, 2000
Colier7	Collier 706C	U	PI 563032	HS		Smith and Frederiksen, 2000
Colier3	Collier Meridian	T		HS		Smith and Frederiksen, 2000
Colier4	Collier	T	PI 19770	HS		Smith and Frederiksen, 2000
Colman1	Colman	T	ASA.52	HS		Sherwood, 1923
Colman2	Colman (Young Meridian)	T		HS		Sherwood, 1923
Cowley	Cowley	T		MS	Collier 706-C, MN1054, MN960, MN 1056, MN 1054, Early Folgers , Hodo, MN 1060	Kresovich et al., 1985
CnAtlas	Cunningham Atlas	T		HS		
DkAmber	Dakota Amber	T	ASA.48	A		
Dale	Dale	K	NSL 74333	MS	Tracy , MN960	Broadhead et al. 1970
Danton	Danton	T	ASA.65	HS		
Della1	Della	K		MS	Atx622, Dale	Harrison and Miller, 1993
Della2	Della	T		MS	Atx622, Dale	Harrison and Miller, 1993
Della3	Della	U	PI 566819	MS	Atx622, Dale	Harrison and Miller, 1993
EFolger	Early Folger	T		HS		

Table 4.1. (Continued)

EllisSo	Ellis Sorgo	T		HS	Leoti, Atlas	Karper, 1949
Folger	Folger	T	ASA.59	HS		
Fremont	Freemont Sorgo	T	Akron, Co	HS		
GaBlueR	Georgia Blue Ribbon	T		HS		Freeman, et al., 1973.
HoneyS1	Honey Sorghum	U		A		Freeman et al., 1986.
HoneyS2	Honey Sorghum	T	PI 181080	A	aka MN2931	
Iceberg	Iceberg Sorgo	T		HS	Orange type	
KColier	Kansas Collier	T	Anthony, Ks	HS		Smith and Frederiksen, 2000
KOrange	Kansas Orange	T	ASA.51	HHS		Smith and Frederiksen, 2000
Keller1	Keller	K		MS	MER 50-1, Rio	Broadhead et al. 1979
Keller2	Keller	T		MS	MER 50-1, Rio	Broadhead, et al., 1981
Leoi	Leoi	U	PI 154995	HS		
Leoti	Leoti	T	ASA.58	HS		
M81E	M81E	K	NSL 174431	MS	Brawley , Brawley, Rio	Broadhead et al., 1981
McLeanS	McLean (Starchy)	T		HS		
McLeanW	McLean (Waxy)	T	ASA.62	HS		
MnAmber	Minnesota Amber	T	ASA.46	A		
Mn1054	MN 1054	U	PI 152965	LMN	Sudan.	Freeman, 1979
Mn1056	MN 1056	U	PI 152967	LMN	Sudan.	Freeman, 1979
Mn1060	MN 1060	U	PI 152971	LMN	Sudan.	Freeman, 1979
Mn1500	MN 1500	U	PI 154844	LMN	Uganda -aka Grassl	Kresovich et al., 1988b
Mn2812	MN 2812	U	PI 167093	LMN	Egypt / Turkey	
Mn291	MN 291	U	Grif 14968	LMN	Extra Early Sumac	
Mn3046	MN 3046	U	PI 195754	LMN	China.	
Mn3083	MN 3083	U	PI 196586	LMN	India / Tiawan	
Mn410	MN 410	U	PI 145619	LMN	S. africa	
Mn4125	MN 4125	U	PI 250583	LMN	egypt	
Mn4466	MN 4466	U	PI 255744	LMN	turkey, Taslik village	
Mn822	MN 822	U	PI 152694	LMN	Kordofan, Sudan.	
Mn856	MN 856	U	PI 152728	LMN	sudan	
Mn960	MN 960	U	PI 534165	LMN	Sudan.	Freeman, 1979
N100	N100	T	PI535785	MS	Waconia, Wray	Gorz et al., 1990

Table 4.1. (Continued)

N108	N108	T	PI535793	MS	Saccharum Sorgo	Gorz et al., 1990
N109	N109	T	PI535794	MS	White Collier, Grain Sorghum Line	Gorz et al., 1990
N110	N110	T	PI535795	MS	Red X	Gorz et al., 1990
N111	N111	T	PI535796	MS	Waconia	Gorz et al., 1990
N98	N98	T	PI535783	MS	Rio, Waconia, Fremont, AN39, N4692	Gorz et al., 1990
N99	N99	T	PI535784	MS	Fremont,Thesis	Gorz et al., 1990
Orange1	Orange	U	PI 2363	HHS		Smith and Frederiksen, 2000
Orange2	Orange	U	PI 533902	HHS	aka MN 604	Smith and Frederiksen, 2000
Orange3	Orange	T	ASA.50	HHS		Smith and Frederiksen, 2000
PI52606	PI52606	K	PI52606	LMN		
P526905	PI526905	K	PI526905	L - zimB		
P527045	PI527045	K	PI527045	L - zimB		
P550604	PI550604	K	PI550604	?		
Ranchr1	Rancher 3	T	Brookings, SD	A		Karper, 1949
Ranchr2	Rancher 3	T	ASA.93	A		Karper, 1949
RedAmbr	Red Amber	T	ASA.49	A		
RedTopT	Red Top Tennesse	T		HS		Winberry, 1980
Rex	Rex	U	PI 534163	HS		Sherwood, 1923
Rio1	Rio	T		MS	Rex, MN 1048	Broadhead, 1972
Rio2	Rio	T		MS	Rex, MN 1048	Colman et al. 1965
RxOrng1	Rox Orange	K		HHS		
RxOrng2	Rox Orange	T		HHS		
WhitMam	White Mammoth	T		G		
Saccaln	Saccaline	T		HS		Vinall et al. 1936
Sapling	Sapling	T	ASA.55	HS		Vinall et al. 1936
Simon	Simon	K		HS		
Smith	Smith	U	PI 511355	MS	MN4004 (Grif 16302), MN 2754, Wiley, MN 48, MN 1056, others	Kresovich and Broadhead 1988a
Sorgras	Sorgrass	U	PI 563222	?		
SucreDm	Sucre Drome	U	PI 197542	LMN		
SgrDrp1	Sugar Drip	U	PI 586435	HS		Freeman et al., 1986.

Table 4.1. (Continued)

SgrDrp2	Sugar Drip	U	PI 146890	HS		Freeman et al., 1986.
SgrDrp3	Sugar Drip	K		HS		Freeman, et al. 1986
SgrDrp4	Sugar Drip	T		HS		Freeman et al., 1986
SgrDrp5	Sugar Drip	T	Oklahoma A&M	HS		Freeman et al., 1986
SgrDrp6	Sugar Drip	T	Oklahoma A&M	HS		Freeman et al., 1986
Sumac1	Sumac	U	PI 63715	HHS		Smith and Frederiksen, 2000
Sumac2	Sumac	U	PI 35038	HHS		Smith and Frederiksen, 2000
Sumac3	Sumac	U	PI 534120	HHS		Smith and Frederiksen, 2000
TxDblSw	Texas Double Sweet	K		HS		
Top76	Top 76-6	K	PI 583832	MS	Brandes, Collier 706-C, MN 1500, MN 1056	Day et al. 1995
Tracy	Tracy	T	NSL 4029	MS	White African, Sumac	Stokes et al. 1953
Umbrela	Umbrella	K		HS		
WcAmber	Waconia Amber	T	ASA.47	A		Smith and Frederiksen, 2000
WxAtlas	Waxy Atlas	T		HS		
WhtAfr1	White African	U	PI 52606	G		
WhtAfr2	White African	T	ASA.60	G		
WhtAfr3	White African	T	Oklahoma A&M	G		
WileyRL	Wiley R Line	K		HS		Stokes et al. 1956
WileySo	Wiley Sorgo	T		MS	Collier, MN 822, MN 2046	Coleman et al., 1956
Wiliams	Williams Sorgo	T	Ky. Certified	MS		Freeman et al. 1973
Wray	Wray	T		MS	Brawley, Rio, MN 856	Broadhead et al. 1978
BTx623	B.Tx623	T		G		
BTx635	B.Tx635	T		G		
BTx631	B.Tx631	T		G		
BTx642	B.Tx642	T		G		
P850029	P850029	T		G		
Macia	Macia	T		G		
Sureno	Sureno	T		G		
ATx623	A.Tx623	T		G		
EA1083	SC599	T	sc599	G		
EA1074	Rio 9188	T	Rio 9188	G		

Table 4.1. (Continued)

EA1084	SC599-6-9188	T	PI 593916			
Forag41		T				
Forag73		T			TX631, Tx2910	
Ramada	Ramada	U	NSL 107377	MS	MER 45-45', MN 1056, MN 1054, MN1060	Freeman et al. 1974
Sart	Sart	U	NSL 91616	MS	Variety from Sudan	Stokes et al. 1951

† Abbreviated name used in later tables and figures.

‡ K: University of Kentucky; T: Texas A&M University; U: USDA/ARS.

§ USDA PI number or additional information to distinguish accessions.

¶ A: amber; G: grain; HHS: historical sweet 1850's HS: historical sweet by 1923; MS: modern sweet; ?: unknown or diverse (exp. forage)

If known, parent lines for modern cultivars with pedigrees, place of origin for collected landrace material. Additional information in reference.

Because the sorghum genome was not annotated, over 100 starch and sucrose metabolism enzymes (Kanehisa et al., 2006) and sugar transport candidate genes from maize, sugarcane, tomato and rice (NCBI, 2008) were also placed on the sorghum genome using BLAST. New SSRs were identified from Phytozome contig sequence using the program Tandem Repeats Finder (Benson, 1999). Primer 3 (Rozen and Skaletsky, 2000) was used to design all primer sequences. Amplicons were sequenced after pretreatment with Exonuclease I and shrimp alkaline phosphatase. All sequencing was performed on sweet sorghum cultivar 'Rio' at Cornell University's Bioresource Center (Ithaca, NY) using a 3730 capillary sequencer. Trace files were investigated for polymorphisms between Rio and grain sorghum 'BTx623' in Sequencher 4.0 (Gene Codes Corp., Ann Arbor, MI).

Genetic distance and principal coordinate analysis: The program PowerMarker version 3.0 (Liu and Muse, 2005) was used to evaluate Fst (Wright, 1965) and create genetic distance matrices (Nei, 1972). Distance matrices were double-centered, used to obtain eigenvectors and then eigenvectors were plotted in NTSYS-pc Version 2.02 (Rohlf, F. J., 1990).

To compare sweet sorghums with the larger sorghum panel of Casa et al. (2008), Nei's 1972 genetic distance matrix was created in PowerMarker using the polymorphic SSRs shared by all accessions in both studies. Eigenvectors were obtained implementing the cmdscale function (eig = TRUE) and then plotted using R (R Development Core Team, 2005). R cmdscale was used rather than NTSYS-pc for this analysis because the data set was so large. Using smaller test data sets, the two Principal Components analyses (PCoA) gave identical results.

Population structure, relatedness, and association mapping: To minimize false positives in association mapping it is important to control for population structure and relatedness. Three programs were used to assign lines membership in the estimated number of populations: Structure, version 2.1 (Pritchard et al., 2000), InStruct (Gao et al., 2007), and NTSYS-pc. Because population structure estimates assume unlinked markers,

SNP assays from the same physical locus were condensed into 208 haplotypic loci. Phase ambiguities were called as missing alleles and loci with more than 20% missing alleles were eliminated. Brix candidate gene markers on chromosome 3 were excluded so including SSRs a total of 241 markers were used. In both Structure and InStruct, five independent runs having 5×10^5 burn-in and sampling iterations were conducted allowing *k* (number of populations) to vary between 1 and 15. For Structure, the ancestry model allowed for population admixture and correlated allele frequencies. For Instruct, population structure and individual selfing rates were inferred. Optimal *k* was identified using the marginal improvements in estimated logarithm of the likelihood of the data, the assignment of individuals to greater than 50% of a population, and on consistency of the five independent runs. *k* was additionally inferred using the DIC criterion in InStruct. Once *k* had been determined for both Structure and InStruct, a run of 5×10^6 burn-in and sampling iterations were used. PCoA eigenvectors from haplotypes were also used as population assignments.

Using the package SPAGeDi 1.2 (Hardy and Vekemans, 2002), a kinship coefficient estimation matrix was created according to J. Nason (described in Loiselle et al. 1995). Association mapping was performed using the GLM and MLM procedure in TASSLE (Bradbury et al., 2007). Six *Q* (population structure) matrices, with different numbers of populations, were separately used for association of brix and height. Brix was additionally tested using MLM in combination with the SPAGeDi created *K* matrix (relatedness / kinship). As criteria for the best *Q* matrix, we used highest model R^2, and the lowest number of positive tests to minimize Type I error. Positive tests were reported using a significance threshold of $p < 1.3 \times 10^{-4}$, based on a Bonferonni correction of .05 divided by 369 tests.

Results

Genetic analysis: One SSR, XCUp55, was not polymorphic in the sweet sorghum panel and was excluded from further analysis, resulting in 45 SSRs shared with Casa et al. (2008). Two additional SSRs, Txp120 (Menz et al. 2002) and a new SSR, discussed later were successfully added. Of the 384 Illumina SNP assays, 329 were successful, and 322 were polymorphic in the sweet sorghum panel. Seventy-seven of the 125 lines were heterozgous at one or more markers. The two lines known to be F1 forage hybrids segregated at the most markers, 41% (Forage 73) and 37% (Forage 41). MN landraces as a group averaged 22% heterozygous markers with only MN960 having no heterozygous markers and Mn1054 having the most (37%). SNP assay results suggested this was exacerbated by pooling tissue from lines, as landraces are often heterogenous.

Lines in the sweet sorghum panel with identical names but different seed sources all had at least one genetic polymorphism (Table 4.2). Of the loci that differed within Sugar Drip almost every possible combination of allele sharing across the six was observed. A few lines had very different names but identical genotypes. N110 and Sugardrip 4 were found to be exactly identical except for one locus with missing data. Rox Orange 2, Saccaline, and Sapling were also genetically identical. The phenotypes within these lines were very similar, so it appears likely the seed came from the same source for CS06 planting.

Table 4.2. Polymorphism between lines with shared names

Cultivar	Accessions	Shared polymorphism of 369 markers
Rio	2	359
Della	3	286
White African	3	282
Chinese Amber	3	194
Sumac	3	183
Orange	3	150
Sugardrip	6	157

PCoA relatedness: To identify accessions for biofuel improvement, it is useful to understand the relationships within the sweet sorghums and between sweet sorghum and grain sorghum's racial types. Nei's (1972) pairwise genetic distance was calculated between lines but was difficult to visually interpret. Genetic relationships were most easily visually interpreted using PCoA (Figure 4.1). The first 12 PCoA eigenvectors explained 35.7%, 21.4%, 7.2%, 6.3%, 5.3%, 4.4%, 4.3%, 3.6%, 3.2%, 3.1%, 2.6%, and 2.4% of the variation. Using the first two principal components with the full SSR and SNP data set, three groupings were observed. These same clusters were observed also when using only SNPs or only SSRs, though a few individuals did shift groups. Plotting additional eigenvectors produced no clear relationships of interest (data not shown).

To objectively assess sweet sorghums genetic relatedness to sorghum racial groups, shared SSRs were used to compare the sweet sorghum panel to Casa et al.'s (2008) pure racial group panel (138 lines - Figure 4.2A), and full diversity panel (382 lines – Figure 4.2B). The sweet sorghum historical and modern syrup group (Figure 4.1) appeared most similar to kafir and to a lesser extent to bicolor. The modern sugar and energy sweet sorghum group appeared most similar to caudatum and possibly guinea types. The amber sweet sorghum group looked most similar to bicolor racial types but is more divergent than most of the material in the Casa et al. (2008) panel. Although the Casa et al. (2008) panel had three times as many lines, the sweet panel appeared to have nearly as much diversity. The sweet panel was comparatively most deficient in accessions to the right side of figure 4.2B, especially the durra race.

Candidate gene identification and sequencing: The primary brix QTL identified in a cross between Rio and BTx623 (Murray et al. 2008a) was localized to a 15Mb sorghum super contig (Phytozome, 2008). Maize shrunken2 (Hamblin et al. 2007a) and a rice hypothetical monosaccharide transporter (NM_001053738 - NCBI, 2008) were the only sugar metabolism genes found to align to this Phytozome contig.

Figure 4.1. PCoA plot of sweet sorghum panel genetic similarity

Nei's genetic distance was calculated from 47 SSRs and 318 SNPs. Three groups were observed, a tight cluster of historical and modern syrup cultivars, modern sugar and energy sorghums with MN landraces, and amber types, which were the most diverse. Grain sorghums did not cluster in any one group.

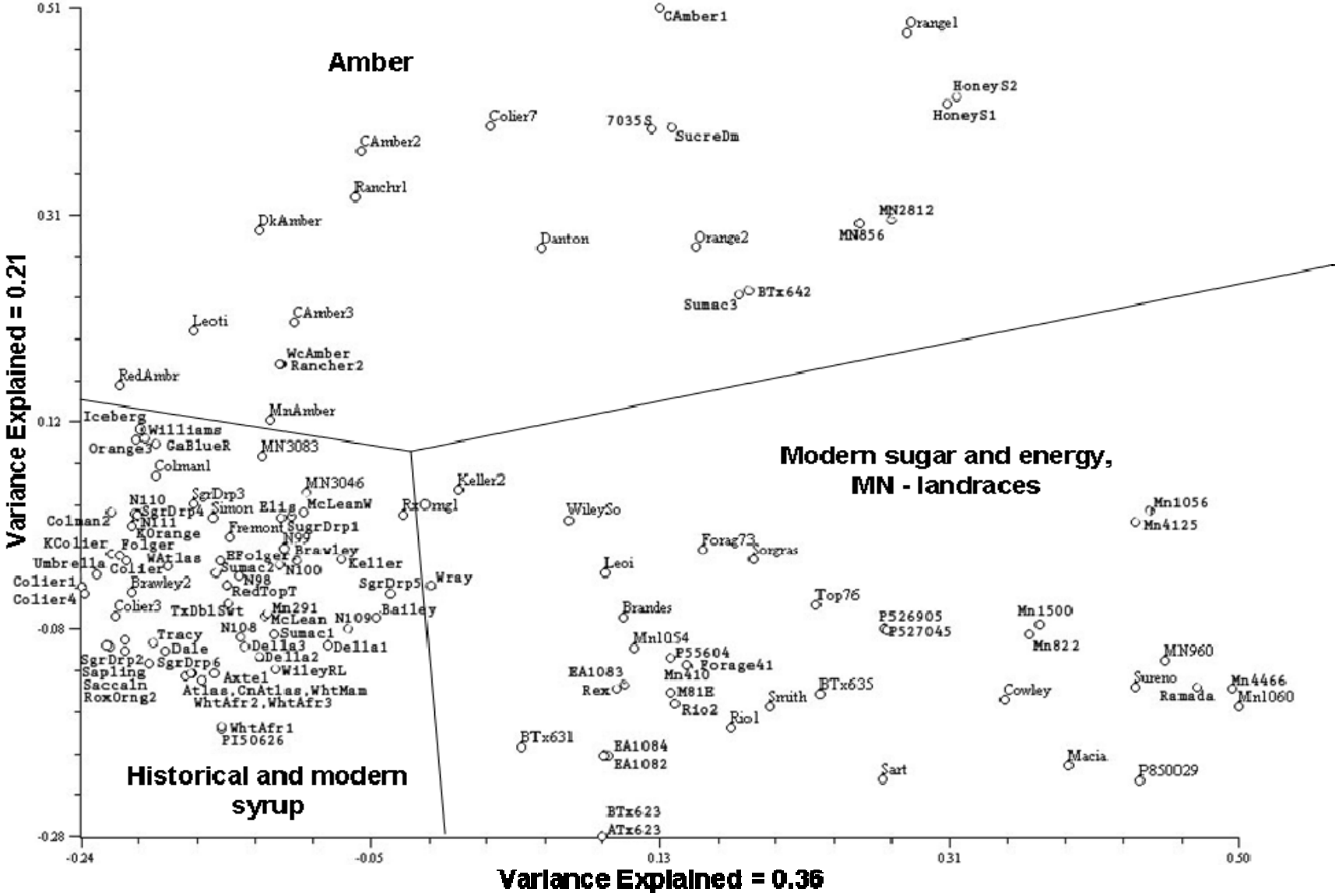

Amber
Modern sugar and energy, MN - landraces
Historical and modern syrup
Variance Explained = 0.21
Variance Explained = 0.36
0.51
0.31
0.12
-0.08
-0.28
-0.24
-0.05
0.13
0.31
0.50
CAmber1
Orange1
HoneyS2
HoneyS1
Colier7
7035S
SucreDm
CAmber2
Ranchrl
DkAmber
Danton
Orange2
MN2812
MN856
Sumac3
BTx642
Leoti
CAmber3
WcAmber
Rancher2
RedAmbr
MnAmber
Iceberg
Williams
Orange3
GaBlueR
MN3083
Colmanl
MN3046
Keller2
SgrDrp3
N110
SgrDrp4
Simon
Elis
McLeanW
RxOmgl
WileySo
Mn1056
Mn4125
Colman2
N111
Fremont
SugrDrp1
KOrange
KColier
Folger
N99
Brawley
Forag73
Sorgras
Umbrella
WAtlas
EFolger
Sumac2
N100
Keller
Leoi
Colieri
Brawley2
N98
RedTopT
SgrDrp5
Wray
Colier4
Colier3
TxDb1Swt
Mn291
McLean
N109
Bailey
Top76
Brandes
P526905
P527045
Mn1500
N108
Sumac1
Tracy
Dale
Della3
Della1
Mn1054
Mn822
MN960
SgrDrp2
SgrDrp6
Della2
P55604
Forage41
Sapling
WileyRL
Axtel
EA1083
Mn410
Sureno
Mn4466
Saccaln
Atlas,CnAtlas,WhtMam
Rex
M81E
BTx635
Cowley
Ramada
Mn1060
RoxOrng2
WhtAfr2,WhtAfr3
Smith
Rio2
Rio1
WhtAfr1
PI50626
BTx631
EA1084
EA1082
Macia
Sart
P850029
BTx623
ATx623

Figure 4.2. PCoA plot of the sweet sorghum panel combined with larger grain sorghum panel of Casa et al. (2008)

Nei's genetic distance was calculated using the 45 SSRs shared by both germplasm panels. Colored letters represent classification of Casa et al. (2008) accessions, legend in upper left. Black letters represent classification of sweet sorghum panel, legend in lower left. **A)** Comparing the sweet sorghum panel to 138 lines of the larger panel classified purely as the primary five races, bicolor (15), caudatum (73), durra (25), guinea (18), and kafir (7). Lines drawn to divide groups within the sweet panel contain roughly the same accessions as in Figure 4.1. **B)** Identical to 4.2A but includes all 382 accessions from Casa et al. (2008) including mixed race types, diverse landraces, and elite breeding lines.
.

Figure 4.2A

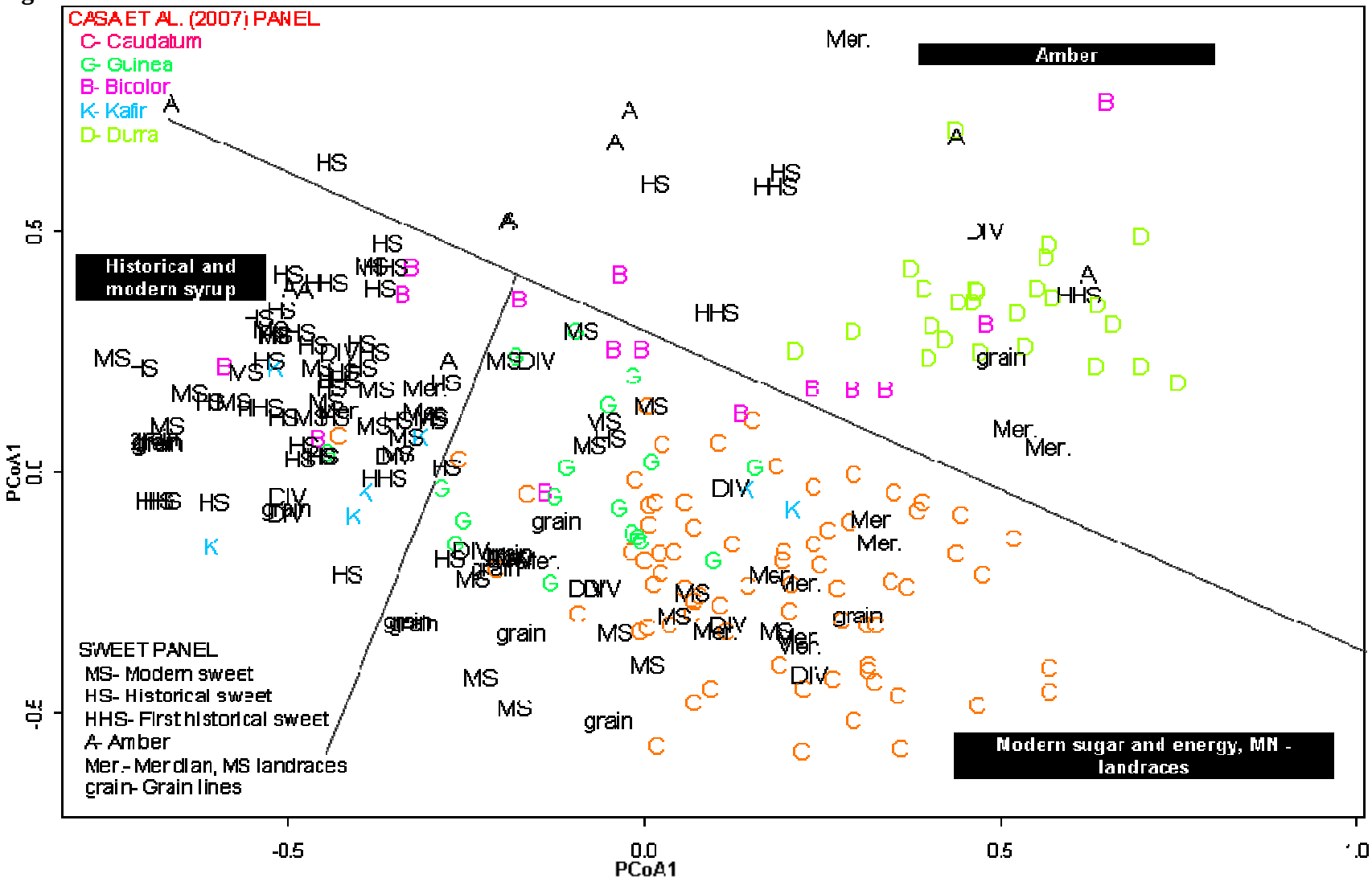
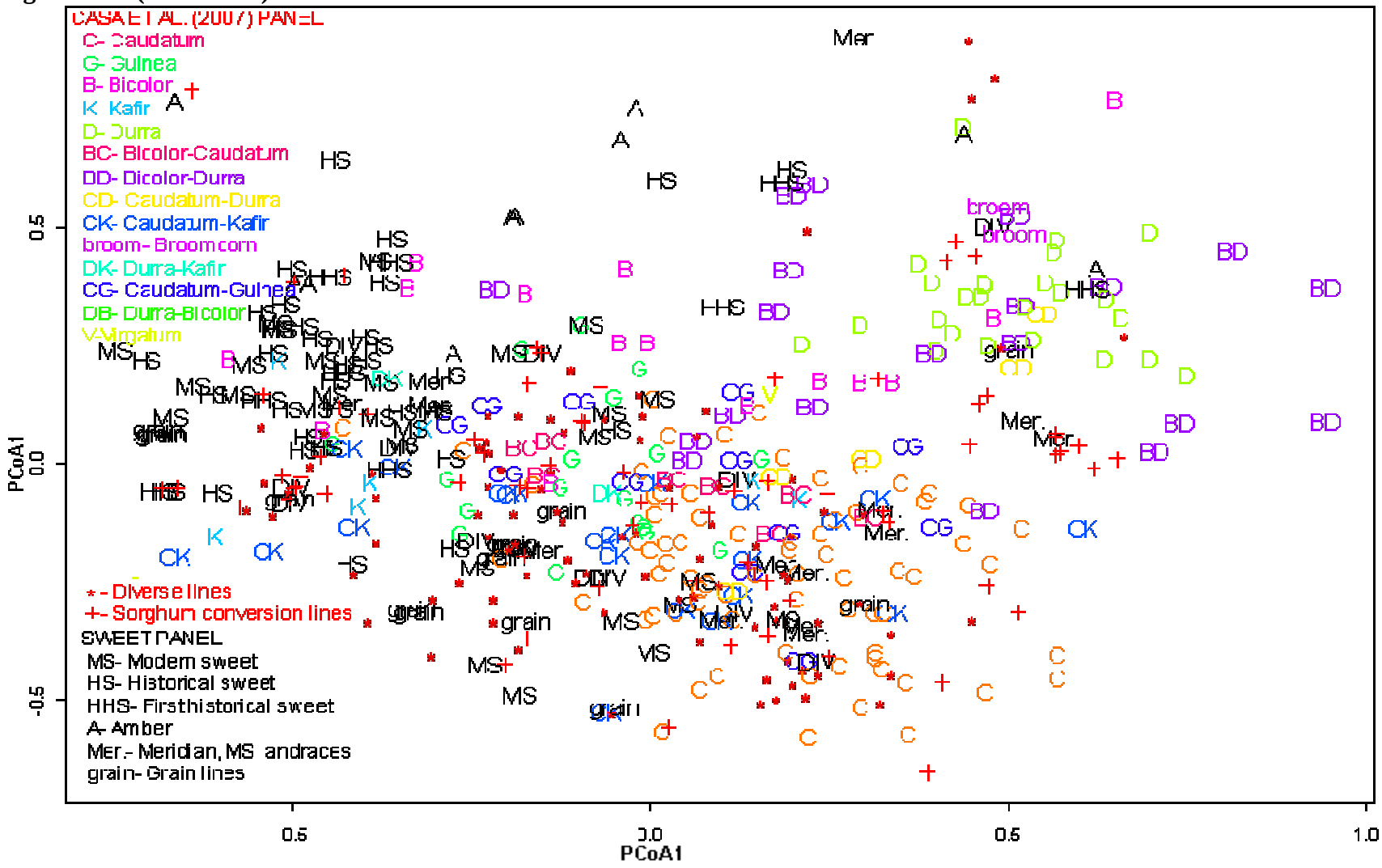

CASA ET AL. (2007) PANEL
C- Caudatum
G- Guinea
B- Bicolor
K- Kafir
D- Durra
Amber
Historical and modern syrup
Modern sugar and energy, MN - landraces
SWEET PANEL
MS- Modern sweet
HS- Historical sweet
HHS- First historical sweet
A- Amber
Mer.- Meridian, MS landraces
grain- Grain lines
PCoA1
0.5
0.0
-0.5

Figure 4.2B (Continued)



Furthermore, these sequences were both located in a 2Mb region flanked by the SSR marker bordering the QTL on the left, and an SSR marker close to the 2LOD peak border on the right (Figure 4.3). The full-length genes (as annotated), the 5' and 3' ends, and nearby regions were sequenced in Rio (~20000bp) and no polymorphisms with BTx623 genome sequence were observed. We then identified nine SSRs spaced through the 2Mb interval. Only one out of the 9 was found to be polymorphic between Rio and BTx623. This marker was included in all analyses.

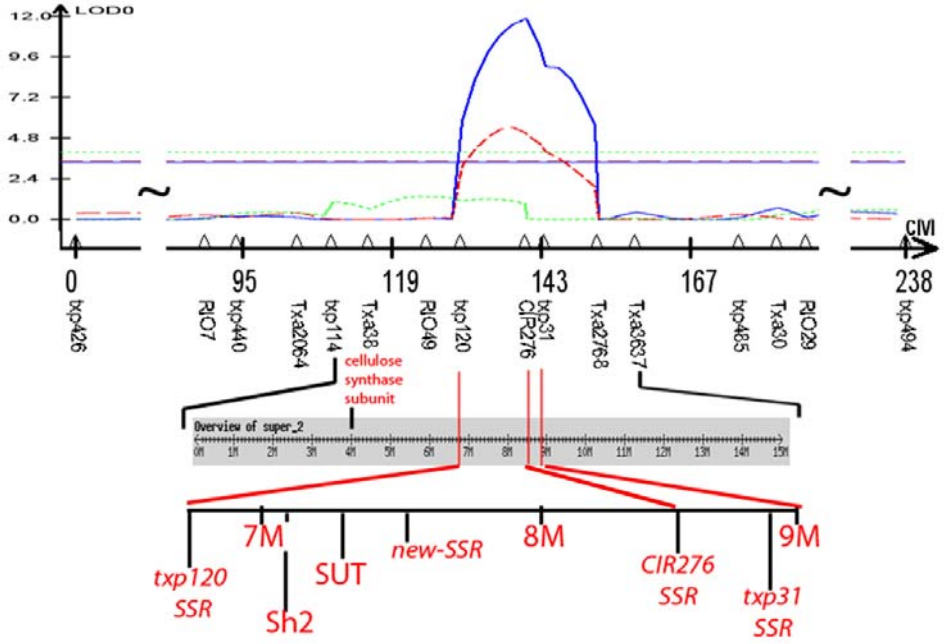


Figure 4.3 Genetic and physical position of chromosome 3 brix QTL and genetic markers.

QTL Cartographer results from Murray et al. (2008a) for the brix QTL on chromosome 3 are on the top. Five SSRs from the QTL region were placed on a 15Mb sorghum genome supercontig shown in the figures center. Of most interest is 2Mb expanded at the bottom to show the position of shrunken2 (*Sh2*), the hypothetical monosaccharide transporter (SUT), and the only new polymorphic SSR that was found in the interval (new-SSR).

Phenotypic analysis: Brix and height values were recorded in all three locations. For the sweet sorghum panel in CS06, brix and HPLC-measured stem sugar had good correlation, with outliers due to bacterial degradation in HPLC samples

(Figure 4.4). Height and brix were positively correlated across locations (Figure 4.5). Height had higher correlations within and across locations than brix, and thus heritability for height was higher. For brix ITH07 was slightly more similar to CS06 than to CS07. ITH07 did not correlate well with CS locations for height, due to photoperiod sensitivity which delayed flowering in some lines.

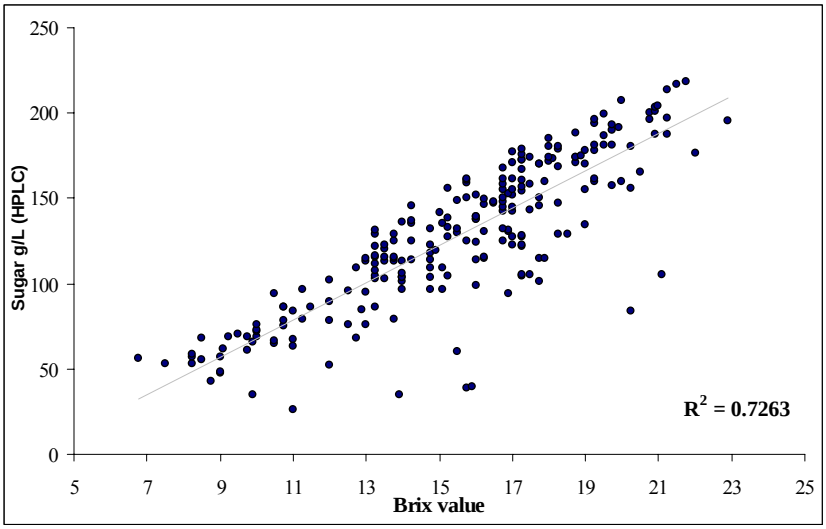


Figure 4.4. Relationship between brix and HPLC measured stem sugar for CS06

A nearly linear relationship exists between these two juice measurements. Samples with high brix and low stem sugar are due to sample degradation.

Population structure and association mapping: To control for false positives in association mapping, *Q* (population structure) and *K* (kinship) matrices were first constructed. Six separate *Q* matrices were calculated using the two most likely population assignments in each of three programs, InStruct, Structure, and NTSYS-pc. InStruct results suggested five or eleven populations were likely with little posterior probability increase after eleven. InStruct DIC criteria also found eleven populations to be most probable. Structure results suggested either four or eleven populations were likely.

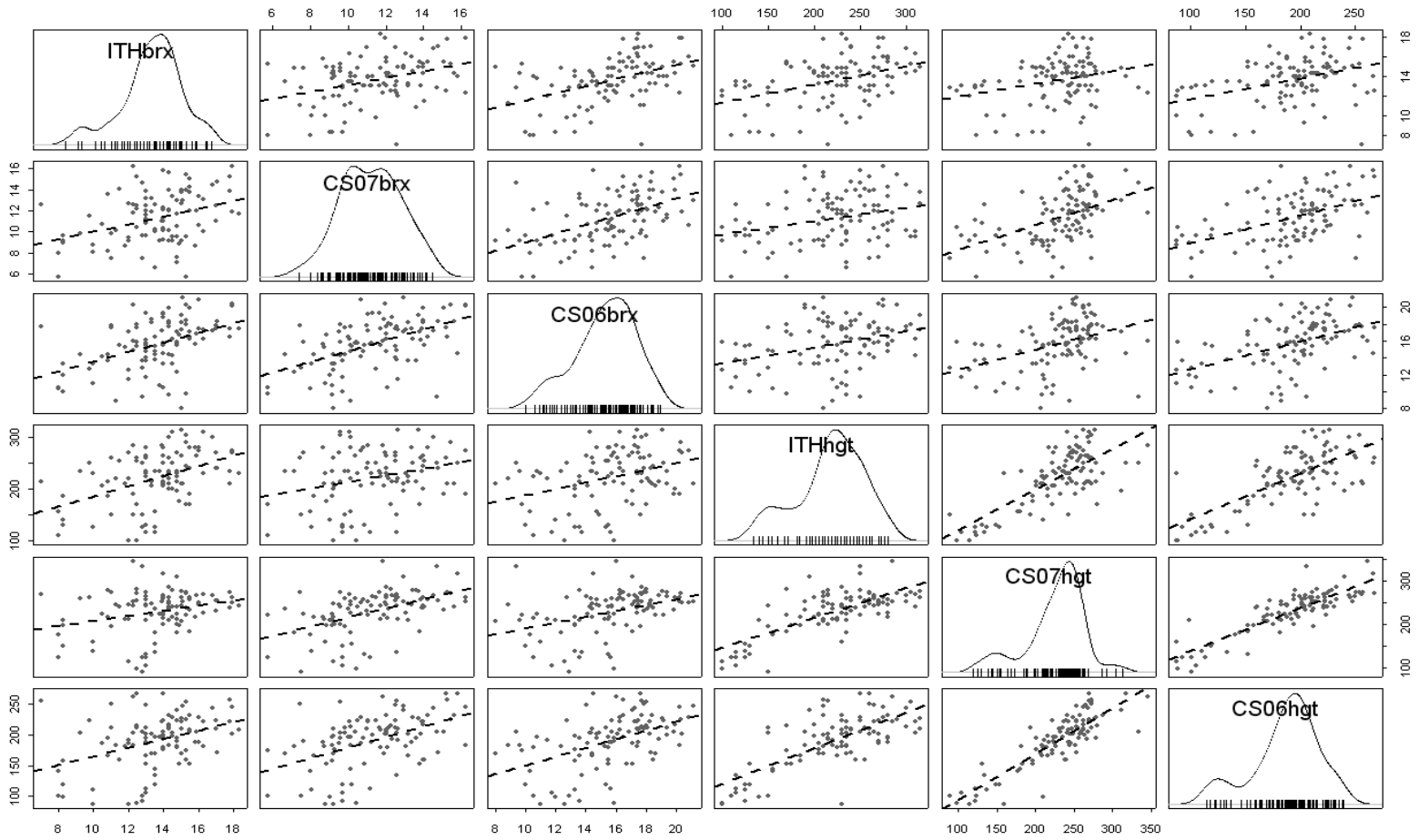


Figure 4.5 Relationship within and between brix and height across three locations.

Units are degree brix and cm for height. Trait names are presented in the center diagonal with a histogram of replicate mean values. Plot regression lines show positive relationships were observed between all sets of measurements but correlation is best within height, specifically between CS06 and CS07.

Structure posterior probability continued to increase marginally past eleven populations, but consistency of runs and population assignment decreased. Because the posterior probability is calculated differently in Structure and InStruct, these cannot be directly compared (H. Gao, personal communication). Using haplotypes for PCoA resulted in eigenvectors very similar to those obtained using individual markers (Figure 4.1). With PCoA, the fifth eigenvector explained less than 5% of the variation and 100% of the cumulative variation was explained by the 12^{th} eigenvector. For association mapping, eigenvectors one through five and, separately, one through twelve were used as *Q* matrix population assignments.

Association analysis: Association mapping was performed for brix and height using the GLM procedure in TASSEL (Bradbury et al., 2007). Of the six *Q* matrices tested, InStruct and Structure using 11 populations had the highest model fit (R^2) and the least number of significant loci (Table 4.3). Using the smaller number of populations for InStruct (k=5) decreased the model fit and for Structure (k=4)also increased the number of positive tests. PCoA eigenvectors performed better than no *Q* matrix at all but very poorly compared to structure analyses. Using PCoA some loci more often failed to converge in TASSEL. In nearly all cases, as more variation was explained by the model, the number of positive tests decreased.

For Brix the MLM model, which included the kinship matrix, *K,* had a better model fit and less positive tests than with *Q* alone. With MLM for Brix, results were nearly identical even if no Q matrix was added. Height was not able to be run with the K matrix on the version of TASSEL tested.

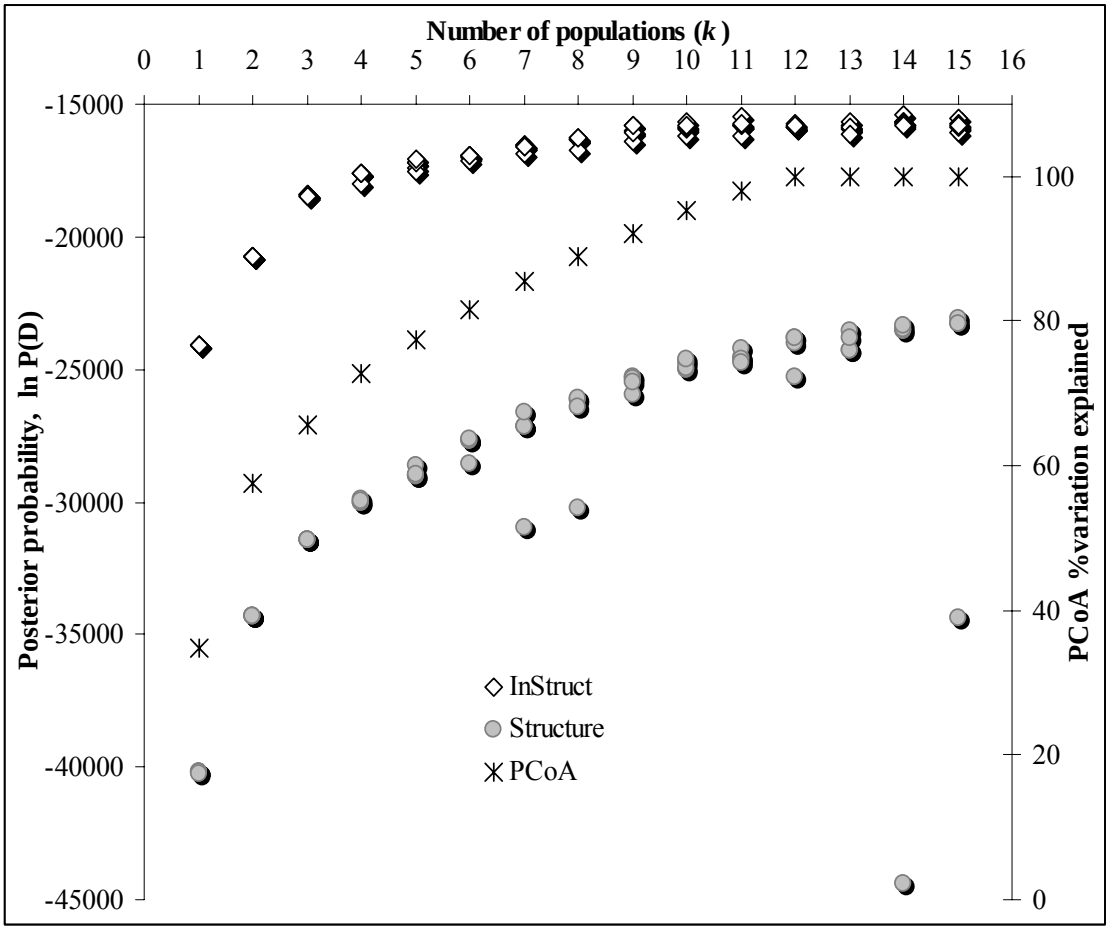


Figure 4.6. Results of population structure analysis using InStruct, Structure and PCoA

Using haplotypes created from markers linked at the same locus, three methods to develop population assignment vectors were used. With all three programs, 4-5 populations and 11-12 populations had the highest population assignments, most consistent posterior probability, and best increase in posterior probability.

Association analysis is reported for the MLM model method using InStruct with 11 populations (Table 4.3). Using a Bonferroni corrected cutoff of 0.05 ($1.3x10^{-4}$), five significant associations were detected for height, one was detected for brix. One marker, SB00016.1, was most significant for height and nearly significant for brix. For brix the only significant marker was G1R.

Table 4.3. TASSEL association of 369 markers across the genome for brix and height

***Q* matrix**		Brix		Height	
GLM	**Populations**	**R^2 model median**	**Positive tests***	**R^2 model median**	**Positive tests***
InStruct	11	0.4	3	0.31	5
InStruct	5	0.29	1	0.17	7
Structure	11	0.4	0	0.32	4
Structure	4	0.22	2	0.1	8
PCoA	12	0.12	13	0.15	3
PCoA	5	0.08	15	0.08	4
None	0	0.04	20	0.03	4
MLM					
InStruct + *K*	11	0.48	2**	0.57	4**
InStruct + *K*	5	0.44	1**	0.53	4**
Structure + *K*	11	0.45	0**	0.59	2**
Structure + *K*	4	0.43	1**	0.53	3**
PCoA + *K*	12	0.47	4**	0.6	2**
PCoA + *K*	5	0.43	3**	0.57	2**
None + *K*	0	0.42	3**	0.52	4**

*at $p < 0.00013$

**some loci failed to converge

Table 4.4. Markers with a significant p_value at 0.001 or in top 5% of Fst in each category

Additional data for the full set of markers can be found online.
Marker significance values were calculated using MLM in TASSEL

Name of SNP	Name of Locus	Chr. Number	p_value Height	p_value Brix	Fst *A priori*	Fst PCoA	Fst Height	Fst Brix
SB00016.1	pSB0945	9	1.89E-11*	1.64E-04	0.31	0.21	0.1	0.15
Xgap72	SSR	6	2.98E-09*	0.0027	0.19	0.28	0.06	0.09
Xtxp343	SSR	4	4.01E-06*	7.29E-04	0.07	0.12	0	0.04
Xtxp265	SSR	6	6.57E-05*	0.0684	0.08	0.15	0.02	0.05
SB00014.3	pSB0301	10	1.06E-04*	0.0117	0.14	0.26	0.08	0.03
SB00215.1	psb1812	3	4.55E-04	0.3014	0.03	0.06	-0.01	0.02
SB00156.1	pSB0289	3	5.19E-04	0.0703	0.1	0.08	0.03	0
SB00154.4	pSB0142	10	5.28E-04	0.3493	0.29	0.53	0.05	0
SB00135.1	pSB1224	2	8.56E-04	0.1051	0.05	0.14	0.01	-0.02
SB00166.1	G1R	1	0.0019	2.97E-05*	0.25	0.57	0.03	0.06
SB00053.1	PRC0271	3	0.1231	0.7493	0.36	0.77	0	0.19
mSbCIR276	SSR	3	0.0385	0.4162	0.09	0.07	0.03	0.18

Table 4.4. (Continued)

SB00200.1	pSB0122	9	0.0052	0.113	0.24	0.27	0.19	0.18
SB00207.1	C2782	9	0.0235	0.13	0.33	0.49	0.08	0.17
SB00176.3	CrtrB2	0	0.1327	0.5122	0.37	0.68	-0.02	0.16
SB00083.1	pSB1015	6	0.0034	0.0431	0.19	0.54	0.03	0.16
SB00083.2	pSB1015	6	0.0217	0.0476	0.18	0.56	0.03	0.16
SB00028.3	AEST056	7	0.4696	0.1893	0.33	0.42	0.03	0.16
SB00176.5	CrtrB2	0	0.0071	0.005	0.31	0.56	0.02	0.15
SB00099.4	pSB1738	9	0.0019	0.0205	0.33	0.53	0.06	0.15
SB00114.1	AGPss	0	0.0085	0.0089	0.25	0.24	0.02	0.14
SB00114.2	AGPss	0	0.0093	0.0102	0.25	0.24	0.03	0.14
SB00137.1	pSB1310	0	0.1068	0.9577	0.29	0.44	0.1	0.14
SB00101.1	pSB1817	6	0.1297	0.9888	0.3	0.45	0.09	0.14
SB00217.1	pSHR0116	10	0.0858	0.1916	0.24	0.15	0.01	0.13
SB00093.2	pSB1469	1	0.1817	0.7204	0.22	0.5	-0.02	0.12
SB00099.1	pSB1738	9	0.0013	0.0102	0.35	0.51	0.05	0.12
SB00078.1	pSB0745	6	0.0074	0.0491	0.26	0.48	0.01	0.11
SB00109.1	R2266	4	0.9256	0.5007	0.26	0.46	0.23	0.07
SB00084.1	pSB1018	1	0.0148	0.0739	0.17	0.25	0.14	0.04
Xcup42	SSR	10	0.0017	0.0363	0	0	0.13	0.06
SB00022.1	pSB1755	7	0.0015	0.0741	0.07	-0.02	0.11	0.08
SB00094.4	pSB1472	1	0.4345	0.2329	0.28	0.46	0.11	0.05
Xcup71	SSR	4	0.1173	0.1155	0.29	0.53	0.11	0.05
SB00161.1	pSB0716	7	0.0068	0.5694	0.12	0.05	0.1	0.08
SB00049.1	pRC0121	7	0.0105	0.8153	0.08	0.1	0.1	0.07
SB00155.3	pSB0193	4	0.1718	0.8109	0.18	0.21	0.1	0.06
SB00025.1	pSB1865	9	0.0352	0.6686	0.13	0.21	0.1	0.06
SB00115.3	SSIIb	0	0.8323	0.0711	0.23	0.52	0.1	0.05
SB00062.2	PRC1203	1	0.0526	0.2442	0.18	0.23	0.1	0.05
SB00198.1	pSB1057	9	0.5932	0.5977	0.03	0.03	0.1	0.05
SB00175.5	CrtrB1	1	0.3604	0.3765	0.03	0.04	0.1	-0.01
SB00052.1	pRC0162	0	0.0013	0.0778	0.26	0.78	-0.03	-0.01
SB00060.1	PRC1149	2	0.3466	0.1986	0.3	0.72	-0.01	0
SB00118.3	gpt	7	0.6391	0.4349	0.28	0.69	0.01	0.06
SB00029.1	C0086	3	0.3813	0.1303	0.24	0.67	0.02	0.01
SB00170.3	SPP1	0	0.0128	0.2413	0.22	0.64	-0.02	0.01
SB00131.5	LDreg4	8	0.844	0.5707	0.32	0.63	-0.01	0.1
SB00097.1	pSB1600	5	0.2062	0.8194	0.28	0.63	-0.02	0.06
SB00131.4	LDreg4	8	0.0141	0.1171	0.24	0.62	-0.01	0.09
SB00106.1	pSB1916	6	0.0056	0.3226	0.24	0.61	0.01	0.09
SB00170.2	SPP1	0	0.6238	0.1064	0.23	0.61	0	0.03
SB00182.3	pSb0243	3	0.0548	0.8621	0.23	0.61	0	0
SB00179.3	CCD4	2	0.0795	0.1933	0.32	0.6	0.02	0.04
SB00162.1	pSB0088	1	0.0323	0.4315	0.19	0.59	0.03	0.01
SB00170.1	SPP1	0	0.0153	0.0741	0.19	0.59	-0.02	0
SB00037.2	CSU653	1	0.4378	0.4718	0.3	0.59	-0.03	0.02
SB00037.1	CSU653	1	0.4748	0.9343	0.31	0.59	-0.03	0.02
SB00179.4	CCD4	2	0.0193	0.4062	0.36	0.58	0.01	0.08

Table 4.4. (Continued)

SB00046.1	HHU62	8	0.3953	0.648	0.47	0.44	0.01	0.09
SB00088.1	pSB1231	3	0.4321	0.1005	0.4	0.23	-0.02	0.04
SB00141.1	pSB1445	4	0.5411	0.3787	0.36	0.4	-0.02	0.02
SB00094.5	pSB1472	1	0.6129	0.9712	0.36	0.37	0.08	0.01
SB00149.2	PHYB	1	0.1176	0.1735	0.35	0.46	0.01	0.04
SB00159.2	pSB0062	1	0.2274	0.2206	0.34	0.38	-0.01	0.08
SB00197.1	pSB0521	10	0.0531	0.846	0.33	0.54	0.03	-0.01
SB00208.2	CSU535	2	0.8499	0.8256	0.33	0.4	0	0.05
SB00141.2	pSB1445	4	0.3536	0.2548	0.32	0.4	-0.02	0.02

* Significant at 1.3×10^{-4}

FST of populations and markers: Wright's (1965) classical Fst (θ) was used to evaluate genetic differentiation between populations in the panel (Table 4.4). Four separate methods were used for dividing the material into populations to address different biological questions.

1) Based on the *a priori* expectation of sorghum types (Table 4.1 - amber, historical syrup, grain, diverse). FST averaged 0.14 across loci (range: -0.04 to 0.47). Markers with high FST would be useful for distinguishing these *a priori* groups and might also be linked to traits important within only one population.

2) Using the three groups identified in PCoA analysis. FST averaged 0 .26 (range: (-0.02 to 0 .77). Markers had higher FST than our *a priori* division and the highest FST would be useful for assigning germplasm with unknown background to these groups.

3) Using a grouping based on brix. Lines in the top half highest brix in CS06, CS07, and ITH07 were in population three, lines in the bottom half for all locations were in population zero. FST averaged 0.03 (range:(-0.03 to 0.19).

4) Using the number of times a line was in the top half of average height for a location, similar to divisions for brix. FST averaged 0.02 (range: (-0.04 to 0.23).

Markers with high FST when separated by brix and height may be linked to the phenotype of interest, and useful for characterizing different germplasms. Relationships between these estimates of Fst and association results may suggest incomplete correction. Fst marker significance did not correspond to significant associations, except in the case of pSB0945.

Discussion

The three primary objectives of this study were to 1) identify genetic relationship within sweet sorghums; 2) identify genetic relationships between sweet and grain sorghums; and 3) perform association mapping in the sweet sorghum panel to confirm QTL for height and brix.

From historical publications on sweet sorghum, it initially seemed likely that sweet types might be closely related to each other and distant from grain sorghums. Two recent publications have suggested otherwise. Casa et al. (2008), using 377 diverse sorghums including 8 sweet cultivars, found that while some sweet sorghums cluster together, all sorghums cluster within grain sorghum groups (A. Casa, personal communication). This finding was supported by Ritter et al. (2007b) who, using AFLPs, showed that 31 sweet sorghums clustered within three of the five clusters containing 64 diverse grain sorghums.

Harlan and deWet (1972) and others have developed classification for sorghums into five major races, bicolor, caudatum, durra, guinea, and kafir. These divisions are mostly based on panicle and grain characteristics as well as the regions of Africa and India where they are commonly found. Sweet sorghums have not been managed based on panicle or grain characteristics, and the referenced origins, African countries of sorghum domestication, South Africa, and China, provide little insight.

Therefore the relationships of sweet sorghums to the major sorghum races was inconsistent.

Results of our study, like Ritter et al. (2007b), identified three distinct groups of sweet sorghum which often are classified together. These major types were syrup (historical and some modern), modern sugar and energy types with associated landrace parents, and amber types. These divisions were supported by PCoA, measures of Fst, phenotypic observations, and structure analysis. Structure analysis and association results suggested that within each of the three sweet sorghum types up to an additional eight additional subpopulation divisions exist.

Within the sweet sorghum panel, the historical and modern syrup population had the best representation and the least diversity. Among sweet sorghums cultivars the historical lines are best known and the modern lines are some of the most common for syrup: 'Orange', 'Sumac', White African', 'Collier', 'Sugar-Drip', 'N98' through 'N110', 'Della', and 'Bailey'. Phenotypically, this material generally had straight, tall, very juicy, medium-large diameter stalks. Across the lines the juice had high average brix, but lower than the sorghums developed for sugar production. Two of the sorghums developed for sugar and having very high brix, 'Keller', and 'Wray', were near classification in this group based on PCoA. The clustering of these syrup types reflects selections from historical material and shared pedigrees from syrup x syrup crosses. Typically, improved syrup sorghums with increased disease and pest resistance and decreased lodging were developed within the same program. InStruct and Structure divided this population into 4 subpopulations of 19, 18, 14, and 12 individuals. An interesting case is 'Sugar Drip' divided into two groups. Based on polymorphism data 'Sugar Drip' was likely heterozygous at some loci which became fixed as different sets of seeds were isolated and maintained separately.

Sugar and Energy: Modern sweet sorghum cultivars for sugar and energy production such as 'Rio', 'Ramada', 'Top76-6', and 'M81E', tended to cluster together with MN landrace lines. Most MN landraces in the panel were specifically chosen because they were in the pedigrees of modern sweet sorghum lines. These MN lines were also from the center of sorghum domestication around Sudan, Ethiopia, and Uganda. This population was very diverse for brix and height. Nearly all of the lines are photoperiod sensitive, and have very thick stalks, some with hard rinds like sugarcane. The modern sugar and energy cultivars, had very high brix while the MN landrace progenitors did not. Many of these lines, especially 'MN1500' produced very high biomass. We initially believed that 'MN1500' was 'Grassl', a cultivar selected out of MN1500, but the high heterozygosity suggested that it is likely the landrace MN1500. In contrast to the expectation that the sweet sorghums using MN lines would have a narrow genetic base, the heterogeneity in these landraces likely contributed to the diversity seen in the modern cultivars. Population analyses further divided this population into groups of 24 (most sugar energy and MN lines), 9 ('Rio', 'Keller', 'Wray'), and 6 (grain and forage).

Amber: Amber and honey sorghums were very distinct from the other two populations but were also very diverse within the population. The weak clustering of amber may be partially the result of a limited number of lines being included. Amber sorghums are not included in published pedigrees of modern sweet sorghum but were among the earliest sweet sorghum in the US. Unlike most sweet lines, amber lines tended to senesce in CS06 and CS07 locations, but did not in ITH07. Possibly as a result, amber lines had relatively higher brix in ITH07 than in either CS06 or CS07. Amber types among the sweet sorghums also had the least consistency of brix between environments with lines having a high brix in only one location. This is why no amber lines were identified as top sugar producers. Structure and InStruct further divided the

ambers into subpopulations of 12 (all but one line with amber in the name and 'Sucre Drome'), 6 ('Honey', '7035S'), 3 (grain sorghums). PCoA suggested that 'Honey' sorghums were most like race durra suggesting a geographic similarity since honey accessions and durra are both from India. The amber population also had some of the most unusual lines, e.g. '7035S' was the tallest line in the panel, had a very large stalk, it was the only line not to tiller at all and to senesce before it flowered in CS06. Sucre Drome was an interesting line in this panel because it was the only one with a 'dry' stalk, a dominant gene that reduced stem moisture by 50% of the panel average and may be useful for cellulosic biofuel.

Sweet & Grain comparison: PCoA was useful to visualize genetic distances between sorghum races, between the sweet sorghum panel and Casa et al.(2008) and between individuals. Using PCoA, races tended to cluster together but were not distinctly separated as observed in rice, or maize (rice-Thomson et al., 2006; maize-Liu et al., 2003; Warburton et al., 2008; Hamblin et al. 2008; or dog/canine – Parker et al., 2004). Rio and BTx623 appeared to be closely related, and both were fairly distant from much of the other material. This suggests that variation found in the bi-parental population investigated in Murray et al. (2008a, 2008b) was more likely to be functional and not confounded by extreme divergence.

The relationships in the sweet sorghum panel using only the SSRs appeared to be similar to what was seen when the 329SNPs were also included. In contrast, the principal coordinate vectors explained far less genetic variation. This discrepancy likely resulted from more rare alleles per locus, fewer loci, and a larger and more diverse germplasm set. From the combined data sets it appeared that the syrup sweet sorghums clustered best with kafirs, and modern sugar energy sorghums and the landraces cluster best with caudatums. Amber types appeared to be poorly represented in the panel of Casa et al. (2008) but clustered most like bicolor types. In general, the

SSR PCoA shows that the panels are structured very differently, the sweet sorghum panel has greater diversity from amber types, the panel of Casa et al (2008) has much more diversity from durra and caudatum types.

Population structure and relatedness in the sweet sorghum panel:

Population structure and relatedness is important in association mapping to avoid spurious associations (Pritchard et al. 2000, Yu et al. 2006)**.** We attempted three separate methods for population assignment of lines, Stucture, InStruct, and PCoA. All three methods suggested that 3 populations was an absolute minimum, and both 4-5 and 11-12 populations were selected as meeting our selection criteria. This consistency was reassuring given that the three methods use different approaches for calculation. Though Structure is widely used for identifying population structure, the program was developed for natural outcrossing populations. The sweet sorghum panel violates Structure's assumption of Hardy-Weinberg equilibrium. InStruct, based on Structure, is a more valid method for a self pollinated domesticated crop such as sorghum, because it relaxes the assumption of Hardy-Weinberg equilibrium (Gao et al. 2007). It was therefore surprising that Structure and InStruct resulted in nearly identical conclusions in this study. Finally, principal component analysis has been proposed to correct for population structure (Price et al., 2006) and similarly PCoA has been used by Cockram et al. (2008). PCoA explained far more variation in this study than in Cockram et al., but the results of this approach were still disappointing for controlling for population structure.

From our results and model fit, it appears that using the kinship matrix (*K* matrix; Yu et al., 2006) better controlled for relatedness than any measure of population structure (*Q* matrix). In fact, we had better fit and fewer positive tests using *K* without *Q* than with any *Q* alone. It seems likely that this will be true for most bred material where admixed diverse crosses are routine.

Brix and height QTL association: Sorghum has much greater linkage disequilibrium than maize extending from a few kb to over a Mb (Hamblin et al. 2005). Associations detected in this panel are therefore probably a result of linkage to a gene of interest rather than variation caused by the marker itself. Only two positive height associations Xgap72, and Xtxp265, were on the same chromosome residing about 10Mb away. QTL for height and or flowering time have been found in this location on chromosome 6, corresponding to the photoperiod sensitivity gene *ma1*(Lin et al.,1995, Rami et al., 1998; Brown et al., 2006, Murray et al. 2008). This gene has undergone extremely strong selection for temperate adaptation in sorghum and detection over a long physical distance was not surprising.

The most significant QTL in this study was found on chromosome 9 for height. QTL for height in this location have been detected both by QTL linkage analysis (Pereira and Lee 1995; Lin et al., 1995; Murray et al. 2008a) and by association analysis. Association analysis in the panel of Casa et al. (2008) detected a peak approximately 400kb away, with significant locus associations on both sides of the marker (SB00016.1) used in this study (Patrick Brown, personal communication). This locus would also be expected to have long range LD given the strength of selection in sorghum for height. Finally, the third most significant marker for height was detected on chromosome 4, which, to our knowledge, has not been previously reported.

The only significant association for brix, on chromosome 1, has also not been previously reported in linkage mapping studies. However, Murray et al. (2008a) did detect a strong peak in this location (the closest marker was Txp482, 5Mb away). This peak explained up to 9% of the variation for brix and sugar, but was slightly below the stringent threshold to be reported as significant (unpublished data). On the physical genome sequence, a sorghum homologue to glucose-6-phosphate isomerase (EC

5.3.1.9) is located ~12kb away, the third closest predicted gene. Although this enzyme has not previously been implicated in stem sugar accumulation, it is known to convert D-glucose 6-phosphate into D-fructose 6-phosphate, both which are important for synthesizing sucrose (Kanehisa et al., 2006).

We also attempted to identify additional markers to support a QTL for Brix on chromosome 3 detected by Natoli et al (2002) and Murray et al. (2008a), but were unsuccessful. Two reasonable candidate genes for brix were investigated but were found to lack polymorphism. Association analysis using 3 SSRs and one SNP in this region did not detect any significant associations. Assuming the QTL results of Natoli et al (2002) and Murray et al. (2008a) are real, three main causes could explain the absence of a brix association on chromosome 3. First, lower heritability of brix than height suggests we should expect lower power for detection. Second, complex epistatic effects with height and/ or genetic background may exist for brix. Epistasis is suggested because various studies have identified both additive and dominant inheritance for brix and linkage mapping has placed brix QTL in at least four separate locations. Epistatic interactions with height QTL have been detected in linkage analysis and in this study two associations for height were nearly significant for brix. Third, given that there appear to be at least three populations of sweet sorghums with multiple levels of relatedness, there may not be adequate control with the *Q* or *K* matrices and/ or enough individuals in each population to have the power to detect brix loci in sorghum. While height appears to have transcended population structure and relatedness, brix does not.

Meaning for germplasm collections and breeding: The results of this analysis suggest that for genetic studies, and/ or core collection development as few as five lines from the sweet sorghum panel could be selected to represent 90% of the SNP alleles identified. Thus, within the sweet sorghum panel, many of the accessions

could be considered redundant for germplasm conservation, especially in the population of syrup cultivars. These differences reflect close pedigrees with similar parentage.

To identify the most informative markers to differentiate the three main populations, population assignments from PCoA were used to calculate Fst. A few of the markers having high Fst (PCoA column in Table 4.3) could be applied to identify a population for sweet sorghums not included in this panel.

How diversity is portioned within sweet sorghums and between sweet and grain sorghums has implications for how germplasm is maintained. One interesting observation about accessions sharing a name, the six 'Sugar Drips' for example, is that the older the cultivar is, the more diverse the lines are. There are two obvious explanations, residual heterozygosity would be greater for landraces than elite cultivars, and over time more outcrossing is likely to occur. Inexpensive DNA markers make testing easy, but to reduce redundancy in core collections duplicates of modern named material should be removed before historical landraces.

For crop improvement understanding the diversity present within the three identified populations is important. For breeding of syrup cultivars, a larger and less diverse selection of elite material from the modern syrup cultivars would be most useful. For breeding energy types for biofuel (lignocellulose and sugar), further selections from within the sugar and energy population and hybrids across populations would be most appropriate.

Conclusion

We have identified three populations of sweet sorghum, with multiple sub populations in each. This information is beneficial to understand the origin of sweet sorghums and for identifying material for further improvement. We have identified a

significant association for brix and identified a nearby candidate gene, glucose-6-phosphate isomerase, to be tested in the future. Our results support epistasis in sorghum, especially with genes for height, but we lack enough power to formally test this relationship. Future work within and across these populations may enable molecular cloning of genes responsible for stem sugar accumulation in sorghum. Understanding the genetic basis for variation in stem sugar may ultimately allow genetic improvement of more complex genomes relative's sugarcane and maize.

REFERENCES

Anas and T. Yoshida. 2004. Genetic Diversity among Japanese Cultivated Sorghum Assessed with Simple Sequence Repeats Markers. Plant Production Science. **7**:217-223.

Aranzana, M.J., S. Kim, K. Zhao, E. Bakker, M. Horto, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, et al. 2005. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet. 1:e60.

Breseghello, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research.Vol. 27:573-580.

Bian, Y.-l., Y. Seiji, I. Maiko, and H.-W. Cai. 2006. QTLs for sugar content of stalk in sweet sorghum (Sorghum bicolor L. Moench). Agricultural Sciences in China 5:736-744.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Broadhead, D.M., O.H. Colman, and K.C. Freeman. 1970 Dale, a new variety of sweet sorghum for sirup production. Miss. State Expt. Sta. Inform Sheet 1099, 2pp., illus.

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

Broadhead, D.M., K.C. Freeman, O.H. Colman, and N. Zummo.1974.Theis, a new variety of sweet sorghum for sirup production. Miss. State Expt. Sta. Inform Sheet 1236, 2pp., illus.

Broadhead DM, KC Freeman, and N. Zummo. 1978. 'Wray' - A new variety of sweet sorghum for Sugar Production. Miss. Agric. For. Exp. Sin. Res. Rep. 1978, 4(1).

Broadhead, D.M., K.C. Freeman, and N. Zummo. 1979. Keller, A new High-Sucrose Sweet Sorghum With Potential for Sugar Production. Miss. Agric. For. Exp. Sin. Res. Rep. 4(8).

Broadhead DM, KC Freeman, and N. Zummo. 1981. M81E A new variety of sweet sorghum. Miss. State Expt. Sta. Inform Sheet 1309, 2pp., illus.

Brown, P.J., P. Klein, E. Bortiri, C. Acharya, W. Rooney, and S. Kresovich. 2006. Inheritance of inflorescence architecture in sorghum. Theor. Appl. Genet. 113:931–942.

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community resources and strategies for association mapping in sorghum. Crop Sci 48:30-40.

Clark, J.W. 1981. The inheritance of fermentable carbohydrates in stems of Sorghum bicolor (L.) Moench. PhD Thesis, College Station, Texas A&M University.

Cockram, J., J. White, F.J. Leigh, V.J. Lea, E. Chiapparino, D.A. Laurie, I.J. Mackay, W. Powell, and D.M. O'Sullivan. 2008. Association mapping of partitioning loci in barley.BMC Genet. 9:16.

Coleman,O.H., I.E. Stokes, K.C. Freeman, J.L. Dean, J.F. O'Kelly, S.P. Crockett, R. Kuykendall, B.C. Hurt, T.E. Ashley, C.L. Blount, W.L. Giles, P.G. Hogg, R.C. Albritton. 1956. Wiley, A New Variety of Sorgo for Sirup Production in Mississippi. Miss. State Expt. Sta. Inform Sheet

Colman O.H., W.R. Cowley, D.M. Broadhead, K.C. Freeman, and I.E. Stokes.1965 'RIO' A New Disease-Resistant Variety of Sweet Sorghum for Southern United States. USDA ARS34-72

Colman O.H., and D.M. Broadhead. 1968. Brandes, a new variety of sweet sorghum for sirup production. Miss. State Expt. Sta. Inform Sheet 1013, 2pp., illus.

Day, J.L., R.R. Duncan, P.L. Raymer, D.S. Thompson, N. Zummo. 1995. Top 76-6 Sweet Sorghum. Crop Sci. 35(4):1213.

DOE-OSTI. 2008. Energy citations database. [Online] Available at http://www.osti.gov/energycitations/index.jsp (verified 08 May 2008).

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15.

Duncan, R.R., D.M. Broadhead, J.W. Dobson Jr., and C.D. Fisher. 1984. Registration of Bailey Sweet Sorghum Crop Sci. 24: 997-998.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164, 1567– 1587.

Fan, J. B., M. S. Chee, and K. L. Gunderson. 2006. Highly parallel genomic assays. Nat. Rev. Genet. 7: 632-644.

Feltus, F.A., G.E. Hart, K.F. Schertz, A.M. Casa, S. Kresovich, S. Abraham, P.E. Klein, P.J. Brown, and A.H. Paterson. 2006. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. TAG. 112:1295–1305.

Freeman, K.C, D.M. Broadhead, and N. Zummo. 1973. Culture of Sweet Sorghum for Sirup Production. USDA Agriculture Handbook No. 441. 30 p.

Freeman, K.C, W.R. Cowley, B.A. Smith, N. Zummo, L. Reyes, D.M. Broadhead, D.T. Rosenow, and K.A. Sund. 1974. 'Ramada': A new Variety of Sweet Sorghum For Potential Sugar Production In South Texas. USDA Agriculture ARS-S-41.

Freeman, K.C. 1979. Germplasm release of sweet sorghum lines with resistance of downy mildew, leaf anthracnose and rust with adequate combining ability to produce progeny with agronomic characters acceptable for commercial sirup and/or sugar production. Miss. Agric. For. Exp. Sin. Res. Rep. 4(2).

Freeman, K.C, D.M. Broadhead, and N. Zummo. 1986. Sweet Sorghum Culture and Sirup Production. USDA Agriculture Handbook No. 611. 55 p.

Gao, H., S. Williamson, and C. D. Bustamante. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635-1651.

Gorz, H.J., F.A. Haskins, and B.E. Johnson. 1990. N97 through N111 Fifteen Sorghum Inbred Lines. Crop Sci 30(3):762

Hamblin, M.T., M.G.S. Fernandez, A.M. Casa, S.E. Mitchell, A.H. Paterson, and S. Kresovich. 2005. Equilibrium processes cannot explain the high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genetics 171:1247–1256.

Hamblin, M.T., A.M. Casa, H. Sun, S.C. Murray, A.H. Paterson, C.F. Aquadro, and S. Kresovich. 2006. Challenges of detecting directional selection after a bottleneck: Lessons from sorghum bicolor. Genetics 173:953-964.

Hamblin, M.T., M.G.S. Fernandez, M.R. Tuinstra, W.L. Rooney, S. Kresovich. 2007a. Sequence Variation at Candidate Loci in the Starch Metabolism Pathway in Sorghum: Prospects for Linkage Disequilibrium Mapping. Crop Sci. 47: S-125-134.

Hamblin, M.T., M.L. Warburton, and E.S. Buckler. 2007b. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE. 2(12): e1367.

Harlan, J.R., and J.W.J.deWet. 1972. A simplified classification of sorghum. Crop Sci. 12:172–176.

Harrison R.L. and F.R. Miller. 1993. Registration of Della sweet sorghum. Crop Sci. **33**:1416.

House, H.R, M. Gomez, O.S. Murty, Y. Sun, and B.N. Verma. 2000. Development of Some Agricultural Industries in Several African and Asian Countries. p. 131-190. *In* Smith et al. (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, Inc., New York.

Houck, O. 1884. Sorghum sugar. American Journal of Pharmacy. 56:1-4.

Karper, R.E. 1949. Registration of sorghum varieties, V. Agron. J. 41: 536-40.

Kanehisa, M., S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. 2006. From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Res. 34:D354–D357. Kresovich, S., H.E. Brown, and D.M. Broadhead. 1985. Registration of Cowley Sweet Sorghum. Crop Sci 25:200

Kresovich, S., and D.M. Broadhead. 1988a Registration of Smith sweet sorghum. *Crop Science* **28:**195.

Kresovich, S., F.R. Miller, R.L. Monk, R.E. Dominy and D.M. Broadhead. 1988b. Registration of Grassl sweet sorghum. Crop Science 28:194–195.

Lipinsky, E.S., S. Kresovich, S., T.A. McClure, W.T. Lawhon. 1977. Fuels from sugar crops. First quarterly report. p.153 Battelle Columbus Labs., OH , USA.

Lin, Y.-R., K.F. Schertz, and A.H. Paterson. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411.

Lingle, S.E. 1987. Sucrose metabolism in the primary culm of sweet sorghum during development. Crop Sci. 27:1214-1219.

Liu, K., and S.V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Maunder, A.B. 1999. History of cultivar development in the United States: From memoirs of A.B. Maunder-sorghum breeder". p. 191-223. In Smith et al. (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons, Inc., New York.

Menz, M.A., R.R. Klein, J.E. Mullet, J.A. Obert, N.C. Unruh, and P.E. Klein. 2002. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP$^{(R)}$, RFLP and SSR markers. Plant Mol. Biol. 48, 483– 499.

Murray, S.C., A. Sharma, W.L. Rooney, P.E. Klein, J.E. Mullet, S.E. Mitchell, and S. Kresovich (accepted) Genetic improvement of sorghum as a biofuel feedstock I: quantitative loci for stem sugar and grain nonstructural carbohydrates. Crop Science.

Murray, S.C., W.L. Rooney, S.E. Mitchell, A. Sharma, P.E. Klein, J.E. Mullet and S. Kresovich (accepted) Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. Crop Science.

Natoli, A., C. Gorni, F. Chegdani, P. Ajmone Marsan, C. Colombi, C. Lorenzoni, and A. Marocco. 2002. Identification of QTLs associated with sweet sorghum quality. Maydica 47:311-322.

Nei, M. 1972. Genetic distance between populations, American Naturalist, 106:283-292.

Parker, H.G., L.V. Kim, N.B. Sutter, S. Carlson, T.D. Lorentzen, T.B. Malek, et al. 2004. Genetic structure of the purebred domestic dog. Science 304:1160-1164.

Price, A.L., N. J. Patterson , R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38:904–9.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rami, J.F., P. Dufour, G. Trouche, G. Fliedel, C. Mestres, F. Davrieux, P. Blanchard, and P. Hamon. 1998. Quantitative trait loci for pain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor l. Moench). Theor. Appl. Genet. 97:605-616.

R Development Core Team. 2005. R: A language and environment for statistical computing. Vienna, Austria. [Online]. Available at http://www.R-project.org (verified 12 Nov. 2007).

Ritter, K.B. 2007. An investigation into the genetics and physiology of sugar accumulation in sweet sorghum as a potential model for sugarcane. Ph.D. diss. University of Queensland, School of Land, Crop and Food Sciences. St Lucia, Australia.

Ritter, K.B., C.L. McIntyre, I.D. Godwin, D.R. Jordan, and S.C. Chapman. 2007. An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. bicolor (L.) Moench, using AFLP markers. Euphytica. 157:161-176.

Rohlf, F. J., 1990: NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System, Version 2.02. Exeter Software, New York

Rooney, W.L., and C.W. Smith. 2000. Techniques for developing new cultivars. p. 329–347. *In* C.W. Smith and R.A. Frederiksen (ed.) Sorghum: Origin, history, technology and production. John Wiley & Sons, New York.

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

Rozen, S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. p. 365-386. *In:* S. Krawetz, S. Misener (ed.) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ.

Salas Fernandez, M.G., M.T. Hamblin, L. Li, W.L. Rooney, M.R. Tuinstra, and S. Kresovich. In press. Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Sci.

Seetharama, N., K. Rao, V. Subramanian, D. Murty. 1987. Screening for sweet stalk sorghums, and environmental effect on stalk sugar concentrations. In 'Technology and applications for alternative uses of sorghum: proceedings of the national seminar'. Parbhani, India, pp 169–179

Sherwood, S.F. 1923 Starch in sorghum juice. Industrial and Engineering Chemistry. 15:727-728.

Schluhuber, A.M. 1945. Inheritance of stem characters: In certain sorghum varieties and their hybrids. J. Hered. 1945 36: 219-222

Stokes, I.E., O.H. Colman, J.F. O'Kelly and others. 1951. Sart, a new variety of sorgo for sirup production in Mississippi. Miss. State Expt. Sta. Inform Sheet 458, 2pp., illus.

Stokes, I.E., O.H. Colman, J.F. O'Kelly and others. 1953. Tracy, a new mid-season variety of sorgo for sirup production in Mississippi. Miss. State Expt. Sta. Inform Sheet 483, 2pp., illus.

Stokes, I.E., K.C. Freeman and others. 1956. New variety of sorgo sirup developed Wiley. Miss. Farm Res. 19(8): 1,7.

Tarpley, L., and D.M. Vietor. 200) Compartmentation of sucrose during radial transfer in mature sorghum culm. BMC Plant Biol. 7:33.

Tamura,K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

Thomson, M.J., E.M. Septiningsih, F. Suwardjo, T.J. Santoso, T.S. Silitonga, S.R. McCouch. 2007. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 114:731-743.

USDA (?Price, C).1958. 'Brawley' New High-Sugar Sorgo Variety Developed For Irregated Areas. USDA 1063-58:1-2

USDA-ARS, NGRP. 2008. National Genetic Resources Program. Germplasm Resources Information Network (GRIN). Nat. Germplasm Resour. Lab., Beltsville, MD. [Online Database] Available at http://www.ars-grin.gov (verified 5 May. 2008.)

Vermerris, W., A. Saballos, G. Ejeta, N.S. Mosier, M.R. Ladisch, N.C. Carpita. 2007. Molecular Breeding to Enhance Ethanol Production from Corn and Sorghum Stover. Crop Sci. 47: S-142-153

Vietor, D.M., and F.R. Miller. 1990. Assimilation, partitioning, and nonstructural carbohydrate in sweet compared with grain sorghum. Crop Sci. 30:1109-1115.

Vinall, H.N., J.C. Stephens, and J.H Martin. 1936. Identification, history and distribution of common sorghum varieties. U.S.D.A. tech. Bull. No. 506.

Warburton, M.L., J.C. Reif, M. Frisch, M. Bohn, C. Bedoya, X. C. Xia, J. Crossa, J. Franco, D. Hoisington, K. Pixley, et al. 2008. Genetic Diversity in CIMMYT Nontemperate Maize Germplasm: Landraces, Open Pollinated Varieties, and Inbred Lines. Crop Sci. 48:617- 624.

Winberry, J.J. 1980. The Sorghum Syrup Industry: 1854-1975. Agricultural History 54:343-352.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395-420.

Wilson, L.M., Whitt, S.R., Ibanez, A.M., Rocheford, T.R., Goodman, M.M. and Buckler, E.S.I. 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell, 16:2719-2733.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping accounting for multiple levels of relatedness. Nat. Genet. 38:203–208.

CHAPTER 5

DEVELOPMENT OF ENERGY SORGHUM TYPES: LOOKING TO THE FUTURE

Summary

Although sorghum does not have the same popularity (or research dollars) as other crops in the US, it is a promising crop and genetic model for biofeedstocks as well as food, feed, and forage. No crop has yet been selected specifically for maximum energy production per hectare and it is clear that all crops, including sorghum, could be improved as biofeedstocks. If cellulosic biofuel gains hold, two sorghum ideotypes have been proposed for biofuel feedstocks; grain sorghums with improved crop residue (dual-purpose crop) and dedicated cellulosic biomass types. Of these, cellulosic high biomass sorghum seems the most promising for maximum energy production per hectare (USDOE, 2006; Farrell et al., 2006; Somerville, 2007; CERES, 2007). For cellulosic types there are many unknowns. For example, it is unclear whether a line that is completely photoperiod sensitive (i.e., never flowers before frost kill in most of the US) would be more productive than a ratooning non-photoperiod sensitive line that is harvested multiple times over the season. This, like many other questions regarding crop improvement for energy production, remains an important and open research question. To address such questions we can look to forage and sugarcane production and harvest systems as models for dealing with biomass.

Altering grain sorghum to produce improved crop residue and developing dedicated cellulosic crops will require major adjustments to complex and quantitative whole plant phenotypes. Simple backcross pedigree selection is unlikely to be as successful as reciprocal recurrent population improvement for dramatic phenotypic changes. For population improvement, selection mapping (Wisser et al., submitted)

will likely be a useful strategy for mapping complex traits. For the dramatic plant compositional and agronomic changes necessary for a biofuel feedstock, I propose the following steps: 1. Create two diverse synthetic populations from wild, landrace, or different improved cultivars (i.e., combine forage, grain, and sweet types) by intermating material with male genetic sterile line(s). Then select desirable agronomic types in a few cycles of intermating and selfing within the population. 2. Select on plant yield and on composition using phenotypic data generated by NIRS. 3. Perform recurrent reciprocal selection concentrating on yield, agronomic characters, and composition. 4. Using selection mapping, identify alleles with significant shifts in allele frequencies before and after selection. 5. Continue selection on yield, agronomic characters, and composition while using molecular markers to push desirable alleles to fixation. From these two pools, lines with good heterotic interactions would be expected. Inbred lines could easily be developed from the two pools.

For the maximum impact in sorghum as both crop and model there are numerous phenotypic, genomic, agronomic, post-harvest, and economic questions which must be addressed. Here I outline what I believe are the most important and interesting phenoptypic and genotypic projects for improvement of sorghum as an energy crop.

Phenotypic traits for biofuel

Height: Taller plants, although more likely to lodge, have the highest correlation with biomass yield. Therefore, for a dedicated sweet sorghum or lignocellulosic feedstock increasing height should be selected while simultaneously selecting for decreased lodging. For grain sorghums, moderate height should be maintained and other ways to increase biomass, such as tillering, should be explored.

Tillering: Tillers serve as additional crop carbohydrate sinks but are also carbohydrate sources and are likely to be advantageous in energy production as they are in forage sorghums. Tillering plants can take advantage of optimal growing conditions even when there is no grain or stem sugar sink to fill. Tillers only compete with a primary stalk for moisture and soil nutrients. For grain sorghum, when under stress, smaller panicles of non-uniform maturity may result and these traits are undesirable. Although genetic diversity is available for tillering, planting density is also an important consideration. Densely planted material often does not tiller but could have just as many stalks per hectare. Increasing planting density, rather than tillering, has been the strategy for maize.

Stem Size: In the Rio x BTx623 population there appeared to be no clear advantage of larger stems except for a correlation with taller plants. In general, larger stem plants had lower yields and poor biomass composition. Stem size was a heritable phenotype with high diversity in the sweet sorghum association panel and it would be beneficial to identify any pleiotropic effects of the genes that affect this trait. Increases in stalk number, arising from tillering or planting density, have a negative relationship with stem size. In forage production, smaller stems are selected due to succulence, increased leaf/stem ratio, increased protein and increased tillering. Additional work on stem size should be conducted, but in general, I believe it is possible to have stems that are either too small or too thick. Small stems may lodge, have high protein, and produce erratic small panicles with little grain that matures unevenly. Stem that are too large take longer to dry down, do not ratoon as well, and may not take be able to respond as quickly to changes in environmental conditions.

Photoperiod sensitivity/ flowering time: Although obvious, in Murray et al. 2008b (chapter 3) we identified that later flowering material tends to produce higher levels of lignocellulosic biomass. Extreme examples of late flowering are caused by

photoperiod sensitivity genes *ma5* and *ma6.* Lines with *ma5* and *ma6* will not flower with more than 12 hours and 20 minutes of daylight (i.e., anywhere in the continental US) (Rooney and Aydin, 1999). For the highest usable energy production, these lines may be superior to photoperiod insensitive material but more work will need to be conducted. One of the greatest drawbacks of using photoperiod sensitive lines is that they will not set seed in most environments, meaning that a winter greenhouse or nursery is necessary to make crosses and produce seed. This requirement increases the number of generations in a breeding cycle. However, the plant epistatic action with *ma5* and *ma6* genes means that parents could potentially have normal flowering in the target environment and only the F_1 progeny will have extreme photoperiod sensitivity. This condition is acceptable as the material that farmers will use will likely be hybrid (Rooney 2007).

Regrowth/ ratooning: The ability to regrow after mowing is a critical trait for forage, turf, and sugar crops. In Murray et al., 2008b (Chapter 3), we observed regrowth yields 30-70% of the initial cutting, and that genetic variation is available. In addition to increasing yields, regrowth can protect the soil from erosion and serve as a reserve pool of soil organic matter. Tillering and regrowth appear to be correlated. With grain sorghum, tillering has typically been selected against and regrowth has not been selected for. Regrowth is an important trait for profitability in forage sorghums and in close relative sugarcane, it will likely be valuable in biomass sorghums. Tall tillering sorghums have much more obvious regrowth phenotypes but it is likely that wild and weedy sorghums have even more regrowth ability.

Dry stalks: Biofeedstock processing and digestion facilities will, by economic necessity, be close to where the feedstock is produced. However, the harvest and transport cost from field to processing plant will still be substantial when up to 75% of the harvest crop weight is water (Murray et al. 2008b). Extensive phenotype data are

available for dry and juicy stalk in grain sorghum and the majority of these lines have been characterized as juicy. Dry stalk is conditioned by a single gene located on chromosome 6 (Bennetzen et al. 2001). Although antithetical to sweet sorghum, the dry allele for high biomass cellulosic varieties has great potential to increase the economic efficiency of biofeedstock transport. Although stem composition did not appear to be different in the only dry line measured in the association study (Murray et. al. 2008c; chapter 4), research needs to be done to determine if the dry stalk gene has pleiotropic effects with reduced dry yield per hectare.

Cold Tolerance: Nearly all sorghum growing regions in the US must possess early- and late-season cold tolerance and frost survival. Cold tolerance has been investigated as an important trait for stand establishment (early-season) with grain sorghum (Tiryaki and Andrews, 2001; Yu et. al., 2004; Franks et al., 2006). Yet, little work has been done for season extension (early- and late-season cold tolerance). Extending either early or late season cold tolerance would increase the days the crop could photosynthesize, potentially increasing total harvestable energy.

Composition: Forage, silage, lignocellulosic biofuel, and building material uses of sorghum share the primary target of high biomass yield. The main differences are in the composition desired. For forage and silage animal feeding, both palatability and high protein are desired. For lignocellulosic biofuel feedstocks and building material low protein is desired to reduce removal of soil nitrogen and potentially increase yield. The main difference between lignocellulosic biofuel feedstocks and building material is high versus low digestibility, primarily increasing lignin and hemi-cellulose content. Thus, after biomass yield has been improved, a NIRS based pipeline for characterizing composition could sort germplasm into the streams of forage/silage, lignocellulosic biofuel, and building material.

Murray et al. 2008b (Chapter 3) found that composition of leaf and stem appeared to be under separate genetic control. For grain producing cultivars, this result suggests that selecting leaf and stem composition traits separately would maximize improvement. It was also found that protein was under separate genetic control in leaf stem and grain and this observation argues that for dual-purpose sorghums with grain is the primary product, grain protein could be increased while biomass protein could be decreased. This approach would address food versus fuel concerns in dual-purpose crops.

"Green" products: As cost and demand of petroleum products increases amid concerns of global warming, consumers can be expected to continue to look towards "green" products. Although the development of sorghum as a biofuel feedstock could be considered a green product, other uses may be developed in the same pipeline. Sorghum crop residue is already being used to produce a structural board, similar to plywood, in China (Kirie 2008). Corn residue has been investigated for particle board production (Ren et al. 2006) and sorghum could also be used. Sales of bamboo flooring suggest that similar domestic products made from sorghum could also be successful. Finally, straw bale homes have been used for centuries, but the effect of composition on bale strength and insulation has not been investigated, especially for sorghum. Although there is no direct market for sorghum crop residue currently, this option would be a good alternative use to biofeedstock digestion. When evaluating biomass composition for biofeedstocks using NIRS it would be simple and complementary to identify composition which may affect structural building and insulating properties.

Perennialism and wild relatives: A desired criterion of biofuels crops is the ability to grow over years to reduce energy use and costs associated with tilling and planting. Two of *S. bicolor*'s closest genetic relatives *S. propinquum* and *S. halepense*

produce rhizomes that can overwinter and produce new plantlets or 'rammets' in spring. Although *S. propinquum* has been shown to overwinter from rhizomes (Paterson et al., 1995), it evolved and grows as a tropical plant. Only *S. halepense* (hypothesized to be a tetraploid interspecific hybridization between *S. bicolor* and *S. propinquum),* is found across different climates worldwide. However, *S. halepense* is also recognized as a noxious weed, at least partially due to the perennial overwintering rhizomes it produces.

Extreme phenotypic variation for rhizome size and number was observed between accessions of *S. halepense* grown in the greenhouse suggesting genetic variation is available (personal observation). Incorporating perennial, overwintering rhizomes into *S. bicolor* could be advantageous both for biofuels and grain crops if the potential for weediness can be suppressed. The Land Institute (Salina, KS) has been successful at backcrossing *S. halepense* rhizome growth into grain sorghum for perennial development without weediness. This material has complete canopy cover one month before annual sorghum but has the disadvantage of being tetraploid (David Van Tassle and Seth Murray unpublished data). Future development of biofuel sorghum may benefit from backcrossing perennial rhizomes into *S. bicolor* at the diploid level, if weediness properties can be repressed.

Incorporating *S. halepense* into a population improvement program may also be beneficial for traits like disease resistance and biomass composition. Research in rice and tomato has identified genes in the wild relative beneficial for agronomic traits (Fridman, 2004; Xie 2006). Sorghums wild relative, *S. halepense,* has much higher molecular and genetic diversity than domesticated sorghum (Caroline Kellogg, personal communication). This diversity is maintained and masked by the fact that *S. halepense* is a primarily outcrossing tetraploid. To my knowledge very little work has tried to exploit this genepool. Working with weedy non-domesticated *S. halepense* is

very difficult. First, successful tetraploid to diploid crosses are rare. Additionally, self incompatibility, seed shattering, inconsistent flowering, poor composition and the potential for perennial "super weeds" are all likely problems.

Genomic

Marker assisted selection for minimizing confounding phenotypes: Using marker assisted selection (MAS) is promising for many traits in sorghum. MAS is generally used to incorporate traits in breeding that are difficult or expensive to phenotype and affected by a few loci. Traits such as stem sugar, cellulose composition, and carotenoid content are expensive traits to phenotype, especially without the use of NIRS. More importantly, what is not often discussed is the importance of MAS in genetic mapping studies to control for major confounding factors such as height, flowering time, and stand density / tillering. For instance, to develop a QTL population that is not confounded by height, the parents must share the same alleles at all height genes. Without additional testcrosses, the genotypes of the parents at these loci cannot be determined by conventional breeding. However, if markers for all major height QTL were known, parents for crosses not segregating for height could be selected. Because height controls and affects many traits, marker assisted selection would be extremely advantageous for higher heritability in genetic mapping. In the studies presented here (Chapters 2 and 3), the parents segregated at a few major height, flowering time, and tillering genes. This segregation was important for increased variation in tissue yield, but likely masked many of the loci of interest for composition.

Identifying additional polymorphism and developing SNP-chips: There are many genomic resources currently available for sorghum (i.e., sequenced genome, EST libraries, candidate gene diversity studies, and bi-parental genetic, physical and

comparative maps). However, there is still a dearth of genetic polymorphism data. This is because the genome sequence, EST libraries and all large scale data collection have been based on a single genotype, BTx623. Solexa "re-sequencing" would be a very cost effective way to identify additional markers. It is likely that over 20,000 SNPs will be required for a whole genome association study (M.T. Hamblin, unpublished data) and potentially more for fine mapping. A recent Illumina® Goldengate assay was developed by M.T. Hamblin (Murray et al. 2008c; Chapter 4) with the extent of SNPs publically available (~384). This genetic marker collection was successful and cost effective per data point generated. Once additional SNPs are identified, a larger Goldengate assay, or perhaps a custom Affymetrix microarray genotyping assay can be developed. This would make genetic data much less time consuming and expensive to collect in linkage, association, and selection mapping populations.

REFERENCES

Bennetzen, J.L., V. Subramanian, J. Xu, S.S. Salimath, S. Subramanian, D. Bhattramakki, and G.E. Hart. 2001. A framework genetic map of sorghum containing RFLP, SSR and morphological markers. P. 347-356 In: R.L. Phillips and I.K. Vasil (ed.) DNA-based markers in plants. 2nd ed. Kluwer Academic Publishers, Boston.

CERES. 2008 [Online]. Available at http://www.ceres.net/ (verified 19 Feb. 2008).

Farrell, A.E., R.J. Plevin, B.T. Turner, A.D. Jones, M. O'Hare, and D.M. Kammen. 2006. Ethanol can contribute to energy and environmental goals. Science 27:506-508.

Franks, C.D., G.B. Burow, and J.J. Burke. 2006. A comparison of US and Chinese sorghum germplasm for early season cold tolerance. Crop Sci 46:1371-1376.

Fridman, E., F. Carrari, Y. Liu, A.R. Fernie, and D. Zamir. 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786–1789.

Kirei sorghum board. 2008. [Online]. Available at http://www.kireiusa.com/ (verified 19 Feb. 2008).

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang. 1995. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. Proc. Natl. Acad. Sci. USA 92:6127-6131.

Ren, H., T. L. Richard, Z. Chen, M. Kuo, Y. Bian, K. J. Moore, and P. Patrick. 2006. Ensiling corn stover: Effect of feedstock preservation on particleboard performance. Biotech. Progress22(1): 78-85.

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

Rooney, W.L., J.Blumenthal, B. Bean, J.E. Mullet. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod. Bioref. 1:147–157.

Somerville, C. 2007. Biofuels. Curr. Biol. 17: R115-R119.

Tiryaki, I., and D.J. Andrews. 2001. Germination and seedling cold tolerance in sorghum: I. Evaluation of rapid screening methods. Agron. J. 93:1386–1391.

Wisser RJ, Murray SC, Kresovich S, Nelson RJ (submitted) Recurrent selection mapping of quantitative trait loci for northern leaf blight resistance in maize. Genetics.

USDOE. 2006. Breaking the biological barriers to cellulosic ethanol: A joint research agenda. DOE/SC 0095. USDOE Office of Science and Office of Energy Efficiency and Renewable Energy, Washington, DC. [Online]. Available at http://genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml (verified 14 Nov. 2007).

Xie, X., M.H. Song, F. Jin, S.N. Ahn, J.P. Suh, H.G. Hwang, and S.R. McCouch. 2006. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon.* Theor. Appl. Genet. 113:885–894.

Yu, J., M.R. Tuinstra, M.M. Claassen, W.B. Gordon, and M.D. Witt. 2004. Analysis of cold tolerance in sorghum under controlled environment conditions. Field Crops Res. 84:21-30.