GENOTYPING BY SEQUENCING FOR IDENTIFICATION AND MAPPING OF QTLS FOR BIOENERGY-RELATED TRAITS IN SWEET SORGHUM

by

Kanokwan Teingtham

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Agronomy & Horticulture

(Plant Breeding and Genetics)

Under Supervision of Professor Ismail Dweikat

Lincoln, Nebraska

August, 2016

GENOTYPING BY SEQUENCING FOR IDENTIFICATION AND MAPPING OF QTLS FOR BIOENERGY-RELATED TRAITS IN SWEET SORGHUM

Kanokwan Teingtham, Ph.D.

University of Nebraska, 2016

Advisor: Ismail Dweikat

Sweet sorghum (*Sorghum bicolor* L. Moench) is a promising bioenergy crop. To increase the productivity of this crop, marker-assisted breeding will be important to advance genetic improvement of sweet sorghum. The objective of the present study was to identify quantitative trait loci (QTLs) associated with bioenergy-related traits in sweet sorghum. We used 188 F_7 recombinant inbred lines (RILs) derived from a cross between sweet sorghum (Wray) and grain sorghum (Macia). The RILs and their parental lines were grown at two locations in 2012 and 2013. Genotyping-by-sequencing analysis of the RILs allowed the construction of a high-density genetic map with 979 single nucleotide polymorphisms. Using the inclusive composite interval mapping of additive QTLs, a total of 29 QTLs for bioenergy-related traits in sorghum were identified, including anthesis date, plant height, head moisture content, biomass yield, stem diameter, brix, grain yield, and 100 seed weight. Major QTLs for anthesis date and head moisture content were detected on chromosome 6, and explained 29.45% and 20.65% of the phenotypic variances (PVE), respectively. Major QTLs for plant height (29.51% PVE) and total biomass yield (16.46% PVE) were detected on chromosome 7, and QTLs for stem diameter (9.43% PVE) and 100 seed weight (22.97% PVE) were detected on chromosome 1. A major

QTL for brix (39.92% PVE) and grain yield (49.14%) PVE co-localized on chromosome 3, was detected consistently across four environments, and is closely associated with a SWEET sugar transporter gene. Additionally, several other QTLs for brix identified in this study or reported previously were found to be associated with sugar transporter genes. The identified QTLs in this study will help to further understand the underlying genes associated with bioenergy-related traits and could be used for development of molecular markers for marker-assisted selection.

The C:N ratio might be an interesting trait for QTL analysis. The variation of C:N ratio among different genotypes of sorghum should be determined. In 2013, Macia, Wray, and four selected RILs with low and high stalk sugar content were determined for carbon and nitrogen partitioning and C:N ratio in stems, leaves, and heads. Carbon and nitrogen partitioning to plant parts was related to the biomass partitioning. Head was the largest portion of carbon and nitrogen in low stalk sugar content lines while stem was the largest portion of those in high stalk sugar content lines. C:N ratio of the stem, leaf, and head were not different between the six sorghum lines.

Acknowledgements

First of all, I would like to express my profound gratitude and sincerest appreciation to my advisor, Dr. Ismail Dweikat, for his patient guidance and continuous support of my Ph.D. study and research.

Profound gratitude and appreciation are also extended to the advisory committee members, Dr. Timothy Arkebauer, Dr. Aaron Lorenz, and Dr. Kent Eskridge for their excellent and helpful advice. Sincere gratitude is given to Dr. David Braun and Dr. Brian Waters for his insightful suggestions in manuscript completion.

There are no words to express the indebtedness to Dr. John Rajewski for his valuable assistance in the field experiments and treasured consultations.

I sincerely thank my laboratory colleagues and friends especially, Niegel La Borde, Mallieswari Gelli, Piyaporn Phansak, Pattama Srinamgoen, Rungravee Boontung, and Veravut Dheraprasart for their valuable help and encouragement.

Very special thanks go to the Royal Thai Government for awarding the scholarship, the Plant Feedstock Genomics for Bioenergy # DE-SC0006810 to DB et al., for the research grant, and Department of Agronomy and Horticulture at the University of Nebraska-Lincoln for giving me the research assistantships.

My deepest gratitude is expressed to my beloved family for supporting, and understanding.

Table of Contents

List of Tables

Page

List of Figures

Page

List of Appendices

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction and Literature Review

Sorghum, *Sorghum bicolor* (L.) Moench, originated in Africa (Doggett 1988; Kimber 2000). Sorghum is a C_4 self-pollinating species in the family Poaceae, and has a haploid chromosome number of 10 (2n=2x=20). Cultivated sorghum is classed as characters of mature heads and spikelets into five basic races: bicolor, guinea, caudatum, kafir, and durra (see Fig. 1) and intermediate races from crosses among the five basic races can be identified (Harlan and de Wet 1972). More than 37,000 accessions of sorghum from 92 countries are maintained at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Sorghum is also classified into four types as conventional uses: grain sorghum, sweet sorghum, forage sorghum, and fiber sorghum (Zegada-Lizarazu and Monti 2012). Grain sorghum is comprised of dwarf varieties rich in starch in the grains; sweet sorghum is tall with a high sugar content in stalks. Forage sorghum is varieties with high protein and fiber content, and fiber sorghum is tall and rich in cellulose and hemicelluloses (Zegada-Lizarazu and Monti 2012). Sorghum is used for multiple purposes including food, feed, fodder, fuel, and fiber (Rao et al. 2014). Sorghum is the world's fourth cereal crop in terms of production quantity and is a staple food crop in semi-arid regions (FAOSTAT 2015; Rao et al. 2014). In recent years, sweet sorghum has served as a potential bioenergy feedstock (Dweikat 2012; Reddy et al. 2005; Rooney et al. 2007; Turhollow et al. 2010; Zegada-Lizarazu and Monti 2012).

Sweet sorghum is a potential bioenergy feedstock

Bioenergy is an alternative and renewable energy derived from biomass resources. In 2005, the U.S. Department of Energy established the billion ton study. The goal of the study was to produce one billion tons of biomass sufficient to generate 60 billion gallons of ethanol a year to replace 30% of current U.S. petroleum consumption with biofuels. It is expected that bioenergy will provide 30% of the world's energy by 2050 (Guo et al. 2015a). Sorghum varieties with juicy and sweet stalks are identified as "sweet sorghum" (Rao and Kumar 2013). Sweet sorghum has been designated as a preferred biomass crop for biofuel because of its high yield in biomass and fermentable sugars (Dweikat 2012; Reddy et al. 2005; Rooney et al. 2007; Turhollow et al. 2010; Zegada-Lizarazu and Monti 2012). Sweet sorghum is highly productive, with low input requirements, and is drought-tolerant (Rooney et al. 2007; Zegada-Lizarazu and Monti 2012). Sweet sorghum accumulates large amounts of carbohydrates in its stalk and produces total biomass as high as 30 Mg ha^{-1} (Bihmidine et al. 2015; Rooney et al. 2007). Sweet sorghum juices contain approximately 16–18% fermentable sugar, which can be directly fermented into ethanol by yeast (Fig. 2) (Ratnavathi et al. 2011). Following juice extraction, the pressed stalk can be compressed into combustible pellets. Thus, sweet sorghum could be used for both biofuel and thermo-electrical energy (Zegada-Lizarazu and Monti 2012). Sweet sorghum has a biomass production capacity close to sugarcane in the tropics while sweet sorghum requires less water than is required for sugarcane (Almodares and Hadi 2009; Monk et al. 1984). Sweet sorghum may be harvested 3-4 months after planting and planted 1-2 times a year (Almodares and Hadi 2009). Regassa and Wortmann (2014)

reviewed that ethanol yield produced from sweet sorghum is high as 3,560 L ha^{-1} and is greater than the other major potential energy crops: sugar beet, maize (grain), wheat (grain), except sugarcane (Table 1). Additionally, output energy per fossil input of sweet sorghum is higher than sugarcane, sugar beet, maize, and wheat (Almodares and Hadi 2009; Dweikat 2012). However, sweet sorghum has limitations for bioenergy production including sugar loss after harvest, transportation and storage of stalks and juice, restriction of the growing season in temperate regions, and low seed yield in seed production (Regassa and Wortmann 2014). In order to overcome these limitations, sweet sorghum should be genetically improved. The successful use of sweet sorghum as bioenergy feedstock will depend largely on the improvement of bioenergy-related traits and subsequent conversion technologies.

Sweet sorghum improvement for bioenergy production

Several sweet sorghum varieties in the United States were released in the 1950s and 1960s by the U.S. Sugar Crops Field Station at Meridian, Mississippi and they were developed from exotic germplasm in Africa (Hunter and Anderson 1997). Presently, more than 2,100 accessions of sweet sorghum have been assembled in the Germplasm Resources Information Network (GRIN, http://www.ars-grin.gov/). Utilization of cytoplasmic male sterility is one of the available strategies for sorghum improvement (Dweikat 2014; Quinby 1974b). Stephens and Holland (1954) discovered cytoplasmic male sterility in sorghum. Greatly, the restorer lines (R-lines) of sweet sorghum were developed, and Wray, which is the sweet sorghum and the elite R-line, has been used in

several sweet sorghum breeding programs (Ali et al. 2008; Dweikat 2014; Ritter et al. 2007b). Many methods of plant breeding have been used in sorghum breeding programs, but the modified mass selection is recommended for practical breeding in some national sorghum breeding programs because the method is simple, inexpensive, and takes a short amount of time (Dweikat 2014). Some of the sweet sorghum cultivars released in the United States are Balley, Dale, Della, M8IE, Rio, Thesis, and Topper 76-6. Sweet sorghum varieties have also been significantly developed in India and China for utilization in these regions (Stefaniak and Rooney 2013).

Maturity is an important trait that should be considered for improvement of sweet sorghum for bioenergy production. Sorghum is a short-day plant; the vegetative growth will remain until the day length becomes shorter than the critical photoperiod, which is 13 hours for sorghum (Miller et al. 1968). Days to anthesis is a measurement of maturity, late maturity of sweet sorghum in temperate regions highlight the risk of freezing that affects stalk yields and sugar (Broadhead 1969). Sorghum flowering time or maturity genes were identified at six loci. The first four genes, *Ma1-Ma4* were recognized by Quinby (1967) and *Ma5* and *Ma6* were recently reported (Mullet and Rooney 2013; Murphy et al. 2014; Rooney and Aydin 1999; Yang et al. 2014). A tropical sorghum variety can be simply converted into a temperate zone variety by substituting one recessive maturity allele for a dominant one (Quinby 1974b). Plant height is another important trait for sweet sorghum production. Plant height of sweet sorghum has a high correlation with biomass yield (Felderhoff et al. 2012; Monk et al. 1984; Ritter et al. 2008). However, lodging can become a serious problem for very tall sweet sorghum

varieties (Monk et al. 1984). Plant height of sorghum is controlled by four dwarfing genes, *Dw1-Dw4* (Quinby and Karper 1953). The dwarfing genes have been used for sorghum breeding to reduce lodging for efficient mechanical harvesting (Quinby 1974a). Additional important traits for sweet sorghum include sugar yield, sugar concentration, sugar quality, lignin concentrations, and agronomic adaptation traits such as stalk rot resistance, seedling cold tolerance, borer tolerance, and drought (Dweikat 2012; Stefaniak and Rooney 2013). Additionally, to be an alternative sustainable energy and food crop, sweet sorghum should be further developed for an efficient multipurpose crop. Crossing between grain sorghum and sweet sorghum could be ideal for sweet-grain sorghum improvement.

Plant breeding is influenced by various approaches from the traditional systems to structural and functional genomics strategies (Ganeshan et al. 2010). Sorghum is an attractive model for functional genomics of grass species because it has a small genome (~730 Mbp) and its complete genome has been sequenced (Paterson et al. 2009). Sorghum genome is smaller and less complex than maize (~2300 Mbp), sugarcane (~7440 Mbp), and wheat (~17,000 Mbp) (Brenchley et al. 2012; Grivet and Arruda 2002; Lee 1996; Schnable et al. 2009). Molecular technology is a significant tool used in plant breeding to achieve desirable traits for crop improvement. Genomics and marker-assisted breeding will be necessary for the genetic improvement of sweet sorghum (Rooney et al. 2007). Genomic regions linked to complex traits can be identified by genetic mapping and quantitative trait locus (QTL) analysis (Shehzad and Okuno 2014).

QTL mapping

QTL mapping with molecular markers is the first strategy in genetic studies. In plant breeding, QTL mapping is an essential step required for marker-assisted selection (Mohan et al. 1997; Shehzad and Okuno 2014). The fundamental idea underlying QTL analysis is to associate genotype and phenotype in a population exhibiting a genetic variation (Broman and Sen 2009). Four steps of QTL mapping are (1) development a population, (2) genotyping the population using molecular markers, (3) phenotyping the population for an interested trait, and (4) QTL analysis using statistical procedures to find markers linked to the QTL (Bernardo 2002).

Populations used for genetic mapping can be a segregating population (F_2 and backcross) or a permanent population (double haploids or recombinant inbred lines). Recombinant inbred lines (RILs) are developed by selfing of individual progenies of the F_2 plants until homozygosity is achieved (F_7-F_8). An RIL population has advantages over segregating populations for many species because of homozygosity; the same RIL population can be used for mapping by different researchers, and linked genes have a greater probability of recombination (Alonso-Blanco et al. 1998; Burr and Burr 1991).

Earlier molecular markers used in genetic mapping are random-amplified polymorphic DNAs (RAPDs), restriction fragment length polymorphisms (RFLPs), and PCR-based DNA markers such as microsatellite or simple sequence repeat polymorphism (SSRs), sequence characterized amplified regions (SCARs), sequence tagged sites (STS) and inter-simple sequence repeat amplification (ISA), amplified fragment length

polymorphic DNAs (AFLPs), and amplicon length polymorphisms (ALPs) (Mohan et al. 1997). However, those marker procedures also have several drawbacks related to safety, being time-consuming, laborious, and the limitation of marker numbers per unit of DNA. Presently, the single-nucleotide polymorphisms (SNPs) technique is a preferred marker system because of the abundance of polymorphisms in plant genomes, and next generation sequencing has simplified the discovery of SNPs (Rafalski 2002).

Several statistical methods have been developed for QTL mapping based on regression, maximum likelihood, and Bayesian models (Broman and Sen 2009; Wang et al. 2012). The simplest model based on marker regression method is single marker analysis; interval mapping is based on maximum likelihood parameter estimation and provides a likelihood ratio test for QTL position. Composite interval mapping increases the precision of QTL detection by combining single marker analysis with multiple marker regression analysis (Broman and Sen 2009; Wang et al. 2012). Wang et al. (2012) proposed a statistical method called inclusive composite interval mapping based on additivity of QTL effects on the phenotype of a trait of interest. In the interval mapping for additive mapping, the phenotypic values are adjusted by all markers retained in the regression equation except the two markers flanking the current interval (Wang et al. 2012). Tools for QTL analysis have been developed and released for researchers such as R/qtl, QTL cartographer, MapQTL, and WebQTL. Recently, Wang et al. (2012) developed a free software for QTL mapping called QTL IciMapping which constructs genetic linkage maps and QTL analysis by simple interval mapping and inclusive composite interval mapping. QTL IciMapping is available for segregating and inbred

populations and nested association mapping populations. Unlike R/qtl, QTL IciMapping is not available for Unix/Linux. Running QTL IciMapping using a basic computer for the numbers of SNPs identified from genotype-by-sequencing (GBS) is time-consuming.

Genotype-by-Sequencing (GBS)

Recently, the cost of a whole-genome sequencing has been dramatically decreased because of the remarkable development of next generation technologies (Elshire et al. 2011). Genotype-by-sequencing (GBS) is a next-generation sequencing (NGS) based method which is robust, cost-effective, and feasible for various species (Elshire et al. 2011; Glaubitz et al. 2014; He et al. 2014). GBS generates large numbers of SNPs that can be used as high-density genotypic information for QTL mapping (Beissinger et al. 2013). A GBS scheme prepared by Poland and Rife (2012) is presented in Fig. 3. The GBS procedure uses restriction enzymes to reduce genome complexity resulting digested DNA then the digested DNA are duplicated using PCR and are ligated with unique barcoded adaptors (GBS libraries). The GBS libraries are sequenced using NGS resulting sequenced reads in raw FASTQ files (raw GBS sequence data) for GBS analysis. GBS analysis includes alignment sequenced reads to the reference genome and identifying SNPs. Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) is a software widely used by many researchers for GBS analysis (Bradbury et al. 2007; Glaubitz et al. 2014). TESSEL-GBS pipeline is presented in Fig. 4 and the documentation for TASSEL-GBS pipeline is available at https://bytebucket.org/tasseladmin/tassel-5-source/wiki/docs/TasselPipelineGBS.pdf. The FASTQ files are the input files for GBS

analysis. Finally, HapMap files are the output of genotypes as identified SNPs. GBS is a desirable method for genotyping. However, GBS is based on NGS, the limitation of NGS is the big data processing. Complex bioinformatics analysis for the amount of data requires high-performance computing. Most of the software for NGS analyses is written to run in a Unix/Linux environment. Many researchers unfamiliar with the Unix command line may be unable to use these tools (Bodi 2011).

QTL mapping of bioenergy-related traits in sweet sorghum

QTL mapping in sorghum based on DNA markers began in the early 1990s (Mace et al. 2009; Madhusudhana 2014). Madhusudhana (2014) reviewed 57 QTL mappings of sorghum published in 1990-2013 and presented that various DNA markers such as RFLPs, AFLPs, RAPDs, SSRs, and SNPs were used for F_2 and RIL populations of the size range 55-370 individuals. Genetic maps of sorghum have been constructed using markers in the range of 37-3418 markers, and map length ranged 283->2750 cM (Madhusudhana 2014). A sorghum consensus map was constructed by Mace et al. (2009), which consisted of 1190 DArTs (Diversity Array Technology) and 839 Non-DArT markers spanning for ten linkage groups of 1,603.5 cM with an average marker density of one marker per 0.79 cM. Recently, Gelli et al. (2016) constructed a 1614 cM linkage map using 131 RILs and 642 SNPs identified from GBS in the QTL mapping for nitrogen use efficiency.

QTLs for bioenergy-related traits in sweet sorghum have been reported, including in th development of plant height (Brown et al. 2008; Guan et al. 2011; Lu et al. 2011;

Madhusudhana and Patil 2013; Murray et al. 2008a; Nagaraja Reddy et al. 2013; Ritter et al. 2007a; Shiringani et al. 2010), grain yield (Nagaraja Reddy et al. 2013; Ritter et al. 2007a; Upadhyaya et al. 2013), biomass (Murray et al. 2008a; Ritter et al. 2007a; Shiringani and Friedt 2011), sugar content (Guan et al. 2011; Murray et al. 2008a; Ritter et al. 2008; Shiringani et al. 2010), flowering time (El Mannai et al. 2011; Higgins et al. 2014; Ritter et al. 2007a; Shiringani et al. 2010; Takai et al. 2012), and stem diameter (Lu et al. 2011; Shiringani et al. 2010). However, most of these QTL mappings were constructed with low-density markers as mentioned above. To improve the precision of QTL analysis, a high density of markers is required. The combination of GBS with a well-defined reference genome will improve the genetic maps (Poland and Rife 2012). Moreover, new bioenergy-related traits of sweet sorghum should be considered for QTL analysis. Carbon and nitrogen partitioning and C:N ratio might be interesting traits related to the biomass accumulation in each part of sweet sorghum. Increasing production of all biochemical components by increasing assimilation of carbon per unit of nitrogen would increase biomass (Lawlor 2002).

Carbon and nitrogen portioning and C:N ratio in sweet sorghum

Sorghum is an efficiently productive C_4 photosynthetic grass, which is well adapted to high temperatures and drought (Dweikat 2012; Stefaniak and Rooney 2013). C_4 plants have the ability to use light, water, and nitrogen effectively for converting carbon dioxide into sugars and organic compounds or biomass (Byrt et al. 2011). Among many requirements, carbon (C) and nitrogen (N) are essential for plant growth and

development. Carbon is fixed by leaves for photosynthesis, while nitrogen, minerals, and water come from roots (Braun and Slewinski 2009; Peoples and Gifford 1997). Carbon and nitrogen partitioning is the process of assimilation and distribution throughout each plant part (Braun and Slewinski 2009; Guo et al. 2015b). Increasing assimilation of carbon per unit of nitrogen increases biomass (Lawlor 2002). Plant biomass contains a relatively constant proportion by weight of carbon (38-46%) and nitrogen (1-7%) and varies by plant species, type, and growth stages of the plant (Pate and Layzell 1981). The relationship between carbon and nitrogen partitioning contributes to the determination of C:N ratio, which affects nutritional quality and increasing biomass (Lawlor 2002; Peoples and Gifford 1997). The C:N ratio is also important in the biological process. Decomposition rate of crop residues, which influences with crop nutrient cycling, is also related to the C:N ratio. Organic materials with a lower C:N ratio (high nitrogen) have higher decomposition rates than those with a higher C:N ratio (Norby et al. 2001). C:N ratio affects the hydrogen production from sucrose in biofuel production (Argun et al. 2008; Lin and Lay 2004). Thus, understanding processes of carbon and nitrogen partitioning are necessary for crop production and crop improvement. The mechanisms of assimilation, translocation, and utilization of carbon and nitrogen in grasses have been the focus of several studies from the production scale to functional genomics (Braun and Slewinski 2009; Irving 2015; Lawlor 2002; Peoples and Gifford 1997). In sweet sorghum, Singh et al. (2012) reported biomass partitioning of two sweet sorghum cultivars: M-81E and Dale. They found that stem accounted for 73% of the total biomass compared to leaves (only 13%), and nitrogen was accumulated in the stem and leaves 30

and 34% of the total nitrogen, respectively. Similarly, Zhao et al. (2012) reported biomass proportion of five sweet sorghum cultivars that stem dry weight was the highest proportion, which ranged from 56-73%, the proportion of the leaf and panicle dry weight were between 19-33% and 4-22%, respectively. Functional genomics of carbon transport and allocation to the stem has been studied in sweet sorghum. Sweet sorghum stores large amounts of sucrose in the stem while grain sorghum accumulates starch in the panicle (Bihmidine et al. 2016; Vietor and Miller 1990). Genes that control accumulation of sugar in sorghum have been characterized as sucrose transporter genes (SUT), tonoplast monosaccharide transporter genes (TST), and SWEET genes (SWEET), (Bihmidine et al. 2015; Bihmidine et al. 2016; Braun and Slewinski 2009). Recently, study findings indicated that SUT gene expression between sweet and grain sorghum are not different, but two TSTs are expressed in different ways (Bihmidine et al. 2015; Bihmidine et al. 2016). Although carbon partitioning and chemical composition of sweet sorghum have been documented, comparison of carbon and nitrogen partitioning and C:N ratio between different sugar content varieties have not been specifically reported. The variation of C:N ratio among different genotypes of sorghum should be determined.

Objectives

The overall goal of the present study was to identify QTLs associated with bioenergy-related traits in sweet sorghum x grain sorghum RIL population using GBS, and to determine the variation of new bioenergy-related traits for QTL analysis. The experiment consisted of two parts.