IDENTIFICATION OF TWO INTERACTING QUANTITATIVE TRAIT LOCI CONTROLLING FOR CONDENSED TANNIN IN SORGHUM GRAIN AND GRAIN QUALITY ANALYSIS OF A SORGHUM DIVERSE COLLECTION

By

WENWEN XIANG

B.S., China Agricultural University, 2007

A THESIS

Submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Jianming Yu

Abstract

Tannin, a second metabolic product in sorghum, has been directly related to resistance to insects and birds. Tannin also impacts sorghum nutritional value. Previous studies have shown tannin content has a positive correlation with early season cold tolerance, an important agronomic trait. Sorghum contains condensed tannins in testa layer below the pericarp. The testa layer tannin is controlled by two complementary genes B_1 and B_2: tannins are present when both genes are dominant but absent when only one or none of these two is dominant. The purpose of this research is to identify and map QTLs associated with the presence of condensed tannins, analyze interaction of QTLs, and provide a potential path to dissect the more complex trait of early season cold tolerance in future studies. A population of 109 $F_{6:7}$ recombinant inbred lines (RILs) developed from the cross of a high tannin sorghum Shan Qui Red (SQR) and non-tannin line Tx430 was used in the mapping study. Two QTLs related to condense tannin presence in testa layer were mapped to chromosome 2 and 4, respectively. Strong epistatic interaction of these two QTLs was detected. The two QTLs together with their interaction explained 74% of the phenotypic variation.

Sorghum grain quality traits, including kernel size, kernel hardness, protein and starch content, are complex traits which are directly related to sorghum nutritional value and market value. Association mapping is a promising method for complex quantitative traits analysis and dissection in plant science. Sorghum grain quality trait association analysis research is purposed to analyze large amount of grain quality data based on a diversity panel. A sorghum *bicolor* panel of 300 lines including germplasm derived from sorghum conversion program and elite commercial lines were established and served as diversity population for the association study. Phenotypic data of grain quality traits were collected by single kernel characterization system

(SKCS) and near infrared reflectance spectroscopy (NIRS). Data analysis proved high diversity within the SB panel. A correlation between tannin presence and kernel hardness was also observed. Quality traits showed high consistence across years and environments.

Table of Contents

List of figures

List of Tables

Acknowledgment

Upon the completion of my Master study, research and thesis, I have many people to whom I want to express my sincere appreciation.

First of all, I would like to express my greatest appreciation to my major professor, Dr. Jianming Yu, for his academic guidance and support through my graduate study, and for his sharing of wisdom with me, both in research and life. My thesis project couldn't have been completed without the mentorship, patience, understanding, and friendship Dr. Yu has given me. His suggestions and advice makes me better understand the importance of time management, the meaning of commitment to research, and the way to make them come true.

I also want to thank all the professors who served as my committee members: Dr. Scott Bean and Dr. Tesfaye Tesso. Thanks to Dr. Bean for sharing his laboratory facilities at USDA-ARS GMPRC and technical support for sorghum grain quality traits data collection. Thanks to Dr. Tesso for his valuable advices and help.

I also want to express my appreciation for Dr. Jeffrey Pedersen at USDA-ARS. I am grateful for his assistance in NIR data collection.

Thanks to all the group members in Yu's Lab: Dr. Zhongwei Lin, Dr. Chengsong Zhu, Dr. Xianran Li, Dr. Yuye Wu, graduate student Sivakumar Sukumaran and Guannan Sun for their kindly sharing of lab skills and research ideas, and also enthusiasm, help, and friendship. Thanks to undergraduate research assistant Christopher Guerra and Spencer Lickteig for help on the bleach test and data collection. It's been good memories for the days we shared.

Also thanks to Brian Ioerger at USDA-ARS GMPRC for his tutoring of grain quality analysis assay. Also the kindness of the people working in sorghum breeding program: Kellan Kershner, Leah Miller and Dr. Newton Ochanda.

Thanks Dr. Dadong Zhang and Mr. Xiaochun Sun for their help for QTL and association software management.

Words alone cannot express the thanks I owe to my parents for their endless support, encouragement and love.

CHAPTER 1-LITERATUR REVIEW

Sorghum

Sorghum (*Sorghum bicolor* [L.] Moench) is a C4 monocot which originated from African dry areas and is ranked as the fifth cereal crop worldwide and third in United States (Sasaki and Antonio, 2009). With high yield ability and good water and heat stress tolerance, sorghum can be successfully cultivated in a very wide range of soil and weather types including stress-prone environments. It is widely planted in tropical, subtropical, and semi-arid regions of developing countries in Asia and Africa, and serves both as feed and food grain. In the United States, sorghum is mainly produced in the Great Plains region from South Dakota to Texas. The top five production states, Kansas, Texas, Nebraska, Oklahoma, and Missouri occupy 85% of the market proportion. In 2008, almost 3.4 million ha were planted in the United States, with 35% of these planted in Kansas (NASS 2008a, 2008b). Sorghum grown in United States is primarily used as feed grain for livestock of poultry, beef and pork industries and materials for ethanol production.

The rising market for increasing human consumption of sorghum is emerging based on use as a demanding of multiple nutrition and gluten-free cereal for people who have celiac disease and consequently can not consume wheat products (Schober et al., 2005). Recently, a unique metabolic product present in sorghum but not other common cereal crops, tannin, was shown to be a good antioxidant. Antioxidants have been proved to have anticancer, anti-inflammatory, and antimicrobial effect, also helpful for therapy of coronary heart disease. With cancer, cardiovascular disorders, and arthritis increasing, sorghum's role as antioxidant source is drawing more attention. Compared to other candidates or traditional plant source, the tannins

obtained from sorghum have merits of large amount and high level of antioxidant activity (Awika et al., 2004).

Sorghum Tannin

Chemical structure and research about sorghum tannin

Compared to other cereal crops, sorghum has a unique chemical component tannin. Tannin sorghums include type II sorghum (tannins present in pigmented testa) and type III sorghum (tannins present in pigmented testa and pericarp), while non-tannin sorghum is classified as type I. Tannin is a plant chemical compound accumulated in sorghum during phenolic metabolism via the phenylpropanoid pathway, which also produces other phenolic compounds such as isoflavones, coumarins, lignins and aromatic amino acids. Tannins are water soluble, with molecular weight ranges from 500 to 3000 and be devided into hydrolysable tannins and condensed tannins (Haslam et al., 1989). The former has a glucose central core esterified with hydroxyl groups and gallic acid residues, rarely present in plants or only present in low amounts. The latter are flavona-3-ol unit oligomers or polymers (epicatechin and catechin) with hydrolysis-insensitive carbon-carbon bonds, also called as proanthocyanidin (compounds that yield anthocyanidin pigments upon oxidative cleavage in hot alcohols), and commonly exist in plants (Kaluza et al., 1980). The typical molecular structures are shown as in Figure 1.1.

Until now, all sorghum germplasm with a testa layer have been found contain condensed tannin, whereas no condensed tannins were found in anatomical structures other than testa layers (Awika et al., 2004). Condensed tannin is easily confused with tannin acid. Tannin acid is commonly used in the leather tanning industry and is said to be toxic to animals and human beings. Condensed tannins are widely distributed in the plant kingdom, including both

gymnosperms and angiosperms. Within angiosperms, they are more common in dicotyledons than in monocotyledons. Plants organisms such seeds, leaves, buds, stems, roots, legumes, and fruits have different amount of tannin. Because of the proximity of aromatic rings and hydroxyl groups and their high molecular weights, condensed tannins are highly effective at quenching free peroxyl radicals and show distinguished antioxidant effect (Waterman et al., 1994; Hagerman et al., 1998).

Almost all wild sorghums have condensed tannin, which is colorless, located at the innermost layer of the seed coat, testa layer. In model species such as Arabidopsis, the testa layer is termed the endothelium. About 10 days after fertilization, the sorghum seed has the color of the underlying endosperm. But at maturity, after seed dry, condensed tannins are oxidized and the testa becomes opaque. (Figure1.2) The United States Department of Agriculture's Federal Grain Inspection Service (USDA-FGIS) classified sorghum into yellow, white, brown and mixed class based on the grain color and tannin content. Sorghums with pigmented testa containing tannin are classified as brown sorghum, even though the pericarp color could be white, yellow, or red. Almost 99% sorghum marketed in U.S. is non-tannin sorghum (Awika and Rooney, 2004). Commercial producers dislike tannin sorghums because their negative effects on feed quality.

First, tannin leads to an unwelcome astringent, puckery flavor in mouth after consumption, the same unfavorable taste as wines and unripe fruits. Second, tannin can bind and produce complexes with various molecules such as amino acids, protein, carbohydrates, polysaccharides, and enzymes involved in protein and carbohydrates digestion. Such binding could precipitate proteins and the other nutritional molecules, produce compounds which are more resistant to digestion in the stomach, and decrease digestive enzyme and digestion rate

activity (Kumar et al., 1984; Salunkhe et al., 1990, Bramel-Cox et al., 1990). It also chelates minerals and metal ions at low pH (Hagerman et al., 1997). As a result, tannins in sorghum reduces food nutritional and caloric availability, decreases livestock feed intake, and leads to poor feed efficiency. Third, previous studies also have shown that tannins may be harmful and destructive to animal and health. Observations such as decreased growth rate of mice and other laboratory animals (Mehansho et al., 1985), reduced poultry egg production (Sell and Rogler, 1984), abased weight gain, decreased nitrogen retention and amino acid availability in rats (Deshpande et al., 1986), and caused abnormal leg bone development growth in chicks (Armstong et al., 1973).

However, recent research evidence strongly indicates the benefit of sorghum tannin to human health: people with type II diabetes and obesity are suggested to consume tannin sorghum, because tannin protein complexes limit digestion rate, release monosaccharides slowly, effectively control blood sugar concentration and caloric availability. Some African farmers prefer tannin sorghum as cereal because it ‘remains in the stomach longer’, and they can do field work a whole day without feeling hungry (Awika and Rooney, 2004).

More importantly, sorghum tannins have recently been shown to be good antioxidants (Zhang et al., 2008). Antioxidants are chemicals which convert free radicals in the human body to harmless molecules by donating electrons. They protect cells against oxidative damage leading to aging, arthritis and cancer. Antioxidants also prevent injury to blood vessel membranes, and optimize blood flow to the heart and brain, which as a result helps lower the risk of cardiovascular disease (Awika and Rooney, 2004) and dementia, including Alzheimer's disease. According to epidemiological evidence derived from previous studies, phytochemicals such as phenolics all have antioxidant effects. The ability to act as antioxidants depends on

extended conjugation, number and arrangement of phenolic substitutents, and molecular weight and polymerization degree of flavonoid oligomers. With molecular weight of greater than five hundred, polymerization and many phenolic hydroxyl groups, tannins turn out to be 15-30 times more effective antioxidants than simple phenolics (Hagerman et al., 1998). (Table 1.1) At the same time, tannins have little or no pro-oxidant activity, although many small phenolics are pro-oxidants (Labieniec and Gabryelak, 2003). Also, because tannin forms strong complexes with protein, carbohydrates, and lipids, they can help avoid oxidative damage during digestion. The potential of tannins to diminish nutrient digestibility must be balanced against their potential to serve as biological antioxidants (Dykes et al., 2006). Furthermore, compared to plant-derived tannin, sorghum tannins have extremely higher antioxidant activity, which consolidate its status as strong dietary antioxidants (Awika et al., 2004). Comparison with other plants is showed in Table 1.1.

Tannins also have favorable agronomic traits since they validate tannins sorghum a good defense mechanism against insects and birds, prevent sorghum from pre-harvest germination, molding, and diseases caused by fungi, bacteria and viruses (Friend et al, 1977; Hahn et al., 1983; Harris et al., 1970; Harris et al., 1973). Pervious studies also showed that a positive correlation exists between tannin and other phenolic compound presence and sorghum early season cold tolerance. Phenolic compounds in the seed, particularly condensed tannin in the seed coat, may contribute to emergence and seedling vigor by suppressing soil-borne pathogens which are often present in cold, wet soils. With a RIL population with 153 lines derived from Shan Qui Red (SQR, cold tolerant) and SRN39 (cold sensitive), phenolic compounds in the seed were found to be positively correlated with seedling emergence, seedling vigor, seedling height, and germination at 22°C (Cisse et al., 1995). The concentration of flavan-4-ols, tannins, and total

phenols in the seed were associated with several markers on a linkage group. With the same population, a later study found that both the same set of markers and additional markers related to coat color are located in a region showing strong association with seedling emergence and vigor (Knoll et al, 2008). Particularly, the presence of a testa layer is associated with seedling vigor, but not emergence.

Screening for Tannin Content

Tannins are complex compounds composed of various units with total molecular weight raged from 500 to 3000. This complexity means that accurate characterization and measurement is not easy to achieve. The presence of condensed tannin is associated with the presence of a testa layer, which is underneath the pericarp and originates from the ovule integuments. The testa layer of sorghum varieties can be absent or present, also can be partially present. Testa layers with condensed tannins are pigmented and present brown color.

From the very beginning, sorghum breeders have a misunderstanding for correlation of sorghum tannin content and the darkness of kernel colors. Red and brown sorghums were classified as tannin sorghum while white sorghums are thought to be non-tannin. However, the darkness is not a suitable indicator parameter for tannin content (Boren et al., 1992). Actually, kernel colors are determined and affected by many factors including 1) pericarp color 2) pericarp thickness 3) presence of pigment testa layer 4) endosperm color 5) endosperm texture 6) glum color and 7) environment. Pericarp colors are controlled by two genes, R and Y, epistastic interactions of these two produce three types of color: red (RRYY), yellow (rrYY), and colorless or yellow (RRyy and rryy). There is also another I gene, which serves as an intensifier, increasing the darkness of red and yellow pericarp. This genetic variation, together with

combination of thickness and color variation produced complex kernel colors which can not serve as evidence for tannin presence. Several red sorghums are non-tannin while several white sorghums are high in tannin content (Table 1.2). High tannin sorghums have a very wide range of seed color, while light-colored varieties might have high levels of tannin content (Waniska et.al, 1992). Qualitative and more accurate methods should be employed to determine the presence and amount of tannin.

A fast and direct method of tannin sorghum classification is the scratch test along with visual observations. With a knife scrapping the pericarp layer, the kernels with a black testa layer is classified as tannin sorghum. However, because the thickness of pericarp layer and testa layer vary in different sorghum accessions, ranging from 8 to 160 um and from 8 to 40 um, respectively, the standard of the scrape intensity should vary according to germplasm sources and accurate classification with the scratch test needs experience to do accurately.

The bleach test is a relatively accurate, inexpensive, and rapid method that can be used for sorghum kernel grading and classification. It is widely accepted as a standard tool to identify sorghum with tannins because its ability to screen several thousand single-head selection using a small amount of labor. It has been used by the United States Department of Agriculture's Federal Grain Inspection Service-Grain Inspection, Packers and Stockyard Administration (USDA-FGIS-GIPSA) from the 1980s to test samples during grading for the presence of tannin sorghum. An alkali/hot water extraction process which remove both the pericarp and testa layer was initiated in early 1970s (Blessin et al., 1971). Armstrong et al. (1974) found that the process effectively removed most of the tannin from sorghum grains. Sorghum grain is immersed in a 3.5% sodium hypochlorite solution (bleach) containing alkali KOH. The kernels were then rinsed with water and deposited on paper towel to dry. After dry out, the kernel colors were rated for

tannin or non-tannin. The solution dissolves away the outer testa layer of sorghum grain, revealing the presence of a black pigmented testa layer in the case of tannin sorghums, or its absence in the case of non-tannin sorghums.

Accuracy of bleach test directly related to the reagents, standards chosen, reaction time, and kernel deterioration levels (Waniska et al., 1992). KOH alone removed all the cell layers outside the endosperm. Bleach alone reacted with the compounds inside the kernels and showed preferential color differences among genotypes. Testa layer of different sorghum genotypes varies in thickness, intensity and color, so the bleach test works well on varieties with good accuracy, but can cause false positives at others. Bleach loses its strength over time, so fresh bleach should always be used for best results. Long treatment times dissolve the testa layers together with pericarp, while limited time may not remove the pericarp completely. Sorghum seeds that have been molded and weathered in the field prior to harvest without testa layer will turn dark or have some dark spots after bleach, lead to false-positives (Dykes et al., 2002; Waniska et al., 1992).

To overcome all the factors affects accuracy of bleach test, samples of tannin and non-tannin sorghum should always include as appropriate standard checks. Also the bleach time, concentration and temperature should be critically controlled. Over bleaching could give false positive results, while inefficient bleaching may fail to recognize tannin sorghum (Table 1.3). Kofoid et al. (1978) carried out bleach test with three concentrations of KOH: bleach (1:6, 1:5, and 1:4) at water bath temperatures (50, 60, and 70 C) for three time periods (10, 15, and 20 min), and reported significant effect for differences of genotypes time, and temperature. There was no single treatment combination that worked for all entries. False positive tannin sorghums do not occur routinely and causes only limited problems. With good knowledge of the bleach test and

carefulness, improper classification can be minimized (Awika et al., 2005). After years of technical modification, USDA-FCIS employs a standard mix jar to carry out the bleach test. Sorghum samples of 15 grams were place into a mixing jar, 15 grams of KOH pellets then added also with 40ml of standard commercial bleach. After 3 minutes mixture, the sorghum is rinsed with warm tap water to remove the KOH-bleach solution. Excess water is removed, and the sorghums are allowed to dry in a dish. Modified bleach test with appropriate tannin and non-tannin standard can be used for sample size reduced to 10 kernels. Tannin grains have black over the entire surface, with the only exception the germ which is lighter in color. Non-tannin grains are either completely white or partially brown.

Both the scratch test and bleach test are qualitative tests. Neither of them is designed to quantify tannin content. Proper confirmatory tannin analysis may be performed in addition to these tests. A colorimetric method, or vanillin/HCl assay, has been used to measure sorghum tannins which react with vanillin in the presence of HCl to give a bright red color. Catechin is used as a standard. However monomeric phenols also react with the reagents and yield false-positives. So, appropriate blanks are included and their values are subtracted to exclude background non-tannin favanol molecules which also have A-ring and a single bond between C2 and C3 reacting with vanillin (Earp et al., 1981). However, significant time is required for this test and for whole day measurements, and a standard curve must be run before sample measurement in the morning and again in the afternoon (Awika et. al, 2004). Slope of the lines should be determined by linear regression on the curve. A low regression value may be caused by old chemicals or unskilled technique. Also weathered grains have significantly higher vanillin-HCl blank readings than normal grains, indicating higher levels background pigmentation do produced which similar to bleach test result. (Awika et. al, 2004)

Methods based on the reducing power of phenolic hydroxyl groups, such as Prussian blue, folin-Denisand Folin-Ciocalteu procedures are also used to measure phenolic acids, flavonoids, and tannins. Because these different solvents and methods extract and measure different chemical parts of the tannin molecule, different tannin values could result. The acidic butanol method is specific for condensed tannin, but is interfered by other sorghum anthocyanidins. Also tannic acid is used as a reference compound in these methods which it is not present in sorghum. Protein precipitation and enzyme inhibition ability could also serve as measurement methods (Hahn et al, 1984), but can not differentiate hydrolyzed tannins and condensed tannin and are too sensitive to reaction conditions and the protein used. Normal-phase HPLC analysis with fluorescence detection efficiently separates tannins according to their degree of polymerization (Awika et al. 2003) Methods comparison is included in Table 1.4.

Tannin Genetics

The presence or absence of the testa layer is controlled by two complementary dominant genes B_1 and B_2. Normally, dominant alleles are needed in both genes to confer tannin presence. Condensed tannins are present when both loci are dominant ($B_1\_B_2\_$) and absent if either or both of the loci are recessive ($B_1\_b_2b_2$, $b_1b_1B_2\_$, or $b_1b_1b_2b_2$). According to classical literature, these two genes obey the rule of independent assortment and should locate at different sorghum chromosomes (Stephens, 1924). The segregation and recombination of these two genes is in accordance with Mendel's genetic laws. With the dominant spreader gene S, pigmented pericarp is present.

For complementary dominance of testa layer and tannin presence, a quantitative genetic model could be developed incorporating the gene action effects of each locus and the effects of

interactions between loci affecting the trait. The model is two-locus model in which each locus has two alleles (Loci B_1 with alleles B_1 and b_1, loci B_2 with alleles B_2 and b_2). Additive effect (a) is half the average difference between genotypic classes measured in other loci background. Dominant effect (d) is the difference between the heterozygote B_1b_1 and the average of the two homozygous genotypes at the B_1 locus (Also the same for B2 locus).

In recombinant inbred line (RIL) populations, the model becomes simple. Since all the genotypes in RIL are homozygous, there is no dominant effect. So, there is no dominant by additive effect, but only additive by additive effect (Table 1.4). In tannin RIL population, tannin is scored as 1 while non-tannin is scored as 0. So, the values of genetic effects are shown in Table 1.5.

As mentioned previously, presence of a testa layer has a positive correlation with cold tolerance. Previous cold tolerance QTL research based on RIL derived from SQR (cold tolerance high tannin kaoliang line) and SRN39 (cold sensitive African caudatum without tannin) identified two cold tolerance QTLs located on chromosome 2 and 4 (Knoll et al., 2008). Another study, based on a RIL population derived from the cross of SQR and Tx430 (non-tannin), identified two tannin QTLs at the regions close to the flanking markers of the cold tolerance QTLs on chromosome 2 and 4 (Ridder, 2004).

Testa layers begin to accumulate colorless condensed tannins from the very early stage of embryo development and turn brown upon oxidative reaction. It is identified in many plant species that condense tannin biosynthesis shares the common flavonoid pathway with the anthocyanins until after the flavan-3, 4-diol step, which is catalyzed by dihydroflavonol-4-reductase (DFR). In arabidopsis, condensed tannin also produced in a testa layer, which more commonly named as endothelium or seed coat (Debeaujon et al., 2003). BAN (BANYULS) gene

has been cloned and shown to be involved in synthesis branch between anthocyanin and condensed tannin (Nesi et al., 2001). The expression of BAN is restricted to the endothelium during early seed development (Devic et al., 1999). BAN either codes for an enzyme involved in the step converting leucocyanidin to catechin or an enzyme catalyzed the step converting catechin to epicatechin, which are leucoanthocyanidin reductase (LAR) and anthocyanin reductase (ANR) respectively. BAN protein is homolog to DFR, but only 20% identical to the LAR enzyme from Desmodium uncinatum (Abrahams et al., 2002), while the relationship of BAN and ANR is still not clear. The flavonoid biosynthesis pathways are shown in Figure 1.3.

Furthermore, in Arabidopsis, nearly 22 loci related to flavonoid compounds production pathways were found through the analysis of seed coat mutants: transparent testa (tt) mutants, banyuls (ban) mutant, and transparent testa glabral (ttg) mutants. Among these loci, TT_1, TT_2, TT_8, TT_{12}, TTG_1, and TTG_2 genes are commonly accepted as directly associated with condense tannin (Koornneef, 1990; Shirley et al., 1995; Zhang et al., 2003). TT8 encodes a helix-loop-helix and TT2 encode a R2R3 MYB protein which regulate BAN and DFR gene expression (Nesi et al., 2001). TT1 is a zinc finger gene (Sagasser et al., 2002). TT12 encode a transporter-like protein which is required for condensed tannin accumulation in vacuoles of testa layer (Mueller et al., 2000). TT_2, TT_8, TTG_1 and TTG_2 act together regulate BAN expression at the transcriptional level, and TTG2 acts downstream of TTG1 to enable tannin biosynthesis (Johnson et al., 2002).

TTG_1 ortholog in sorghum is located on chromosome 4. It is related to pale aleurone color 1 (PAC 1) locus in maize, which required for anthocyanin pigment in the ***aleurone*** and scutellum of maize seed and has similar effect of tt1 in arabidopsis. Consequently, TTG_1 could

be considered as a good candidate for one of the condense tannin complementary genes (Figure 1.4).

QTL Mapping

Molecular Markers and Mapping populations

Markers used for linkage map constructions include restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeat (SSR). SSR markers, also known as microsatellites, are consists of nucleotide repeats of 1-6 base pairs sequence in length. Because of merits such as codominant, high occurrence frequency, whole genome distribution, high polymorphism level, high reproducibility, and amenability to simple and inexpensive polymerase chain action (PCR) assays, SSR is widely used in various plant species.

To construct genetic linkage maps for QTL mapping, a suitable population is very important. The commonly used man-made populations include F2, backcross (BC_1), double haploid (DH) and recombinant inbred lines (RIL). These populations are all derived from a bi-parental cross and segregating for a desired trait. DH and RIL populations are preferred because of their genome stabilization, which facilitate multiple year and location experiments and guarantees accurate phenotypic value estimation. But time required for RIL population development is relatively long, even when winter nurseries in tropical areas are used as a time saving approach. Comparatively, DH population could be constructed within three years. However, since the recombination while DH production are limited to one or two cycle, its

genetic background is not so favorable compared with RIL which have undergo several round of recombination. There is no successful DH population produced for sorghum till now.

QTL Mapping Methods

Quantitative trait locus (QTL) mapping technologies are used to estimate the number of loci controlling genetic variation in a segregating population and to determine the map positions of these loci in the genome. QTL analysis methods based on statistics have been developing for years. Single marker analysis based on the likelihood ratio is the most original and simplest one. However, it can only be used to detect QTL effect without estimate QTL location on the chromosome. Interval mapping (IM) method based logarithm of the odds (LOD) calculation was employed to detect both the QTL effect and location in every two markers interval (Lander and Botstein, 1989). Where the maximized LOD score exceeds the threshold value is determined to be the most possible location of an objective QTL. Ghost QTLs are potential problems because IM could not eliminate the effects of other QTLs in nearby chromosome regions**.** Zeng et al. (1994) modified interval mapping and published composite interval mapping (CIM), which separates makers by a detecting window. Markers outside the detecting window serve as cofactors to remove the effect from other QTLs in the genome. Minor effects QTLs could also be detected using this method. Both IM and CIM are likelihood-based approach and require a suitable threshold to declare a QTL with significant effect. The permutation test has been widely accepted for threshold value determination. This test assigns trait data randomly to the individuals of certain populations repeatedly, and selects the best data set for threshold set up

(Churchill and Doerge, 1994). Zeng et al. (1999) developed the multiple intervals mapping (MIM) method that fits a multiple QTL model including epistasis. It provides the benefit of simultaneously searching the number, positions and interaction of QTLs. Compared to CIM, MIM has merits of improving statistical power for multiple QTL detection, facilitation for QTL epistasis analysis.

Genetic Maps and QTL Analysis in Sorghum

Sorghum is a diploid with haploid chromosome number of 10 (2n=2x=20). The physical size of the sorghum genome is 730 Mb (Paterson et al., 2009). This makes the sorghum genome the second smallest genome of cereal crops, about 60% larger than rice, but only one-fourth the size of the genomes of maize or human. Sorghum is closely related to major crops of tropical origin such as maize, sugarcane, and pearl millet, and thus provides a better roadmap for study of these crops at the DNA level. In 1990s, genome mapping of sorghum began and several genetic maps of sorghum were constructed. The initial marker used for map construction was restriction fragment length polymorphism (RFLP) markers. Amplified fragment length polymorphisms (AFLP) markers which have highly polymorphic information content, great multiplex ratio, ease of automation use, and the ability to develop large number of markers are also developed in sorghum (Menz et al. 2002). Simple sequence repeat (SSR) or microsatellite markers were integrated into genetic maps thereafter (Bhattramakki et al., 2000).

Several QTLs for important agronomic traits in sorghum have been mapped using molecular markers, including disease resistance, tolerance to environmental stress, leaf phenotypic traits, seed and panicle characteristics, and plant status (height, flowering, maturity, tillering) (Subudhi et al., 2000).

Tannin QTL Analysis in Other Plant Species

Condense tannins in seeds attract research attention not only in sorghum but also other valuable model plants and economical crops including Arabidopsis, canola, tea and the common bean (Wittkop et al., 2006). Using AFLP and SSR markers, Lipsa et al. (2009) identified 14 QTLs for condensed tannin in canola seed coat (similar to the testa layer in sorghum) based on a double haploid population. In this research, 10 QTLs were in charge of oligmeric condensed tannin and 4 QTLs were associated with polymeric condensed tannin. Caldas and Blair (2009) identified condensed tannin QTL in the common bean based on three RIL populations. They found that seed tannin concentration in the common bean associated with Mendelian genes for seed coat color and pattern. Seed coat condensed tannins were determined with a butanol-HCl method and a total of 12 QTL were identified on separate linkage groups in the three RIL populations with individual QTL explaining from 10 to 64% of the phenotypic variation for this trait. Loci on linkage groups B3 and B10 were associated with the Mendelian genes Z and Bip for partly colored seed coat pattern, while a QTL on linkage group B7 was associated with the P gene which is the primary locus for the control of color expression in beans. Another study in common bean using AFLP markers also found one putative QTL associated with tannin content in the common bean seed which explain 42% of the phenotypic variance (Caldas et al. 2000). There are also tannin QTL identification researches for leaf tannin (Doyle et al., 1987). Tanaka et al. (2000) employed RAPD markers in a two clone derived population, and find one QTL effect tannin biosynthesis or degradation pathway in tea.

Sorghum Conversion Program and Sorghum Diversity Panel

As the very early domesticated sorghum plants were selected and dispersed, genetic adaptation and intercrossing followed by selection and continued intercrossing in isolated ecosystems gave rise to new and stable sorghum biotypes. However, strong selection for domestication-related traits created a severe genetic bottleneck and reduced diversity in domesticated sorghum compared with wild relatives (Hyten et al., 2006). In addition, modern breeding practices have further constrained the amount of extant diversity in crop species and limited genetic gains in breeding programs. Most sorghum commercial varieties grown in United States are less than 5 feet height, named dwarf types, which facilitate modern mechanical harvest. Diversity shortage and germplasm consistency makes U.S. commercial sorghum lines susceptible to disease, pests and other environmental stress.

Simultaneously, sorghum has a large range of genetic variability available in Africa where domestication first occurred. Due to the early introduction of sorghum to Asia, further diversity could be found in Asia. Landraces and wild relatives of cultivated sorghum from Africa and Asia are rich sources of resistance to stresses such as high temperature, drought, insects, and diseases. They may also provide a source for food, feed, and industrial products quality trait improvement.

Around the 1960s, CMS A1 was used extensively for sorghum hybrid seed production, at the same time the tropical originated, photoperiod sensitive accessions were cultivated in U.S. temperate regions. Collection and conservation of sorghum germplasm should be accelerated to prevent the extinction of landraces and wild relatives of cultivated sorghum. To address this concern, in 1963, the USDA in cooperation with Texas A&M University initiated sorghum conversion program (SCP), which introduced novel genetic variation from exotic, tropical germplasm into modern U.S. cultivars (Stephens et al., 1967). After reciprocally crossing

tropical lines temperate and elite line, selecting day-neutral flowering and reduced height progenies, tropical lines were converted to photoperiod insensitive line with early maturing and short stature characters. Backcrossing of the progenies to the tropical parents were carried out until the resultant lines were fixed for temperate alleles at major maturity and height controlling loci but with nearly 90% of the tropical genome (Lin et al., 1995). About 850 converted tropical lines have been released by the SCP and these germplasm have allowed breeders to exploit novel variation for insect and disease resistance, drought tolerance, heterosis, and grain quality. As a result, most of the U.S. sorghum hybrids grown today have some tropical germplasm in their pedigrees (Gabriel, 2005).

A sorghum diversity panel was recently assembled. This panel is comprised of 377 accessions, including 228 converted tropical lines produced by the SCP, and 149 important breeding lines and their progenitors from all around the U.S. A subset of 300 lines from this panel was selected for sorghum grain quality traits research by association analysis (Hamblin et al., 2007).

Sorghum Grain Quality

Definition of grain quality depends on the grain type and its end use. It includes a range of properties that can be defined in terms of physical, sanitary, and intrinsic characteristics. Physical characters include moisture content, kernel weight, kernel size, total damaged kernels, and broken kernels. Grain quality is also related to fungi count, insects and insect fragments, rodent excrements, foreign material, toxic seeds, pesticide residue, odor and dust. Oil content, protein content, hardness, density, and starch content are classified as intrinsic characteristics (Henry and Ketlewell, 2007). The quality properties of a grain are affected by its genetic traits,

the growing period, timing of harvest, grain harvesting and handling equipment, drying system, storage management practices, and transportation procedures (Mazur et al., 1999).

Grain hardness is an important attribute in the processing of cereal grains and in the end-use products such as breads and snack foods (Bettge and Morris, 2000). It also plays a role in plant defense against molds (Jambunathan et al., 1992), weathering, and insect attack (Waniska, 2000). For sorghum, hardness is reported to be significantly related to cooking quality parameters such as adhesion, cooked grain texture, alkali gel stiffness (Cagampang et al., 1984), porridge quality (Akingbala and Rooney, 1987), and production of high-quality couscous granules (Aboubacar and Hamaker, 1999). Milling quality of sorghum grain has been related to grain hardness as well (Rooney and Waniska, 2000). Commonly, large sorghum kernels are harder than small ones and related to higher quality grain (Lee et al., 2002).

Sorghum kernels are round, small in size; vary from about 3 to 4mm in diameter. Variation in kernel size occurs not only between cultivars but within a cultivar obtained from a different location or season (Wills and Ali, 1983). Large sorghum kernels with corneous endosperm are usually preferred for human consumption and associated with desirable physical and chemical quality parameters: harder, higher in protein concentration, and lower ash, higher milling yields, higher water absorbance flour, brighter white color, and larger particle size (Lee et. al, 2002). Small kernel sorghum, which has a greater possibility to be hard and not easy for milling, is also not popular in the grain marker because of these qualities.

Sorghum kernel weight is determined by kernel growth rate and total duration of grain filling, also related to grow position within the sorghum panicle (Gabriel et al, 2005; Buffo et al., 1998). Sorghum kernel weight contributes greatly to final yield determination. The two

components for weight: kernel moisture content and kernel density are correlated with milling value.

Sorghum kernel color varies from dull white, yellow, and brown to red, which is also a big component for sorghum grain quality. Because usually the seed with a red coat has a good chance of high tannin content which is not good for food and feed use, light color are more preferred. Lighter flour is more favorable in markets.

Chemical quality parameters such as protein, starch and mineral contents are directly factors play roles in sorghum nutritional value, which are certainly consider as important grain quality components.

Grain Quality Analysis

The Single Kernel Characterization System (SKCS) is a device for whole grain physical properties measurements originally used in wheat analysis. After modification and calibration of parameter detection, adjusting new slope and bias values for kernel rejection criteria, SKCS has been employed to provide countable and accurate parameters for sorghum kernel parameters (Bean et al., 2006). Individual grains are crushed between a serrated rotor and a crescent, and parameters for kernel hardness, diameter, moisture and weight were sent to computer at the same time (Martin et al 1993). Typically 300 kernels are analyzed for each sample, and both the average and the standard deviations for the 300 kernels are reported. Moisture content obtained by SKCS for sorghum is skewed, less than air oven measurement, and not suitable for accurate usage. The SKCS accurately predicted weight relative to weight kernels on an analytical balance while predictions of kernel thickness were highly correlated to digital caliper measurements, with only 20% underestimated value (Bean et al, 2006).

Near infrared reflectance spectroscopy (NIRS) is a spectroscopic method utilizing the near infrared region of the electromagnetic spectrum (from about 800 nm to 2500 nm) to acquire the goal of rapid and nondestructive determination of the concentration of physical and chemical constituents in agricultural materials. Grain quality traits, such as protein content, moisture, starch could be predicted by NIRS. NIRS is based on molecular overtone and combination bands vibrations. The hydrogen containing bonds (O-H, C-H, N-H, S-H, P-H) and hetero-nuclear bonds such as C=O are abundant in nutritional molecules and have high anharmonicity and strong overtone absorptions. In addition to the chemistry of a material, near-infrared spectra are also influenced by the physical structure of a material. The size and shape of the particles, the void between particles and the arrangement of particles affect the length of light transmission passing through a sample and thereby influence reflectance. Near-infrared spectra are difficult to interpret directly because the molecular overtone and combination bands seen in the NIRS are typically very broad, which leads to complex spectra and increases the difficulty of assign specific features to specific chemical components. As a result, multivariate calibration is required for quantitative analysis of sample constituents by NIRS.

Multiple wavelength calibration techniques are often employed to extract the desired chemical information. These calibration methods include principal components analysis, partial least squares, step wise multiple linear regression, Fourier regression, locally weighted regression are the most used ones. None of these proposed calibration techniques have achieved universal acceptance because calibration model that works well for one application may be unacceptable for another. So, careful development of a set of calibration samples and application of multivariate calibration techniques is essential for near infrared analytical methods.

Association Mapping and Candidate Gene Approach

Grain quality traits are complex traits controlled by quantitative traits loci. QTL mapping developed for 20 years could map QTL within 10 to 30cM (Salvi et al., 2005). Near isogenic lines (NIL) are then used to mendalize the trait, which transfers quantitative data into qualitative data and lead to QTL positional clone (Doerge, 2002). However, plants with large genome with large amounts of repeat sequences make the fine mapping and positional clone problematic. Furthermore, the extent of QTL effect directly controls the mapping efficiency, no small effect but only main-effect QTLs were cloned. Association mapping, also called linkage disequilibrium mapping, is another approach for quantitative trait locus discovery and dissection. Compared with linkage analysis, association have four merits 1) time saving, the mapping population are always natural populations, which save the time for population construction 2) high throughput, alleles within the same loci can be analyzed at the same time, 3) high extension, identification could be based on single gene level, 4) detect more QTLs than linkage mapping and even QTLs with minor effects (Flint-Garcia et al., 2005).

Association mapping is based on linkage disequilibrium, the random combination of alleles at different locus. If the rate of certain allele company with another allele of other loci on the same chromosome is larger than rate of two alleles show up together after random segregation, the two alleles is said to be in Linkage disequilibrium. Linkage disequilibrium (LD) is derived from polymorphism produced by mutation, and break by recombination (Hamblin et

al., 2005). Biological and historical factors, the outcross rate of species, chromosome location, population size, selection intensity of chromosome segments, and gene drifting, also affect LD extension and distribution. Self-pollinated species have high level of LD. Regions near the centromere have low recombination rate and high LD, while the regions on chromosome arms show relatively higher recombination rate and low LD. The population subdivision and mixture increase LD. Selection on specific genes will at the same time reduce diversity of this gene loci and the locus nearby, as a result increasing LD.(Wang et al., 2005)

Association analysis could be employed by two approaches, the whole genome approach and the candidate genes approach. The former scans the whole genome to search for potential mutant locus explaining phenotypic variation. For the later, discovery of candidate genes related to target traits with assistances of bioinformatics and biochemistry methods is critical. Both approaches request large amount of statistical analysis and computation (Falush et al, 2003).

Whole genome association mapping, strictly in concept, need thousands of markers and population with as many unrelated individuals as possible. It is only practically feasible when large amount of research funding available. The candidate gene approach can be combined with QTL analysis (Krrakman et al., 2004). If one gene locates within a certain QTL region, and its function is associated with phenotypic traits, this gene is probably a candidate gene of the QTL. If the sequence of the research species is already available, linkage analysis could locate objective QTL within 3 to 5 cM, then through employing functional prediction bioinformatics and relative physiological and biochemical analysis, most genes inside the objective region could be excluded. By association analysis of several candidate genes, the final objective gene is easily determined (Meuwissen et al., 2005). Furthermore, traditional QTL linkage analysis can not discover alleles which are not present in either of the parental lines, or alleles without

polymorphism between parental lines, and limits the number of QTLs detected. Candidate gene association analysis based on natural populations with great diversity and large genetic variations could overcome this problem. In all, candidate gene approach shortens the mapping time and decreases the resources required for gene identification (Doebley et al., 2005).

Candidate gene association analysis also plays an important role in gene function verification (Doerge, 2002). Though transformation is widely accepted as gold standard for gene function verification, it is difficult for quantitative genes which only control one step of complex metabolic pathways. For example, carotenoid production in gold rice is controlled by four genes together. Only when all the four genes were transferred into rice genome together, the rice would present the golden color. Without clear dissection of metabolic pathway, functional verification of such genes is unavailable. Association analysis could avoid such drawbacks.

Candidate gene association analysis could also promote the development of marker assistant breeding by detection of functional markers. Recently, makers used in MAS are developed from linkage mapping results, and proved to be tightly linked to objective genes. However, linkage mapping results are only based on particular population constructed by researchers, which might not discover the best allele, consequently the best selection result might not be achieved using MAS. Furthermore, recombination and genetic drift could lead to objective gene lose. A functional marker which is derived directly from gene coding region could solve such problems (Doerge, 2002). Exploiting a functional marker requires 1) candidate genes with known function and alleles sequence information; 2) availability of traits investigation, sequence comparison and analysis in multiple populations with different background using association analysis. With more SNP markers available in plant model species

and rapid development of bioinformatics and biochemistry, candidate gene association mapping will becomes hot point of plant genome research quickly.

Metabolic Pathways and Candidate Genes

Sorghum grain quality traits are determined by different metabolic pathways, the learning of the pathway and relative enzyme is very critical for trait dissection by association mapping candidate gene approach. Until now, fifteen candidate genes in starch metabolism pathway have already been sequenced. They influence synthesis or regulation of ADP-glucose pyrophosphorylase, debranching enzyme, starch synthesis, granule bound starch synthesis, sucrose synthesis, phosphoglucomutase, and Glucose-6-phosphate translocator. More candidate genes for other metabolic pathways related to grain quality traits is on the way of discovery.

Figure 1.1 Molecular structures of tannin related compound. From left to right, top to down: gallic acid, epicatechin, catechin, hydrolysable tannin, condensed tannin, and tannin acid.

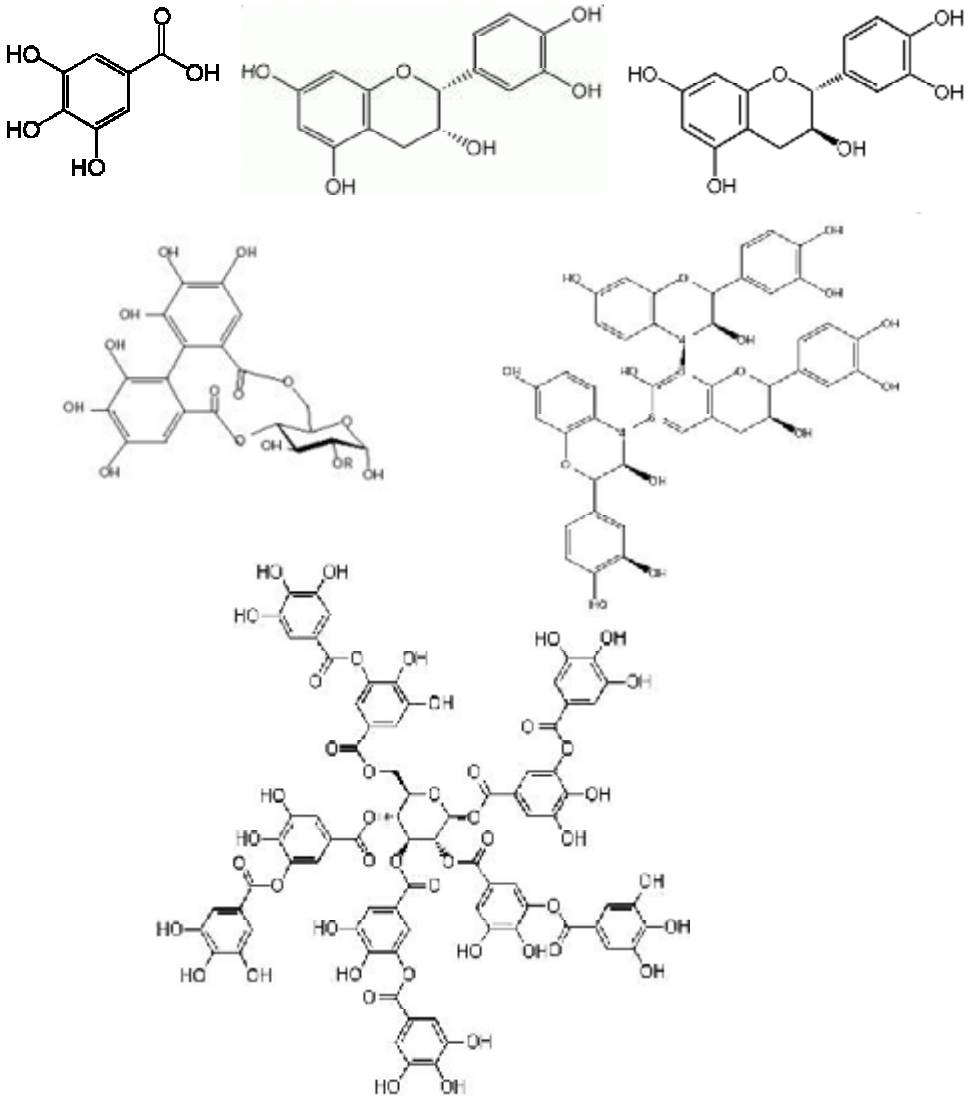


Table 1.1 Antioxidant activity (ORAC) levels of tannin sorghum compared to common fruits (Adapted from Awika et al. 2004)

Plant	ORAC
Tannin sorghum bran	2400-3100
Blueberries	87-870
Strawberries	356-400
Plums	452-600
Grapes	100
Watermelon	15
Orange	80-150

Figure 1.2 Fluorescence photomicrograph of cross-sections of a non-tannin (left) and a tannin sorghum kernel (right). Ep- epicarp; M- mesocarp; CW- cell wall; En- endocarp; Al- aleurone; E- endosperm cell. (Adapted from Earp et al., 2004)

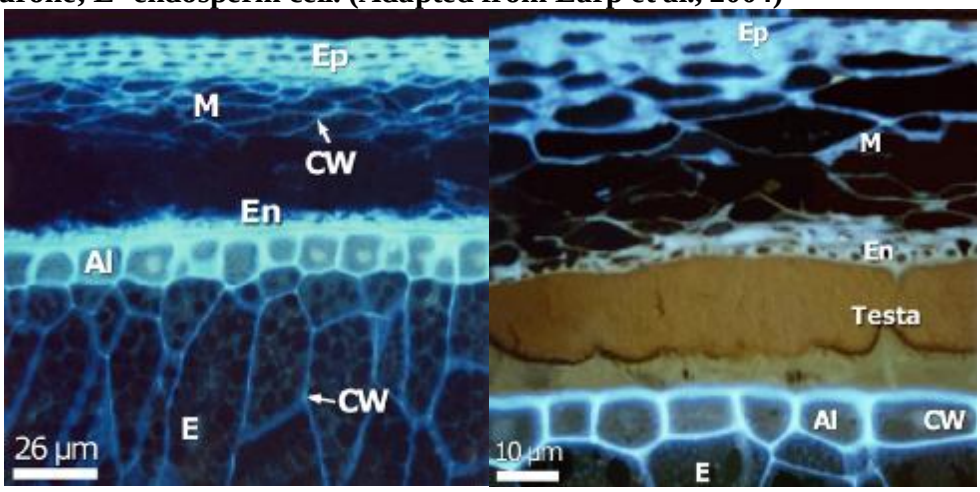


Table 1.2 Kernel characteristics, their genotypes, grouping and tannin content. (Adapted from Earp et al., 1981).

Entry	Source	Pericarp color	Testa[a]	Testa color[b]	Group[c]	Tannin (%[d])	Pericarp genotype[c]	Testa genotype[c]
Jinfu No1	China	Red	A	–	I (S)	0.2	RRYY	b1b1b2b2SS
BTX 378	USA	Red	A	–	I (S)	0.1	RRYY	b1b1b2b2SS
MR 724	ICRISAT	White	A	–	I	0.0	RRyy	b1b1b2b2ss
MR 723	ICRISAT	White	A	–	I	0.0	RRyy	b1b1b2b2ss
2219A	ICRISAT	White	A	–	I	0.0	RRyy	b1b1b2b2ss
CK60B	USA	White	A	–	I	0.1	RRyy	b1b1b2b2ss
TX623B	USA	White	A	–	I	0.1	RRyy	b1b1b2b2ss
Daber	Africa	White	A	–	I	0.1	RRyy	b1b1b2b2ss
Hegri/durra	USA	Yellow	A	–	I	0.1	rryy	b1b1b2b2ss
P721N	USA	White	A	–	I	0.1	RRyy	b1b1b2b2ss
KafinamB	ICRISAT	White	P	p	II (P)	0.1	RRyy	B1B1B2B2tptp
Feterita	Africa	White	P	P	II (P)	0.2	RRyy	B1B1B2B2tptp
IS0135	USA	White	P	B	II (B)	1.2	RRyy	B1B1B2B2TpTp
IS2319	USA	White	P	B	II (B)	1.4	RRyy	B1B1B2B2TpTp
IS8544	USA	White	P	B	II (B)	5.8	RRyy	B1B1B2B2TpTp
IS8768	USA	White	P	B	II (B)	2.9	RRyy	B1B1B2B2TpTp
Lishihuang	China	Yellow	P	B	III	4.2	rrYY	B1B1B2B2TpTp
Sanchisan	China	Red	P	B	III	4.4	RRyy	B1B1B2B2TpTp
Xin No. 7	China	Red	P	B	III	4.4	RRyy	B1B1B2B2TpTp
Jin No. 5	China	Red	P	B	III	5.1	RRYY	B1B1B2B2TpTp
HC356	China	Red	P	B	III	3.6	RRYY	B1B1B2B2TpTp
HM65	China	Red	P	B	III	4.1	RRYY	B1B1B2B2TpTp
S-37	Africa	Red	P	B	III	4.1	RRYY	B1B1B2B2TpTp
BR64	USA	Red	P	B	III	12.8	RRYY	B1B1B2B2TpTp

A: Absent; P: Present; Testa color was determined by visual observation of kernels with pericarp scratched off with a pocket knife

Table1.3 Bleach test problems and their suggested solutions.

Problem	Suggested solution
Kernels have black tips	All kernels have a black hylar region (where the kernel attached to the plant)
Kernels are white	Compared to standard sorghum checks a. All kernels do not have tannin b. Shaking time too long or the bleach is strong.
Kernels are red to brown	Compared to standard sorghum checks a. All kernels do have tannin b. Pericarp is not removed completely. (Bleach is old or shaking time is not enough.)
Kernels have dark speckles or spots	Compared to standard sorghum checks. Kernels may be damaged by insects, disease or weathering. (Darkness of the spots depends on the damage extent)
Kernels are black	Compared to standard sorghum checks. All kernels have tannin.

Table1.4 Methods used to determine tannins. (Adapted from Hahn et al., 1984; Hagerman et al., 1997)

Tannin	Standard	Reagents	Time	What is Measured
Vanillin	Catechin	4% HCl,	24 hr	Leucoantho-cyanidins
		1% vanillin in methanol	20hr	Condensed tannins
Prussian blue	Catechin	FeCl3 in HCl	1 min	Total phenols
Folin-Denis	Tannic acid	Folin-Denis reagent	5 hr	All reducing substances
Protein Precipitation	Tannic acid	FeCl3, alkaline detergent	20 min	Condensed tannins
Folin-Ciocalteu	Catechin or gallic acid	Folin-Ciocalteu reagent	1 hr 2 hr	Total phenols
Polymerization Degree		4% HCl	20 min	Anthocyanidins
		0.5% vanillin in acetic acid		Condensed tannins
		4% HCl in acetic acid		Leucoathocyanidins
Acid butanol	Catechin	5% HCl in 2-butanol	1 hr	Condensed tannins

Table 1.5 Condensed tannin complementary genes theoretical additive effects and their interaction.

	B_2-locus genotype		
B_1-locus genotype	B_2B_2	b_2b_2	Marginal mean
B_1B_1	$m+a^1+a^2+aa$	$m+a^1-a^2-aa$	$m+1/2a^1$
b_1b_1	$m-a^1+a^2-aa$	$m-a^1-a^2+aa$	$m-1/2a^1$
Marginal mean	$m+1/2a^2$	$m-1/2a^2$	m

Figure 1.3 Condensed tannins and anthocyanins metabolic pathways. (Adapted from Petit et al., 2007).

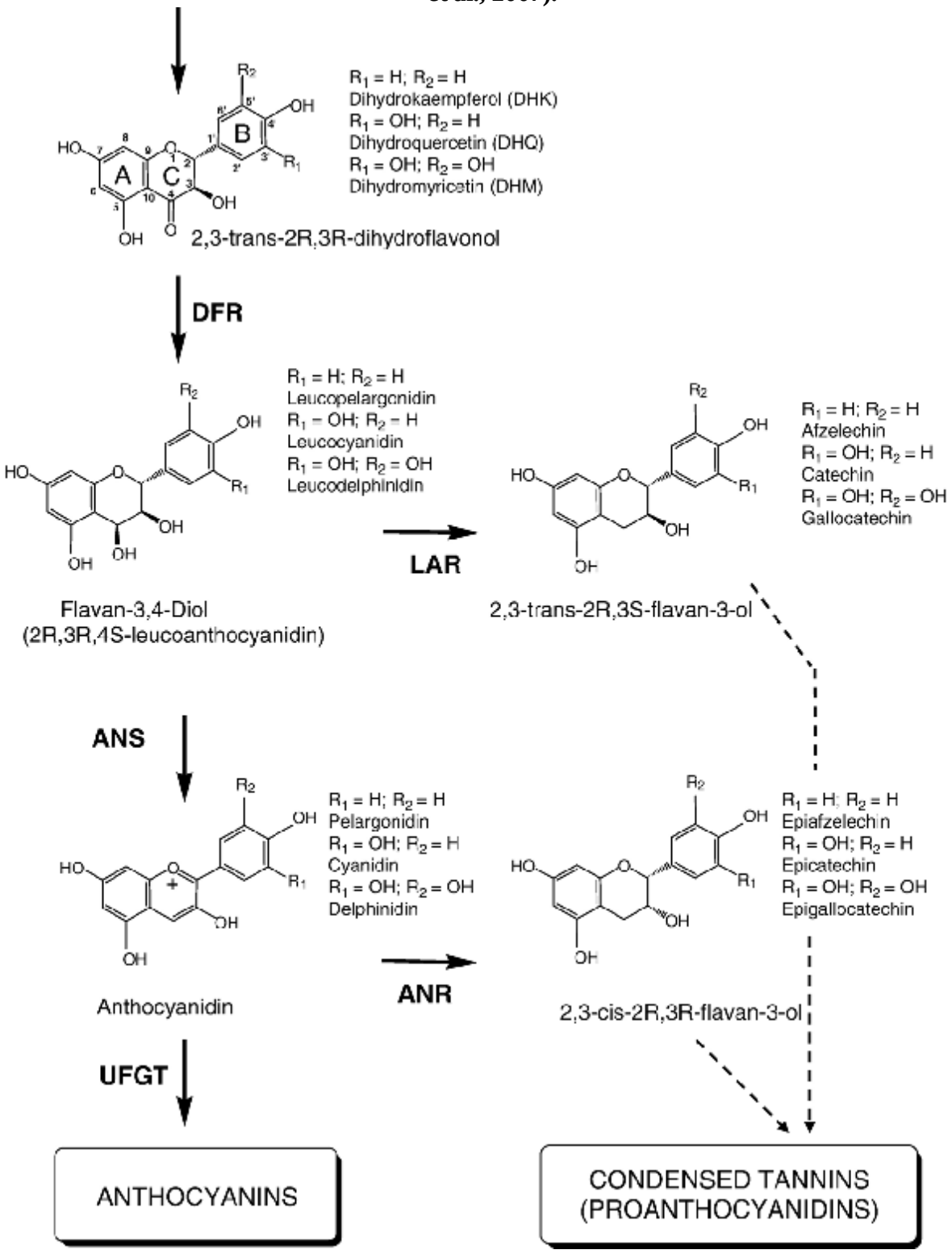
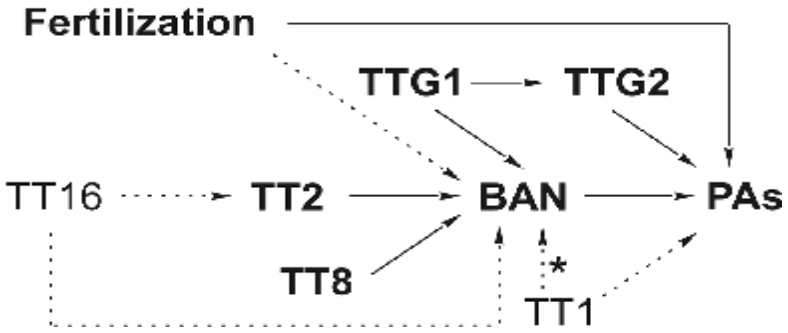


Figure 1.4 Tannin genetic regulatory pathways. Arrows indicate positive regulations. Gene activities prevailing in both the endothelium and the chalaza/micropyle are shown with boldface letters and solid arrows, and endothelium -specific activities are indicated with lightface letters and dashed arrows. The star indicates that the activity is necessary in a few cells at the endothelium base.



CHAPTER 2 – TWO INTERACTING QTL CONTROL FOR THE CONDENSED TANNIN IN SORGHUM GRAIN

Abstract

Condensed tannin is a unique secondary metabolic product that accumulates in the sorghum seed testa layer underneath the pericarp. The formation of the testa layer and presence of condensed tannin were shown to have antinutritional effects for human consumption and be related to sorghum resistance to stress. But recent studies also demonstrated the antioxidant activity of sorghum tanning for industry usage. Classical literature indicated that the presence of condense tannin is controlled by two unlinked genes acting in a complementary dominant epistasis pattern in an F_2 population. But the exact genes and their genomic locations remain unknown. Our objective is to identify the quantitative trait loci (QTLs) controlling the presence and absence of condense tannin in the sorghum grain. Towards this goal, a high tannin line, Shan Qui Red, and a non-tannin line, Tx430, were crossed. The resulting recombinant inbred line (RIL) mapping population was phenotyped with a bleach test and genotyped with simple sequence repeat (SSR) markers for QTL analysis. Two QTLs with significant epistatic interaction were identified, with one on chromosome 2 and the other on chromosome 4. Multiple interval mapping indicated these two QTLs and their additive by additive epistasis explained about 74% of the overall phenotypic variation, which agreed with the binary trait analysis. Further research is being conducted to fine map these QTLs for gene tagging.

Introduction

Sorghum is the fifth cereal crop in the world (Sasaki and Antonio, 2009). It is a C_4 crop with high photosynthetic efficiency and high water use efficiency. It has good tolerance to extreme environments like high temperature, drought, and poor soil conditions. It's also a good gluten-free food resource for people who have celiac diseases and can not consume wheat products (Schober et al., 2005). Compared to other cereal crops, sorghum is unique for its secondary metabolic product, tannin, present in tesat layer of some accessions. Tannin is a favorite agronomic trait but, has negative impacts on feed quality. Recently, it was shown that tannin is a good antioxidant, which is a potential medical treatment for cancer and cardiovascular disease (Awiak et al., 2003). Previous studies have also found a strong relationship exists between tannin presence and seed emergence in cold environment, which is related to sorghum early season cold tolerance and grain yield (Knoll et al., 2008).

Sorghums with pigmented testa layer contain condensed tannins. Testa layer is underneath the outmost pericarp layer and covers the endosperm. In most of sorghum accessions, it is either absent or present. The testa layer (and thus condensed tannin) presence is controlled by two complementary dominant genes B_1 and B_2. Condensed tannin is present when both loci are dominant ($B_1\_B_2\_$) and absent if either or both of the loci are recessive ($B_1\_b_2b_2$, $b_1b_1B_2$, and $b_1b_1b_2b_2$). Previous research using AFLP makers mapped two QTLs on chromosome 2 and 4 respectively. Efforts were employed to convert these AFLP makers into STS markers, but problems of indetectable and unrepeatable band types occurred. To benefit tannin gene dissection in future using high throughput research approach, commonly used simple sequence repeat (SSR) maker is in need for QTL position and effect identification.

To identify the location of the condensed tannin QTLs in sorghum genome and their effects, a RIL population with 109 plants were constructed by cross of high tannin SQR and non-tannin Tx430. SSR markers which are commonly used in both linkage mapping and association mapping were employed for QTL dissection. Multiple interval mapping method was utilized for complementary QTLs interaction exploration.

Methods and Materials

Plant materials

A population of 109 $F_{6:7}$ recombinant inbred lines (RILs) were developed from the cross of a high-tannin sorghum line, Shan Qui Red (SQR), and a non-tannin commercial line Tx430. As a Kaoliang sorghum, SQR was derived from cool temperature and high latitude regions in China with relatively broad leaves covered with waxy blooms. SQR produces red color seeds with high tannin content, presents a relatively high germination rate, rapid growth and strong early season cold tolerance compared to commercial lines in the US. But it also contains many agronomically unfavorable traits, such as tall plant height, open panicle, and susceptibility to many diseases. On the contrary, Tx430 is a typical "dwarf type" adapted U.S. line which is less than 5 feet in height and suitable for harvesting with combines, served as pollinator in hybrid production. But Tx430 has no tannin and is susceptible to early season cold stress.

Tannin phenotyping

Condensed tannin presence in testa layer was identified using bleach test. $F_{6:7}$ RILs population grown in Manhattan, 2005, was tested for three duplications. Data sets of RILs grown in Manhattan in 2001-2005 were also compared for validation and better identification of tannin

presence purpose, among which Vanillin/HCl assay data also obtained for 2001, 2002, and 2003 RILs and served as comparative parameters.

The bleach solution is made by 15 grams KOH pellets with 40ml of fresh standard commercial bleach. Each 1.5ml Eppendorf® tube with 10 sorghum kernels was filled with 1 ml of bleach solution and shook for 3 minutes for complete reaction. Then, kernels were poured into a tea strainer and rinsed with 70˚C water to remove the KOH-bleach solution. After removing excess water, kernels were placed in plate for completely dryness. Color screening was processed after all the kernels were dried. Tannin grains have black over the entire surface, only with the exception at germ place which is lighter in color. Non-tannin grains are either completely white, or partially brown.

Sorghum DNA extraction

Two-week-old fresh leaf tissue was collected from a single plant which represents most of the plants within an RIL line morphologically. Those tissues were placed in 1.5ml Eppendorf® tubes, lyophilized for 48 hours in a freeze drier (Thermo Fisher, Waltham, MA), and ground to fine powder in a Mixer Mill (Rheinische Strasse 36, Germany) by shaking tubes with a 3.2-mm stainless-steel bead at 25 times/sec for 5 min. Genomic DNA was extracted from parents and RILs by the revised cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987). For each tube with 0.15g leaf tissues, 800ul 2x CTAB buffer was added. Samples were incubated at 75^{o}C for 1hr, to guarantee sufficient reaction, tubes were shook up and down twice during the 1 hour. Next, 800ul of chloroform/isoamyl alcohol (24:1) was added, mixed well, then centrifuged at 12,000 rpm for 10 min. After centrifuging, 600ul of aqueous phase was transferred to a new tube. Then 500ul of cool isoamyl alcohol were added. Tubes with DNA precipitation

were centrifuged at 12,000rpm for 10 min. DNA was washed and purified with 70% alcohol. And then dissolved DNA in 200ul 0.1X TE.

Maker selection and primer design

Single sequence markers located on chromosome 2 and 4 were first selected based on previous researches (Bhattramakki et al., 2000; Menz et al. 2002; Knoll et al., 2008; Li et al., 2009). Markers named with prefix of Xtxp are commonly used SSR makers. Markers named with a five-digit number are makers designed from public genomic sequence contigs of sorghum by other sorghum research groups (Li et al. 2009). Markers named with prefix W are newly designed markers in this study.

In total120 new SSR makers were designed in this research. Maker design was based on sorghum bicolor genome sequence download from www.phytozome.com. The following marker design principles were followed: 1) PCR primers generally ranged in length from 18-25 basepaires and the amplified DNA fragment products size should be in the range of 120-250 base pairs. 2) The GC content of the whole primers, both forward and reverse, should be within range of 40% to 60%. 3) Avoid sequences which would produce internal secondary structure. 4) The 3’ ends of the primers should not be complementary to avoid the production of primer-dimers in the PCR reaction. 5) The primer’s 3’end should be end with G and C, because the hydrogen bonds produced by G and C is much stronger than hydrogen bonds between A and T, which facilitate steady prolonging of the DNA amplified fragment, but three G or C nucleotides in a row near the 3’ end of the primer should be avoided. 6) Both primers should have similar G＋C content so that they anneal to their complementary sequences at similar temperatures. The temperature difference between forward and reverse primers should be less than 4°C.

PCR amplification and genotyping

A 20-μl PCR mixture contained 2 μl of 10X NH4 buffer (Bioline Inc. Taunton, MA), 2ul 2.5 mM of $MgCl_2$, 200 μM of each dNTP, 2ul primer, 1 U of Taq DNA polymerase, and 50 ng template DNA. GeneAmp® PCR system 9700 was used for PCR amplification with the touchdown program. Initially PCR mixture was denatured at 95°C for 5 min; the first five cycles started with 1 min of denaturing at 96°C, 5 min of annealing at 68°C with a decrease of 2°C for each subsequent cycle, and 1 min of extension at 72°C; in the next five cycles, the annealing temperature started at 58°C for 2 min with a decrease of 2°C for each subsequent cycle; the last 25 cycles ran 1 min at 96°C, 1 min at 50°C, and 1 min at 72°C with a final extension at 72°C for 5 min and 4°C for 5 min.

Marker surveys by agrose gel electrophoresis were first conducted to identify polymorphic markers from an available SSR marker set. Polymorphic markers between the parental lines were further analyzed for linkage mapping and QTL analysis in the populations. For SSR markers which didn't show good polymorphism on agrose gel, another PCR mixture was used for phenotyping by ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). This mixture contained 1.2 μl of 10X NH4 buffer (Bioline Inc. Taunton, MA), 2.5 mM of $MgCl_2$, 200 μM of each dNTP, 100 nM of forward tailed primer, 200 nM of reverse primer, 100 nM of M13 fluorescent-dye labeled primer, 1 U of Taq DNA polymerase, and 50ng template DNA.

Because the PCR products of most primers in this research are low in molecular weight, 4% agrose gel is used for agarose gel electrophoresis. 10X TBE was used as buffer. Bromophenol blue is used as gel loading buffer dye. Separated PCR amplicons were stained with ethidium bromide (10mg/ml) and visualized by UV lights on the gel.

The separation and detection of SSR products with relatively small amplicon size were operated by capillary gel system ABI 3730 DNA analyzer. The data was analyzed and scored using GeneMarker® version 1.6 (SoftGenetics LLC, State College, PA).

Linkage and QTL analysis

Linkage maps were constructed by using MAPMAKER/EXP version 3.0 (Whitehead Institute, 1993). LOD threshold value was set at 3.0 to group linked markers in linkage groups. Kosambi mapping function was used to convert recombination frequencies into genetic distance in centimorgan (cM) and create the genetic map.

Composite interval mapping (CIM) and Multiple Interval Mapping (MIM) methods were employed to map QTL by QTL Cartographer version 2.5. Composite interval mapping was implemented by starting with default values of five cofactors obtained by a forward regression to control genetic background. To block a chromosome region between the nearby markers and the test site, a window size of 10 cM was used. In CIM, the walking speed for genome wide QTL scan was set at 1.0 cM and the LOD thresholds to declare a significant QTL were determined based on the result of 1000 permutations. Coefficient of determination (R^2), which was the proportion of total phenotypic variance explained by a QTL, was determined based on the R^2 of the single marker that was the closest to the target QTL. All loci that had significant main effects were tested against all other markers to detect significant interactions ($P < 0.01$). MIM was used to determine the additive effect of QTLs detected by CIM and any epistatic (additive by additive) interaction between QTLs. Forward and backward selection on markers were used for initial MIM model select method. Criteria of model selection was set at significant level 0.05, with BIC criteria of g(n)=ln(n). QTL number range was set to 2. Since based on the classical theory, the

complementary genes should be two. Afterward, the QTL interaction term was selected to present in the result. Model refine is used to detect the optimalizing positions of two tannin QTLs.

Result

Phenotypic data evaluation

Because of weathering and insect and pest bites, test kernels might have dark speckles or spots, or partially colored, which could make false positive without careful classification (figure 2.1). Also testa layer thickness variation may cause inaccurate results while the small concentration bleach was applied. USDA-FGIS standard bleach protocol is used for large size sample screening with specific equipment. Comparative bleach tests were set up to verify feasibility on the SQR/Tx430 population. Comparative tests were: 1) standard USDA protocol: KOH 15 grams, fresh bleach 40ml. 2) KOH 7.5 grams, fresh bleach 40ml 3) KOH 15 grams, fresh bleach 20ml, H_2O 20ml 4) KOH 15 grams, fresh bleach 20ml, H_2O 20ml. A subset included 20 RILs was used for the bleach test result comparison. The standard USDA protocol was proved to be the best combination of condensed tannin presence identification.

The seed kernels of 109 SQR/Tx430 RILs and their parents together with other 10 high tannin lines were bleached to test presence or absence of pigmented testa layer with tannins. The parental lines turn out to be significant different in bleach test, while SQR and all the high tannin lines present dark black color and classified into tannin presence, Tx430 was totally white after bleach, classified as non-tannin. RILs segregated for tannin presence and absence. In total, there are 41 tannin lines and 68 non-tannin lines out of the 109 (Table 2.1; Figure 2.3). To minimize the scoring error, three duplications of Manhattan 2005 RILs were carried out and compared with seeds of the previous year RILs population (Manhattan 2001 and Manhattan 2004). Tannin

samples were score as 1 and non-tannin ones were scored as 0. Total and tannin and non-tannin sample number of each population after bleach score is showed in table 2.1. Only 10 RILs out of the whole population showed differences in tannin screening. After careful examination and comparison, combined tannin data were used for final analysis.

Marker analysis and QTL identification

Because tannin QTLs and related cold tolerance QTLs have been previously located on chromosome 2 and 4 based on genetic analysis (Ridder, 2004; Knoll, 2008), 229 microsatellite markers (SSR) covering sorghum chromosome 2 and 4 were selected to screen the parental lines. Those markers include markers which were proved linked to tannin QTLs and cold tolerance QTLs in previous research. Marker analysis identified twenty six polymorphic SSRs between SQR and Tx430 (Table 2.2; Figure 2.4; Table 2.5).

Linkage analysis detected two linkage groups located on Chromosome 2 and 4 associated with the tannin trait. The first linkage map covered a total genetic distance of 94.2 cM with 14 makers. The second linkage group covered a total genetic distance of 79.82 cM with 12 makers (Figure 2.2). Two QTLs were identified on chromosome 2 and 4 separately by IM, CIM and MIM. A slight difference in QTL locations was observed between IM and CIM on chromosome 2, which the QTL peak shifted towards left in IM than in CIM. Since CIM reduced the genetic noise and enhances the QTL detection power by incorporating cofactors in the model, the confidence intervals determined by one-LOD reduction from the maximum LOD value tend to be smaller in CIM than in IM. For this reason, the results from CIM were used for further result interpretation. CIM detected two major QTLs with major effect (R^2= 0.08 and 0.40) on the chromosome 2 and chromosome 4 (Paterson et al., 2003).

Marker W072 and marker W075 were the closed markers to QTL_1 at chromosome 2, and 53842 and W015 were flanking markers for QTL_2 at chromosome 4. W072 was the most closely linked marker to QTL1 at 1.2 cM, while W075 is in the vicinity of 2.8 cM. The marker 53842 was mapped at 1.6 cM proximal to QTL_2, and the maker W015 was 2.1 cM distal of QTL_2.

MIM also used to detect the interaction and its effect to QTL location and gave a pretty good result consider the additive effect and epistastic interaction effect it presented. According to complementary dominance theoretical values for RIL population, the QTL_1, QTL_2 and their additive by additive interaction effect should be 0.25, 0.25, and 0.25 respectively. The actual values detected by the MIM were 0.2615, 0.2317, and 0.2552 respectively (Table 2.4).

Discussion

Bleach test

We chose the bleach test because it is widely accept as an efficient method for tannin measurements with USDA identified standard. Bleach test is easy to conduct with time and cost savings. However, many environment factors like weathering, mold, and insects bites may cause dark spots on the bleached kernels and affect the accurate and repeatability of the bleach test. Also because the pericarp and testa layer thicknesses vary in different sorghum germplasm, and a partially testa layer may be present, and the bleach concentration may be suitable in one accession but too strong or weak in other sorghum accessions. All these factors may lead to false positive results. Furthermore, USDA-FGIS bleach test standard is used with special equipment for large amount of seeds. Seeds proportion or percentage is an important identification standard for classification. In our research only ten seeds were tested for each sorghum sample, which may cause classification bias. To verify the efficiency and accuracy of the test methods,

comparison experiments with different bleach solution concentrations were set up. Data for RIL population of several years were compared, the frequency distribution (tannin: non-tannin ratio) was consistent in the three RIL populations tested. Such result indicated that the standard bleach test still is the best identification protocol for a sample size of 10 kernels.

Because the testa layer is derived from tissue laid down by the maternal parent (maternal effects), testa color phenotypes of one generation were determined by seeds genotypes of the previous generation. For example, $F_{2:3}$ testa color phenotypes were determined from F_2 seeds and $F_{6:7}$ phenotypes from F_6 seeds. So, when the RIL population is pure enough, the tannin presence should be consistent in successive generations. Also the data comparison research for several generations might give more heritability cures and dissect heritability of the tannin genes in-depth.

In this research, tannin trait was scored as 1 for presence and 0 for absence, which is classified as binary trait. Further research employing vanillin/HCl assay with a continuous data set may provide a distribution graph of tannin phenotypic result and lead to more accurate identification of tannin QTL position and effect (Walton et al., 1983). A critical issue for vanillin/HCl assay data collection and management is that, a standard threshold value needs be chosen to classify seeds into tannin and non-tannin. Test values of the parental lines and also the background phenol compound values should be taken into consideration for the threshold set up.

Candidate genes for tannin presence

TTG_1 is an important candidate gene for condensed tannin production in Arabidopsis seed coat (Zhang et al., 2003). Previous comparative genome research revealed that TTG_1 is also similar to pale aleurone color 1 (PAC1) locus in maize, which is required for anthocyanin

pigment in the aleurone and scutellum of maize seed. Consequently, TTG_1 is considered as a good candidate for one of the condense tannin complementary genes. However, the flanking maker linked to this gene was used to screen SQR and Tx430, no polymorphism was shown between the two parental lines. Also, the PAC1 position in sorghum genome on chromosome 4 is located far from the QTL_2 position detected in this research, which indicated that sorghum tannin gene accumulation and regulation mechanisms are different from the complex system of Arabidopsis and maize and might be more simple, which is good for marker assistant selection and consequently sorghum breeding.

As mentioned early, candidate gene association mapping combined with QTL mapping is a good approach for target gene identification. This approach asks for available candidate genes related to tannin metabolic pathway or regulation factors in charge of tannin accumulation in testa layer cells. Therefore, the two QTL peak regions identified in this research were explored for potential candidate genes by comparing sorghum sequence at www.phytozome.com. However, for both the two region and their extension regions which are 200kb in length, there were only five genes with defined function. All the five function defined genes are unrelated to any metabolic pathways related to tannin presence. There are 15 more unknown genes included in these two regions. More cooperative research by genome comparison via bioinformatics tools or biochemical metabolic researches is needed for further dissection of tannin genes.

Mapping result validation

Population size and suitable QTL detecting methods are important factors affecting precision and accuracy of QTL mapping (Doerge, 2002). In the beginning of this research, a subset of 94 RILs was used for initial scoring. A ghost QTL peak was found near the QTL on

chromosome 2 by both IM and CIM. After increasing the population size to 109, this ghost QTL disappeared, especially in the CIM mapping result, suggesting that the QTL detection power may increase with population size and a large population significantly improves precision for QTL detection. CIM appears to be more reliable in predicting QTL as it considers the background effect by taking cofactors into account, with which the total variance that is determined by other linked QTL is reduced, therefore increasing the relative variance explained by the target QTL.

In this research, heterozygous band types which included the bands types of both the two parents were identified in genotypic screening. Band types which were different from both parents and could only be scored as missing data also present in four markers. Such results indicated the 109 plant RIL population was not pure enough. Purification of the RIL population was needed. Furthermore, according to the Bevis effect, small populations tend to emphasis major QTL effects explained by a limited number of QTLs, while large population would predict more QTLs with relative less effect for each individual QTL. To verify the two complementary genes hypothesis and validate the QTLs detected by SQR/Tx430 RIL population, bigger mapping populations and more population resources should be explored.

According to the classical complementary dominant gene hypothesis, many US sorghum lines are found to be fixed with the B_1 gene but not the B_2 gene. Collecting such materials would provide good background for tannin gene mapping. Utilizing mapping population derived from cross of B_1 fixed lines with high tannin lines, single QTL for B_2 gene is easier to be dissected since complex interaction term is removed. Subsequence QTL detection for B_1 gene could also be simplified.

Previous research based on RIL population derived by the cross of SQR and SRN39 discovered cold tolerance QTLs (Knoll et al., 2008). Previous QTL mapping using AFLP makers

located the two tannin QTLs near the region of SSR marker Xtxp 211 and Xtxp 212 on chromosome 2 and chromosome 4, respectively, which is the flanking marker of cold tolerance QTL. If the same population could be employed for tannin QTL validation, the relationship between cold tolerance QTLs and tannin QTLs would be clearer. Comparison study of proximal markers of these two groups of QTLs could provide more direct QTL positional evidences. As a result, lead to dissection of the two correlated traits: whether they are controlled by pleiotropy or they just tight linked in a short chromosome segment.

Deployment of tannin QTL

Both the QTL_1 and QTL_2 were located within a relatively small interval of approximately 3 cM, which should be helpful for further high-resolution mapping, assuming that there are enough polymorphic markers at the distal of long arm of chromosome 2 and 4. If the best markers show consistently high polymorphism in more collections of germplasm, they could be readily used for marker-assisted breeding. Breeders could choose to select the allele from a non-tannin tannin parent or select against the allele from a high tannin parent to decrease the level of tannin sorghum percentage for food and feed sorghum breeding, in which non-tannin is preferred. Wild sorghum could serve as new germplasm and tannin genes could be knocked out, and then elite lines cross to the wild ones to enlarge the genetic diversity of sorghum for breeding and selection. Further research may find other minor QTL or QTL in gene regulation regions which control the tannin content or metabolic pathways related to accumulation of tannin in testa layer cell vacuole and select high tannin sorghum for industry antioxidant production usage. As mentioned before, most sorghum cultivars in the US are fixed with the B_1 gene but lack of B_2 gene. Breeding work could be addressed by selecting sorghum possess B_1 by parents and

progeny for a tannin relationship studies. Afterwards, fix B_2 gene. By this way, breeding circle would be shorten and resource saving.

Figure 2.1 Typical bleach test results. From left to right, top to bottom: white(non-tannin), black (tannin), brown (tannin), white with spots (non-tannin), tender yellow with spots (non-tannin), partially brown with spots (non-tannin).

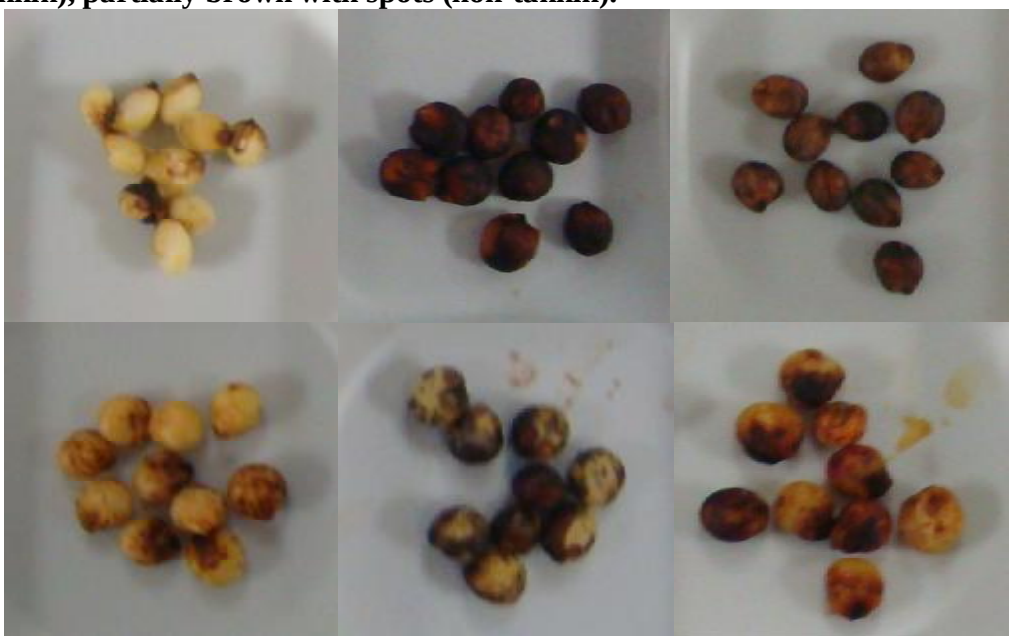


Figure 2.2 Comparative bleach tests. From up to down, left to right 1) standard USDA protocol: KOH 15 grams, fresh bleach 40ml. 2) KOH 7.5 grams, fresh bleach 40ml 3) KOH 15 grams, fresh bleach 20ml, H_2O 20ml 4) KOH 15 grams, fresh bleach 20ml, H_2O 20ml.

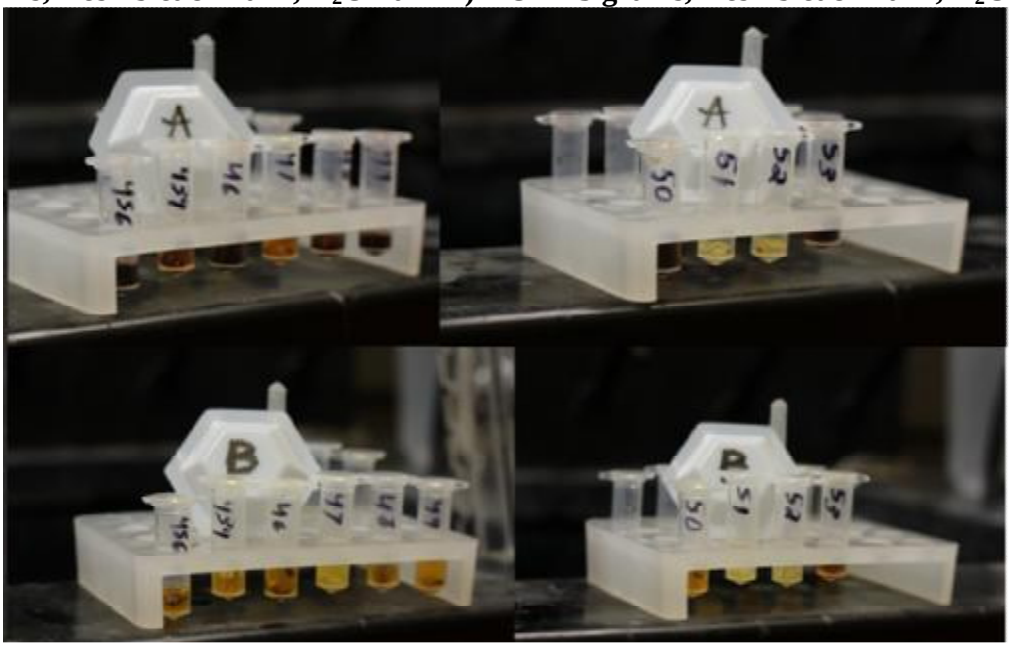


Table 2.1 Bleach test results.

Seed Set	Total Num.	Tannin	Non-tannin	Expected Ratio	X^2
MN01	109	47	62	1:3	19.09
MN04	99	41	58	1:3	13.84
MN05-1	109	44	65	1:3	13.73
MN05-2	109	46	63	1:3	17.20
MN05-3	109	43	66	1:3	12.14
MN05	109	41	68	1:03	9.25

Table 2.2. Segregation ratios of SSR marker alleles in SQR/Tx430 RIL population

Marker	Total	Tannin	Non-tannin	Observe Ratio	Expected Ratio	X^2
Xtxp297	101	53	48	1.10	0.33	1.78
Xtxp211	104	65	39	1.67	0.33	5.33
304	104	60	44	1.36	0.33	3.18
W089	109	63	46	1.37	0.33	3.22
W091	109	74	35	2.11	0.33	9.52
W100	109	63	46	1.37	0.33	3.22
W105	109	67	42	1.60	0.33	4.78
W072	100	55	45	1.22	0.33	2.37
W075	104	61	43	1.42	0.33	3.53
35100	105	53	52	1.02	0.33	1.41
W057	106	64	42	1.52	0.33	4.25
Xtxp4	108	67	41	1.63	0.33	5.08
9142	107	72	35	2.06	0.33	8.91
Xtxp201	107	65	42	1.55	0.33	4.42
Xtxp21	106	66	40	1.65	0.33	5.20
W017	106	75	31	2.42	0.33	13.05
W015	106	70	36	1.94	0.33	7.79
53842	108	75	33	2.27	0.33	11.28
W013	108	70	38	1.84	0.33	6.83
W012	106	68	38	1.79	0.33	6.36
33522	105	66	39	1.69	0.33	5.54
71443	104	69	35	1.97	0.33	8.05
59974	104	73	31	2.35	0.33	12.26
Xtxp327	106	70	36	1.94	0.33	7.79
37422	108	67	41	1.63	0.33	5.08
3484	105	67	38	1.76	0.33	6.13

Table 2.3 QTL flanking markers statistical analysis.

Source	DF	SS	MS	F value	Pr > F
Model	3	19.203	6.4009	96.88	<.0001
53842	1	6.9675	6.9675	105.45	<.0001
W072	1	3.3584	3.3584	50.83	<.0001
53842*W072	1	3.3584	3.3584	50.83	<.0001
Error	104	6.8714	0.0661		
Corrected Total	107	26.0741			

Table 2.4 Composite interval mapping (CIM) and multiple interval analysis (MIM) of QTLs associated with testa layer condensed tannin presence.

Chromosome	Method	Position(CM)	Marker	LOD	Additive Effect	Threshold	R^2
2	CIM	71.2	W072	14	0.192	2.5	0.13
2	MIM	62.7	W072	15.5	0.262	2.5	0.08
4	CIM	70.8	53842	30	0.265	2.5	0.39
4	MIM	37.4	53842	23	0.232	2.5	0.40

Figure 2.3 QTLs detected by CIM. Cofactors selection was conducted with "unlinked marker control" procedure.

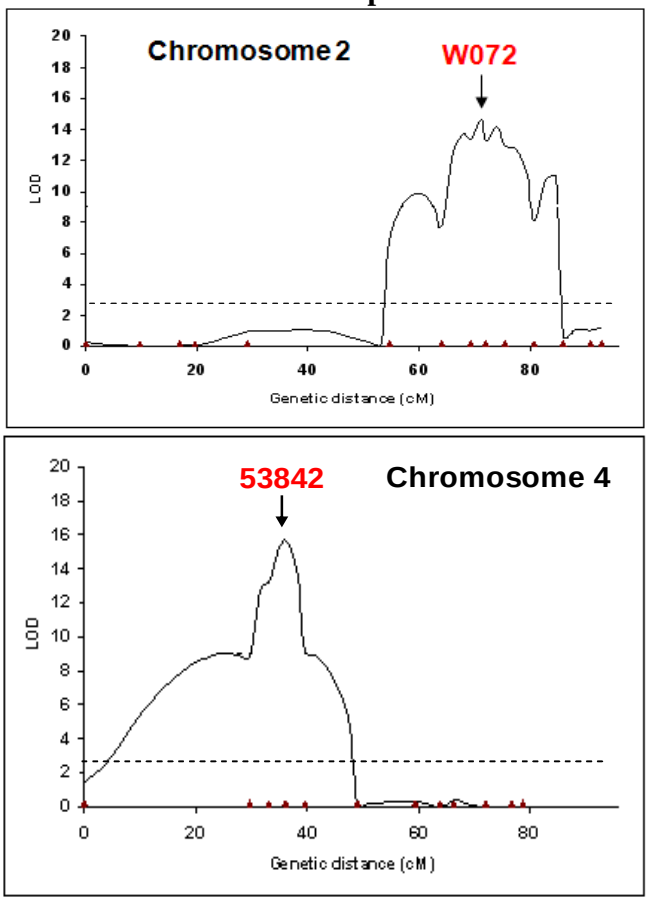


Figure 2.4 QTLs detected by MIM. The number of QTLs was set to two to start the searching for both main effects and interactions.

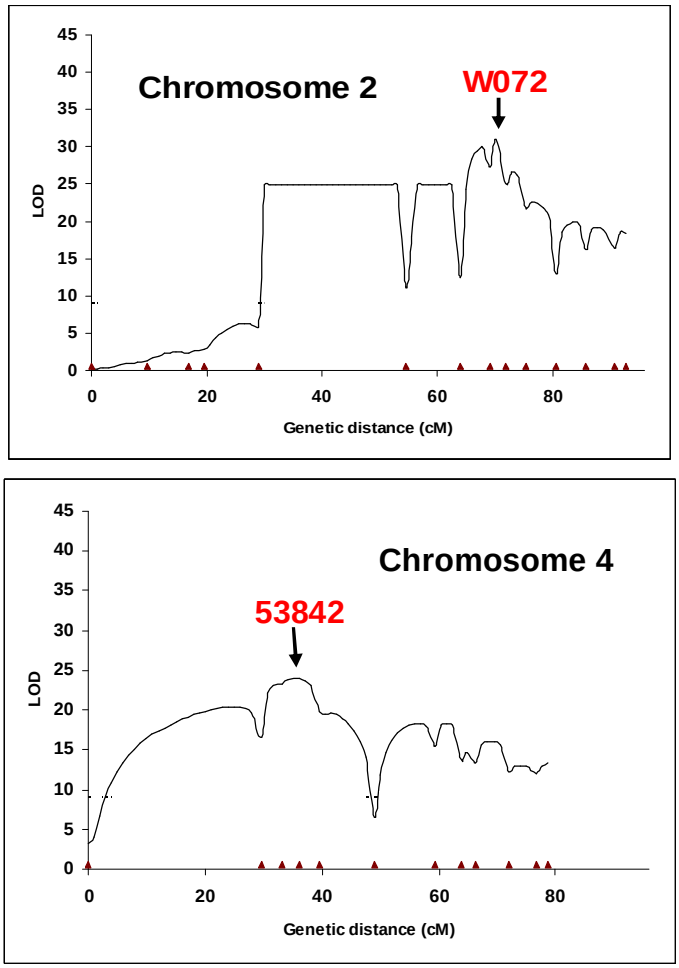
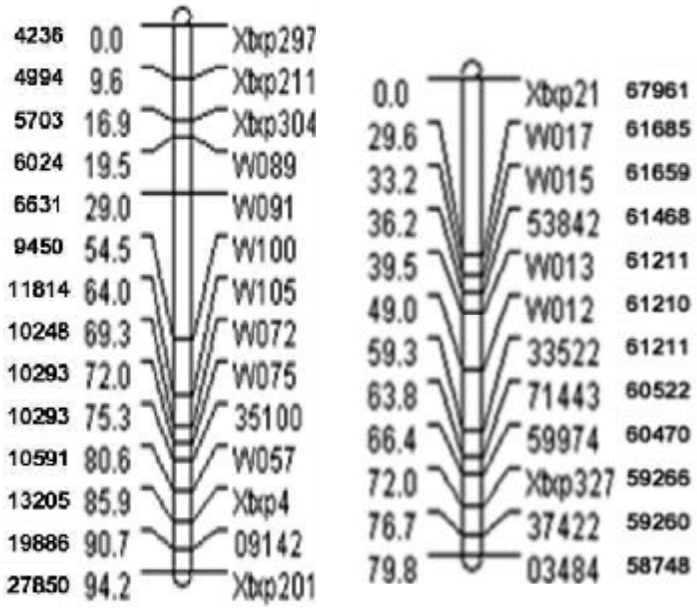


Table 2.5. Characteristics of sorghum bicolor tannin genes and SSR loci that were analyzed and the primer sets used to amplify them.

Marker name	Genetic distance	Physical position	Sequence of forward primer	Sequence of reverse primer
Xtxp297	0	4236926	GACCCATATGTGGTTTAGTCGCAAAG	ATTGTTGATTTAGGCGAAGATTGTGC
Xtxp211	10.4	4994787	GTTAGAAATCATTGGCCGTTGA	GTTGCGAATAAAAGGTAATGTG
304	18.3	5703328	GAAGAGGGGCTTTTATGT	CTTTCACACCCTTTATTCA
W089	21.4	6024685	GGCGTGTACTAGCTTCCTC	AGGGTTCCCGTCTTCAAT
W091	31	6631092	AGAATGATGTGCTCTACAGACA	ATTTTGACACGATAGGACTGA
W093	41.5	6869388	AGAAGTAACCTTTGAGTTTGGA	GTGGTCACAACTAGCTTATTGA
W099	66.9	-*	AGAGTCTCCGTGATTCACC	AGAACGCCATGCTCATG
W100	80.1	9450855	GTTTCTCTCTCGCTCAAAAG	AGCGTACAGTACCTCTCATCA
W105	90.6	11814002	GTGACTGGGATGGAGGTT	GAGATACGCTAGGCTCAAAA
W072	93.3	10248319	CTCTATGGCGACTAAGATGG	TAGCCCACTCTAAACCCTAAC
W075	93.9	10293489	CATCATCAGGCTTTGGAAGG	GTGGCACAAGTAGTAACAGG
35100	98.8	10293749	AAGATATTCCAGCTTGTGGA	GCAACGAACGTAGTATAGGG
W057	103.7	10591745	AGAAGCATTGGCACTATACC	TGGGAGAACATGTCTGAAAG
Xtxp4	108	13205419	AATACTAGGTGTCAGGGCTGTG	CTTGGTTGTTGCGGTTACAT
9142	110.8	19886670	AGAACTGGTCCAAGTCCAG	AACTAGCATCGATCGACCTA
Xtxp201	121.8	27850987	GCGTTTATGGAAGCAAAAT	CTCATAAGGCAGGACCAAC
Xtxp21	0	67961718	CGACCAAATCTATGGCAGCTC	ACCTCGTCCCACCTTTGTTG
W017	29.7	61685362	GTCAATGTGCCATACAGTAC	CACCTATATATGAGGTACGG
W015	32.5	61659852	GATCCGGTCGTTTGGTTTG	AACGAACCCACGCAATGTC
53842	35.5	61468888	TATAGCACGTTTTTAACTCGTG	TAGTCTAAAAGCCCGTTTCA
W013	38.8	61211508	CCTCGGTTGGTTCTTGATTG	CATTGGACGAAGAAGCTCG
W012	48.3	61210467	GGTAATGCCGTGTTAGTGG	CACACACGCATATGTTTTGG
33522	58.6	61211486	GGTGAGTGCGACTACGAC	GTTCCTCGGTTGGTTCTT
71443	62.7	60522623	AGAGGGAAGCAGCTGACT	GCCAGAAGACCGACTACAT
59974	65.1	60470490	GACGATGCCATTAGCTTATT	CCGGTTGTTGAGAAGTTG
Xtxp327	70.1	59266753	GTGAGCGTGAGCAGTGGT	GCGGTGTACAGCTTCGTC
37422	74.9	59260599	TCCCCATGGTCCTAAAGT	TATTTAGTAGCAGCCCCAAG
3484	78	58748445	ACAAATTGGCAATGCTAAGT	CAAACCAACATAGCTGTAAATG

*** The physical position of this marker could not be validated after initial design.**

Figure 2.5 Linkage groups on chromosome 2 and 4. Physical distances are shown by kb.



CHAPTER 3 - GRAIN QUALITY ANALYSIS OF THE SORGHUM DIVERSITY PANEL

Abstract

Grain sorghum has high photosynthetic efficiency and high water use efficiency as a C_4 crop. Its good tolerances to extreme environments and gluten-free statue make it an attractive grain. Grain quality traits are directly related to sorghum nutritional value and end use efficiency. To optimize its food and feed potential, evaluation of sorghum accessions is first step. A sorghum collection, sorghum *bicolor* panel (SB panel), with three hundred lines derived from sorghum conversion program or elite breeding lines and their progenitors was used for grain quality evaluation. The Single Kernel Characterization System (SKCS) and near infrared reflectance spectroscopy (NIRS) are employed for grain quality data collection. Results indicated the SB panel contains wide range of diversity traits which have high heritability and have a good potential to be used for association analysis.

Introduction

Sorghum is a C_4 crop native to tropical and subtropical regions and ranks as the fifth cereal crop in the world and third in United States (Sasaki and Antonio, 2009). Sorghum is among the most efficient crops in conversion of solar energy and use of water. It is also known as a high-energy, high yield potential, good stress tolerant crop to various environments including high temperature, drought, and poor soil condition. Its wide uses and adaptation make

its status in cereal crop family irreplaceable. For people in European and African where celiac disease is present, gluten-free grain sorghum is a good food substitute resource (Schober et al., 2005).

Most sorghum commercial varieties grown in United States are dwarf types which are less than 5 feet in height. This facilitate combine harvest but lead to diversity shortage (Hamblin et al., 2006) and sorghum lines similar susceptive to disease, pests and other environmental stress. The sorghum conversion program (SCP) initiated in 1963, introduced novel genetic variation from exotic, tropical germplasm into modern U.S. cultivars (Stephens et al., 1967). After crosses between tropical lines and elite lines, and subsequent backcross and selection for day-neutral flowering and reduced height, about 850 converted tropical lines with photoperiod insensitive, early maturing and short stature features were released. Recently, a sorghum diversity panel was recently assembled. This panel is comprised of 377 accessions, including 228 converted tropical lines produced by the SCP, and 149 important breeding lines and their progenitors from all around the US (Casa et al., 2008). A subset of 300 lines from this panel was selected for sorghum grain quality traits research by association analysis.

Grain quality includes a range of physical, sanitary, and intrinsic characteristics (Henry and Ketlewell, 2007). Factors including genetic inheritance, growing period, harvest timing, drying, storage management, and transportation procedures all affect grain quality. This research was focused on genetic inheritance traits classified as physical and intrinsic grain quality characteristics. For example, kernel hardness, size, weight, protein, fat, and starch content. The single kernel characterization system (SKCS) is used for whole grain physical properties measurements (Martin et al., 1993). Near infrared reflectance spectroscopy (NIRS) is employed

to provide rapid and nondestructive determination of the concentration of certain constituents in sorghum grain.

Grain quality traits are complex traits controlled by quantitative traits loci (QTL). Association mapping, also called linkage disequilibrium mapping, is a promising approach for quantitative trait locus discovery and dissection (Zhu et al, 2008). Compared with linkage analysis, association mapping have merits of time saving, high throughput, high extension, high possibility of detection of genes or QTLs with minor effects. Because populations used for association analysis are natural populations with complex historical backgrounds, population structure, kinship and other potential factors should be dissected to avoid biased analysis results. The objective of this study is to determine the genetic diversity of sorghum diversity panel with 300 lines.

Materials and Methods

Sorghum diversity panel

A sorghum diversity panel for association mapping was recently assembled. This panel was comprised of 377 accessions, including 228 converted tropical lines produced by the SCP, and 149 important breeding lines and their U.S. progenitors. We selected 300 lines from the panel to study grain quality traits. Accessions with name prefix of SC are derived from SCP. Accessions which are elite US lines are named using their PI number. Seeds of the accessions were obtained from the USDA-ARS, PGRCU and planted in random complete block design in Manhattan, KS and West Lafayette, IN with two replications in 2007 and 2008. Seeds were harvested from 10 selfed sorghum heads for grain quality traits analysis. The trait means were employed to dissect the availability of association analysis within this panel.

Single kernel characterization system (SKCS)

The single-kernel characterization system (SKCS) was designed to test 300 kernels in about 3 min and determine the sample average and standard deviation for several physical parameters. The SKCS singulates individual kernels, weighs them, and then crushes them between a toothed rotor and a progressively narrowing crescent gap. As a kernel is crushed, the force between the rotor and crescent and the conductivity between the rotor and the electrically isolated crescent are measured (Martin et al., 1993). The main parameters processed were hardness index, weight, moisture, and kernel thickness on an individual kernel basis. Mean of the four parameters and their standard deviations are then calculated from the single kernel data obtained on the entire 300 kernel sample. The SKCS was first used for wheat kernels and thereafter calibrated to be suitable for sorghum grain researches (Bean et al., 2006). Compare to traditional method for kernel traits data collection, SKCS has the merits of time saving, accuracy and automation. In this research, SKCS provided kernel hardness, size and weight determination for 290 accessions, in total 580 samples.

Near infrared reflectance spectroscopy (NIRS)

Near infrared reflectance spectroscopy (NIRS) is a rapid measurement technique utilizing spectrum near infrared region to determinate the concentration of most organic and some inorganic compounds in tissue. NIRS is a simple rapid, accurate, reliable, and repeatable method for analyzing several components of grains simultaneously .

NIRS works by passing an infrared radiation through a sample and directly measure and determine the amount of radiation that is absorbed at a particular wavelength. When the infrared

radiation is absorbed by organic molecule, the energy is converted into molecular motions called vibrations. These molecular vibrations would allow the absorption bands to give rise throughout most of the infrared region of the spectrum. NIRS directly measure the absorbance at each wavelength. Chemical functional groups with C-H, O-H, and N-H produced absorption bands easy to be detected under near-infrared spectrum (Delwiche et al., 1996). As a result, NIRS is effective to predict compounds containing these functional groups. NIRS provide predictions of starch, fat, protein, fiber, and phosphorus in grain and it is very useful when limited sample amount available. Sometimes, the molecular overtone and combination bands seen in the near-infrared spectra are very broad, which leads to complex spectra and increases the difficulty of assign specific features to specific chemical components. Utilizing calibration technique to careful develop a set of calibration equations is essential.

A total of 15 g of seeds were ground in a cyclone mill (UDY Corporation, Fort Collins, CO) with a 1-mm screen. A total of 538 samples (269 accessions) from this population were analyzed by NIRS. To obtain accurate data from NIRS, the system was calibrated based on values obtained from chemical analyses of a subset of samples. For developing calibration equations, 114 of the most informative grain samples were chosen. Grain samples were then analyzed for starch, fat, crude protein, and acid detergent fiber content by wet chemistry methods. Based on fiber, fat, and crude protein value, data of digestible nutriments metabolizable energy, net energy for gain, net energy for lactation, net energy for maintenance were also calculated.

Near infrared spectroscopy equations for each grain trait were developed. Trait values from a randomly selected group comprising 62 of the 114 samples were used to produce the calibration equations while values from a second group, the remaining 52 samples, were used to evaluate the derived equations. In all, 66 equations, each with different wavelengths and math

treatments, were tested for each trait. Based on low standard errors of prediction and high R^2, the equations that maximized the prediction of trait values were retained. The best calibration equation were then evaluated using the full subset of 11 samples, and was used to predict composition values from the NIRS spectra of all grain samples (n = 538).

Results and Discussions

The SKCS analysis data are consistent for years (Table 3.1 and Figure 3.1). The hardness and weight data are especially consistent in duplications of the same year. While year comparison, both the two parameters have slightly increased. For minimum diameter characters, the seeds obtained in 2008 were almost two times larger than the seed produced in 2007, which averaged 0.7 and 1.5 respectively. But there was no obvious change observed for maximum values. The standard deviations obtained for all the samples which were calculated based on 300 kernels. The maximum values for the standard deviations for hardness, weight and diameter are 35.977, 17.144, and 0.844 respectively, all within the normal ranges. Moisture data were also provided by the SKCS, however the predicted data was not accurate enough, so were excluded (Bean et al., 2006).

Kernel weight values varied from 15 to 40 ms or more in all the four duplications. Kernel diameter values also presented a wide range from 0.7cm to 3.2cm. Hardness is the most diverse parameter, with the maximum value of 126.46 and minimum value of 3.37. For standard test value for wheat, samples with hardness index of over 100 are classified as hard wheat, while kernels with hardness index less than 40 are considered as soft wheat (Pearson et al., 2009). The over 120 hardness index extension demonstrated the large diversity within the SB panel.

Consistency of the SKCS data for seed sets 2007 and 2008 were compared. The hardness, weight and diameter data for the two years had high correlation coefficient of 0.82, 0.71, and 0.73. Negative correlation coefficient observed for hardness and weight, and hardness and diameter, which indicated that hard sorghums tend to be smaller and lighter. Correlation of weight and diameter is not strong, with the correlation coefficient of only 0.59.

Also, based on the bleach test results of 2007 SB panel data, we found the sorghum accessions with lower hardness values were more likely to have tannin while accessions with higher hardness values tend to be non-tannin. A R^2 of 0.72 were observed for tannin and kernel hardness. Such relationship does not exist between tannin content and two other grain quality traits (kernel weight and kernel diameter).

In total data of 287 lines in the SB panel were obtained from the NIRS. All the chemical traits predicted by NIRS displayed a wide range of diversity. Fiber, Ca, protein, fat, p and starch contents ranged between 2.95-9.37, 0.04-0.08, 10.82-19.26, 2.26-4.35, 0.36-0.63, 61.74-74.38, respectively. The starch contents displayed large variations in SB panel, which had a standard deviation of 2.1. Ca content was the most consistent trait in the SB pane, with standard deviation of 0.06 (Table 3.3). Correlation of the six parameter obtained by NIRS is showed in Table 3.4.

NIRS provided good data with perfect consistence. High R^2 values were obtained for various traits using a validation set of 52 samples. The R^2 for starch, crude protein, fat, acid detergent fiber, and P content were 0.985, 0.884, 0.978, 0.914, 0.877, and 0.985 respectively. Based on fiber, fat, and crude protein values, parameters for feed use efficiency were also calculated: total digestible nutriments, metabolizable energy, net energy for gain, net energy for lactation, and net energy for maintenance have R^2 of 0.886, 0.895, 0.887, 0.887 and 0.887 respectively. Since all the predict R^2 values are higher than 0.85. The calibration equation proved

to be successful in indicating accurate test values for all the test characters. Out of the 62 random samples used for prediction, there was only one sample with a H value of 3.004, all other predicted samples showed a global H value less than 3.0. Based on all the statistical parameters mentioned above, NIRS was demonstrated efficient and accurate for chemical grain quality traits prediction of our SB panel.

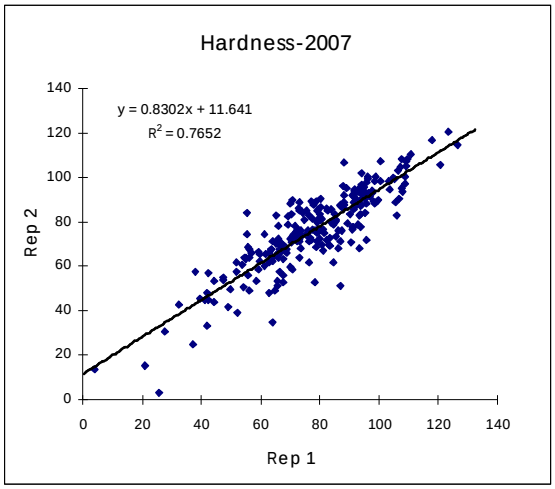
To dissect the potential relationship of physical and intrinsic traits, we analyzed the correlation of all the traits data obtained by both SKCS and NIRS (Table 3.3). Protein and P content has the highest correlation coefficient of 0.75, out of all the traits comparison. This is reasonable since P is the fundamental component of amino acid and protein. Negative correlation between protein and fat, protein and starch were observed, indicated that the total amount of the three essential nutritional components in sorghum kernel may be conserved in different sorghum varieties. Both protein and starch content positively related to kernel weight, this is consistent with the fact that protein and starch are the main nutritional compounds in sorghum kernels. Kernel hardness didn't show significant correlation with any of the other intrinsic traits.

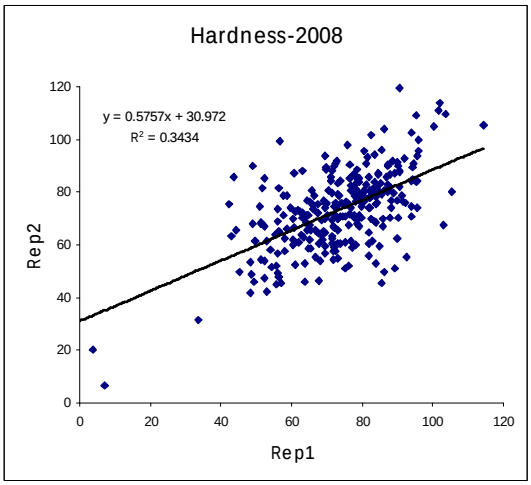
In future, SNP data and software including STRUCTURE, SPAGeDi, TASSEL will be employed for association analysis of these grain quality traits.

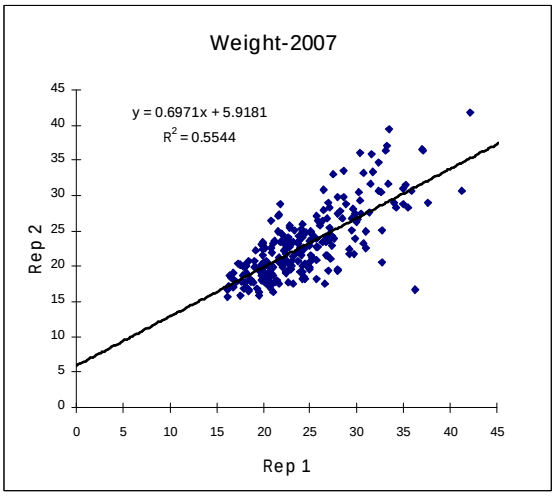
Table 3.1 Statistics for hardness, weight and diameter of the sorghum diversity panel.

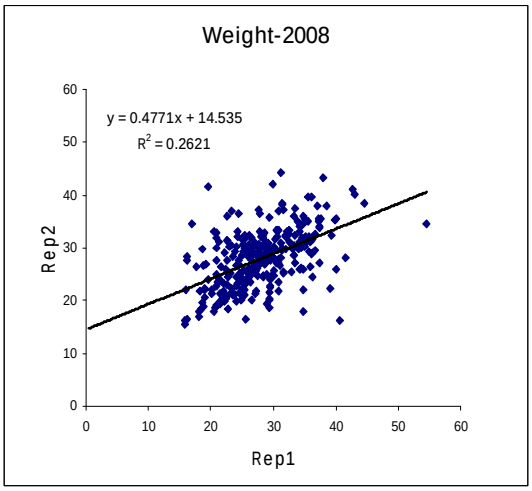
Variable	N	Mean	Std Dev	Minimum	Maximum
Hardness07	300	73.0	14.4	0.4	119.4
Weight07	300	27.9	5.4	15.7	48.4
Diameter07	300	2.2	0.3	1.5	3.1
Hardness08	290	78.2	18.2	8.7	122.4
Weight08	290	23.4	4.7	15.1	42.0
Diameter08	290	1.6	0.3	0.7	2.7

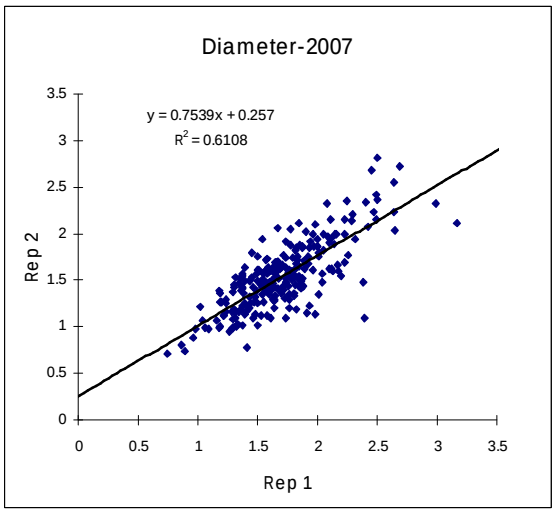
Figure 3.1 Correlation of data for kernel hardness, kernel weight, and kernel diameter in 2007 and 2008.











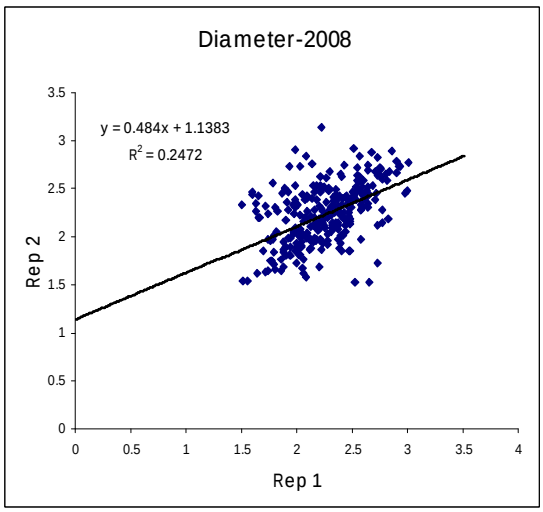


Table 3.2 SKCS data correlation coefficient for 2007 and 2008.

	Hardness 07	Weight 07	Diameter 07	Hardness 08	Weight 08	Diameter 08
Hardness 07		-0.25	-0.28	0.82	-0.21	-0.22
		<.0001	<.0001	<.0001	0.0004	0.0001
Weight 07	-0.25		0.88	-0.36	0.71	0.65
	<.0001		<.0001	<.0001	<.0001	<.0001
Diameter 07	-0.28	0.88		-0.36	0.59	0.73
	<.0001	<.0001		<.0001	<.0001	<.0001
Hardness 08	0.82	-0.36	-0.36		-0.31	-0.3
	<.0001	<.0001	<.0001		<.0001	<.0001
Weight 08	-0.21	0.71	0.59	-0.31		0.89
	0.0004	<.0001	<.0001	<.0001		<.0001
Diameter 08	-0.22	0.65	0.73	-0.3	0.89	
	0.0001	<.0001	<.0001	<.0001	<.0001	

Table 3.3 Statistics for NIRS of the sorghum diversity panel, 2007.

Variable	N	Mean	Std Dev	Minimum	Maximum
Fiber	287	5.08	0.86	2.95	9.37
Ca	287	0.06	0.01	0.04	0.08
Protein	287	14.61	1.44	10.82	19.26
Fat	287	3.18	0.35	2.26	4.35
P	287	0.46	0.04	0.36	0.63
Starch	287	68.86	2.10	61.74	74.38

Table 3.4 NIRS Data correlation for 2007.

	Fiber	Ca	Protein	Fat	P	Starch
Fiber		0.05	0.37	0.03	0.28	-0.73
		0.37	<.0001	0.57	<.0001	<.0001
Ca	0.05		0.22	0.52	0.34	-0.26
	0.37		0.0001	<.0001	<.0001	<.0001
Protein	0.37	0.22		-0.1	0.75	-0.72
	<.0001	0.0001		0.11	<.0001	<.0001
Fat	0.03	0.52	-0.1		0.12	-0.21
	0.57	<.0001	0.11		0.04	0
P	0.28	0.34	0.75	0.12		-0.63
	<.0001	<.0001	<.0001	0.0417		<.0001
Starch	-0.73	-0.26	-0.72	-0.21	-0.63	
	<.0001	<.0001	<.0001	0.0004	<.0001	

Table 3.5 SKCS and NIRS Data correlation for SB panel, 2007.

	Hardness	Weight	Diameter	Fiber	Ca	Protein	Fat	P	Starch
Hardness		-0.31	-0.3	-0.24	0.23	0.08	0.29	0.12	0.04
		<.0001	<.0001	<.0001	<.0001	0.18	<.0001	0.05	0.45
Weight	-0.31		0.89	-0.13	-0.04	0.08	-0.17	-0.03	0.05
	<.0001		<.0001	0.03	0.5	0.17	0	0.57	0.37
Diameter	-0.3	0.89		-0.12	-0.09	0.03	-0.21	-0.07	0.11
	<.0001	<.0001		0.04	0.11	0.64	0.0004	0.23	0.06
Fiber	-0.24	-0.13	-0.12		0.05	0.37	0.03	0.28	-0.73
	<.0001	0.03	0.04		0.37	<.0001	0.57	<.0001	<.0001
Ca	0.23	-0.04	-0.09	0.05		0.22	0.52	0.34	-0.26
	<.0001	0.5	0.11	0.37		0.0001	<.0001	<.0001	<.0001
Protein	0.08	0.08	0.03	0.37	0.22		-0.1	0.75	-0.72
	0.18	0.17	0.64	<.0001	0.0001		0.11	<.0001	<.0001
Fat	0.29	-0.17	-0.21	0.03	0.52	-0.1		0.12	-0.21
	<.0001	0.0033	0.0004	0.57	<.0001	0.11		0.04	0.0004
P	0.12	-0.03	-0.07	0.28	0.34	0.75	0.12		-0.63
	0.05	0.57	0.23	<.0001	<.0001	<.0001	0.04		<.0001
Starch	0.04	0.05	0.11	-0.73	-0.26	-0.72	-0.21	-0.63	
	0.45	0.37	0.06	<.0001	<.0001	<.0001	0.0004	<.0001	

References

Aboubacar, A., and Hamaker, B. 1999. Physiochemical properties of flours that relate to sorghum couscous quality. Cereal Chemistry 76:308- 313.

Abrahams, S., G. J. Tanner, P. J. Larkin, and A. R. Ashton. 2002. Identification and Biochemical Characterization of Mutants in the Proanthocyanidin Pathway in Arabidopsis. Plant Physiology 130:561-576

Akingbala, J. O., and L. W. Rooney. 1987. Paste properties of sorghum flour and starches. Journal of Food Processing and Preservation 11:13–24.

Armstrong W. D., J. G. Rogler, and W.R. Featherston. 1974. Effect of tannin extraction on the performance of chicks fed bird resistant sorghum grain diets. Poultry Science 53:714-720.

Awika, J.M., L.W. Rooney, and X. Wu. 2003. Screening Methods to measure antioxidant activity of sorghum (Sorghum bicolor) and Sorghum Products. J. Agric. Food Chem, 51:6657-6662

Awika, J.M. and L.W. Rooney. 2004. Sorghum phytochemicals and their potential impact on human health. Phytochemistry 65:1199-1221.

Bean, S.R., O.K. Chung, M.R. Tuinstra, J.F. Pedersen, and J. Erpelding. 2006. Evaluation of the single kernel characterization system (SKCS) for measurement of sorghum grain attributes. Cereal Chemistry 83:108–113.

Bettge, A.D. and Morris,C.F. 2000. Relationships among grain hardness, protein fractions and end-use quality of wheat. Cereal Chemistry 77:241-247.

Bhattramakki, M.D., J. Dong, A. K. Chhabra, and G. E. Hart. 2000. An integrated SSR and RFLP linkage map of Sorghum bicolor (L.). Genome 43:988–1002.

Blessin, G.W., R.A. Anderson, W.L Deatherage, and G.E. Inglett. 1971. Effect of alkali dehulling on composition and wet-milling characteristics of sorghum grain. Cereal Chemistry 48:528-532.

Boren B, and R.D. Waniska. 1992. Sorghum seed color as an indicator of tannin content. J. Appl. Poultry Res. 1:117-121.

Bramel-Cox, P.J., M.A. Lauver, and M.E. Witt. 1990. Potential gain from selection in grain sorghum for higher protein digestibility. Crop Science 30:521–524.

Buffo, A.R., C.L. Weller, and A. Parkhurt. 1998. Relationship among grain sorghum quality factors. Cereal Chemistry 75(1):100–104.

Cagampang, G. B., A. W. Kirleis, and J. S. Marks. 1984. Application of small sample back extrusion test for measuring texture of cooked sorghum grain. Journal of Food Science 49:278–280.

Cagampang, G. B., and Kirleis, A.W. 1984. Relationship of sorghum grain hardness to selected physical and chemical measurements of grain quality. Cereal Chem. 61:100-105.

Caldas G.V., and M.W. Blair. 2009. Inheritance of seed condensed tannins and their relationship with seed-coat color and pattern genes in common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics 119:131–142

Casa, A.M. 2008. Community resources and strategies for association mapping in sorghum. Crop Science 48:30-40.

Cisse N. 1995. Heritability estimates, genetic correlation, andidentification of RAPD markers linked to seedling vigor and associated agronomic traits in sorghum. Ph.D. thesis. Purdue University, West Lafayette

Debeaujon, I., N. Nesi, P Perez, M Devic, O Grandjean, M Caboche, and L Lepiniec. 2003. Proanthocyanidin-Accumulating Cells in Arabidopsis Testa:Regulation of Differentiation and Role in Seed Development. Plant Cell 15:2514-2530.

Delwiche, S.R., K.S. McKenzie, and B.D. Webb. 1996. Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. Cereal Chem. 73:257-263.

Deshpande, S.S., and M. Cheryan. 1986. Tannin analysis of food products. CRC Critical Reviews in Food Science and Nutrition 24:401–449.

Devic, M., J. Guilleminot, I. Debeaujon, N. Bechtold, E. Bensaude, E., M., Koornneef, G. Pelletier, and M. Delseny. 1999. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J. 19:387–398.

Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2005. Maize association mapping population:a high resolution platform for quantitative locus dissection. Plant J. 44:1054-1064.

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Genetics 3:43-52.

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

Dykes, L., J.M. Awika, C.M. McDonough. L.W. Rooney, and R.D. Waniska. 2002. False positives for tannin sorghum in non-tannin sorghum using the bleach test.

Dykes, L. and L. W. Rooney. 2006. Sorghum and millet phenols and antioxidants. J. Cereal Sci. 44:236-251.

Earp, C. F., J. O. Akingbale, S. H. Ring, and L.W. Rooney. 1981. Evaluation of several methods to determine tannins in sorghum with varying kernel characteristics. Cereal Chem. 58 (3):224-238.

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

Flint-Garcia, S.A., A.C. Thuillet, J. M. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Friend, J. 1981. Plant phenolics, significance and plant disease. Progress in Phytochemistry 4:197.

Friend, J. 1977. Phenolic substances and plant diseases. Recent Adu.Phytochem.12:557. Giner-Chavez, B.I. 1996. Condensed tannins in tropical forages - Ph. D. Thesis. Cornell University, Ithaca, NY, USA.

Gabriel, K. 2005. A study of heterotic relationships in sorghum. Ph.D. diss. Texas A&M Univ., College Station (Diss. Abstr.3202314).

Hagerman, A.E., Y. Zhao, and S. Johnson. 1997. Methods for determination of condensed and hydrolyzable tannins. American Chemcial Society Symposium Series 662:209–222.

Hagerman, A. E., Riedl, K. M., Jones, G. A., Sovik, K. N., Ritchard, N. T., and Hartzfeld, P. W. 1998. High molecular weight plant polyphenolics (tannins) as biological antioxidants. Journal of Agricultural and Food Chemistry 46:1887–1892.

Hahn, D.H., J.M., Faubion. and L.W. Rooney. 1983. Sorghum phenolic acids, their performance liquid chromatography separation and their relation to fungal resistance. Cereal Chemistry 60:255–259.

Hahn, D. H., L.W. Rooney, and C. F. Earp. 1984. Tannins and phenols of sorghum. Cereal Foods World 29(12):775-779.

Hamblin, M.T., M.G. Salas Fernandez, A.M. Casa, S.E. Mitchell, A.H. Paterson, and S. Kresovich. 2005. Equilibrium processes cannot explain 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. 2007. Sequence variation at candidate loci in starch metabolism pathway in sorghum:Prospect for linkage disequilibrium mapping. Crop Science. 47:S125-S134.
Harris, H. B. and R.E. Burns. 1970. Influence of tannin content on pre-harvest seed germination in sorghum. Agronomy Journal 62:835-836.
Harris, H. B. and R.E. Burns. 1973. Relationship between tannin content of sorghum grain and pre-harvest seed molding. Agronomy Journal 65:957-959.
Haslam E. 1989. Plant polyphenols. Cambridge University Press, Cambridge, UK. Chemistry and significance of condensed tannins. Ed. by Hemingway R.W. and Karchesy J.J. Plenum Press, New York.
Henry, R. J., and P. Kettlewell. 2007. Cereal Grain Quality. Kluwer Academic Publishers Jambunathan, R., M. S. Kherdekar, and J. W. Stenhouse. 1992. Sorghum grain hardness and its relationship to mold susceptibility and mold resistance. J. Agric. Food Chem. 40:1403-1408.
Hyten, D.L., Q. Song, Y. Zhu, I.Y. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker, and P.B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA 103:16666–16671.
Johnson, C. S., B. Kolevski, and D.R. Smyth. 2002. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell 14:1359-1375.
Kaluza, W.Z., R.M. McGrath, T.C. Roberts and H.H. Schroder. 1980. Separation of phenolics of Sorghum bicolor (L.) Moench grain. Journal of Agricultural and Food Chemistry 28:1191-1196.
Knoll, J.,N. Gunaratna, and G. Ejeta. 2008. QTL analysis of early-season cold tolerance in sorghum. Theor Appl Genet. 116:577–587.
Kofoid, K.D., J.W. Maranville, and W.M. Ross. 1978. Use of a Bleach Test to Screen Single-head Sorghum Selections for the Presence of a Testa Layer. Agron J. 70:775-779.
Koornneef, E., M., G. Pelletier, and M. Delseny. 1999. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J. 19:387–398.
Krrakman A.T.W., R.E. Nika, P. Stam, and F.A.V. Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446.
Kumar, R, and M. Singh. 1984. Tannins, their adverse role in ruminant nutrition. Journay Agricultural food Chemistry 32:447.
Labieniec, M, and T. Gabryelak. 2003. Effects of tannins on Chinese hamster cell line B14. Mutation Research.539:127–135.
Lander, E., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.
Lee, W., J. Pedersen, and D. Shelton. 2002. Relationship of sorghum kernel size to physiochemical, milling, pasting, and cooking properties. Food Res. Int. 35:643-649.

Li, M., NYuyama, L Luo,M. Hirata, and H. Cai. 2009. In silico mapping of 1758 new SSR markers developed from public genomic sequences for sorghum. Molecular Breeding 24:41–47.

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs a¤ecting plant height and maturity across the Poaceae, in reference to an interspecies sorghum population. Genetics 141 :391-41.1

Lipsa F. D., R. J. Snowdon, W. Fried. 2009. QTL analysis of condensed tannins content in Brassica Napus L. Research Journal of Agricultural Science 41 (2):274-279.

Martin, C. R., and Steele, J. L. 1996. Evaluation of rotor-crescent design for sensing wheat kernels hardness. Trans. ASAE 39:2223-2227.

Maxson E.D., and L.W. Rooney. 1972. Two methods of tannin analysis for Sorghum bicolor (L.) Moench grain. Crop Science 12:253.

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 Molecular Biology 48: 483–499.

Meuwissen, T.H.E., and M.E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics 155:421-430.

Mueller, L.A., C.D. Goodman, R.A. Silady, and V. Walbot. 2000. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid binding protein. Plant Physiol. 123:1561–1570.

NASS. 2008a. 2008 Acreage Summary [Online]. National Agricultural Statistics Service. http://www.nass.usda.gov/QuickStats/indes2.jsp. Accessed: October 17, 2009.

NASS. 2008b. Kansas 2008 Crop Production Summary [Online]. National Agriculture Statistics Service. http://www.nass.usda.gov/QuickStats/PullData_US,jsp. Accessed: October 17, 2009.

Nesi, N., I. Debeaujon, M. Caboche, and L. Lepiniec. 2001. The Arabidopsis TT2 gene encodes an R3R3 MYB domain protein that acts as a key determinant for the proanthocyanidin accumulation in developing seed. Plant Cell 13:2099–2114.

Paterson A. H. et al. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457:551-556.

Pearson, T. C., D. L. Brabec, and H. Dogan. 2009. Improved discrimination of soft and hard white wheat using SKCS and imaging parameters. Food Qual. 3:89–99.

Reed, J.D. 1995. Nutritional toxicology of tannins and related polyphenols in forage legumes. J. Anim. Sci. 73:1516-1528.

Ridder, D. 2005. Early-season cold tolerance in grain sorghum: the relationship with seed characteristics and the evaluation of molecular tools for breeding. Master thesis. Kansas State University

Rooney, L.W. and F.R. Miller. 1982. Variation in the structure and kernel characteristics of sorghum. In: Proceedings of the International Symposium on Sorghum Grain Quality, Oct. 28-31, 1981.

Rooney, L. W., and Waniska, R. D. 2000. Sorghum food and industrial utilization. Pages 689-750 in: Sorghum: Origin, History, Technology, and Production. C. W. Smith and R. A. Frederiksen, eds. John Wiley and Sons. New York.

Rooney, L.W., and J.M. Awika. 2004. Specialty sorghums for healthful foods. in:Specialty Grains for Food and Feed. Abdel-Aal, E. and Wood, P. (Eds.). American Association of Cereal Chemists. St. Paul, MN. pp. 283-312.

Sagasser, M., G.-H. Lu, K. Hahlbrock, and B. Weisshaar. 2002. A. thaliana TRANSPARENT TESTA1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes Dev. 16:138–149.

Salunkhe, D. K., J. K. Chavan, S.S. Kadam. 1990. Dietary tannins:Consequences and remedies; CRC Press: Boca Raton, FL.

Salvi S., and R. Tuberosa. 2005. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10:297-304.

Sasaki, T., and B.A.Antonio. 2009. Sorghumin sequence. Nature 457:547–548.

Sastry, L.V.S., J.W. Paulis, J.A. Bietz & J.S. Wall, 1986. Genetic variation of storage proteins in sorghum grain: Studies by isoelectric focusing and high-performance liquid chromatography. Cereal chemistry 63:420–427.

Schober, T. J., M. Messerschmidt, S. R. Bean, S Park, and E. K. Arendt. 2005.Gluten-Free Bread from Sorghum: Quality Differences Among Hybrids. Cereal chemistry 82:394-404.

Schulze, T.G., and F. J. McMahon. 2002. Genetic Association Mapping at the Crossroads:Which Test and Why? Overview and Practical Guidelines. American Journal of Medical Genetics 114:1-11.

Sell D., and J.C. Rogler. 1984. The effects of sorghum tannin and methionine levels on the performance of laying hens maintained in two temperature enviroments. Poultry Science 4:63-109.

Shirley, B.W., W.L. Kubasek, G. Storz, B E.ruggemann, M. Koornneef, F.M. Ausubel, and Goodman, H.M. 1995. Analysis of Arabidopsis mutants deficient inflavonoid biosynthesis. Plant J. 8:659 – 671.

Stephens, J.C. 1924. A second factor for subcoat in sorghum seed. Journal of the American Society of Agronomy. 25:340-342.

Stephens, J.C., F.R. Miller, and D.T. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Science 7:396.

Subudhi, P.K. D.T. Rosenow, and H.T. Nguyen. 2000. Quantitative trait loci for the stay-green trait in sorghum (Sorghum bicolor L. Moench): consistency across genetic backgrounds and environments. Theor. Appl. Genet. 101:733–741.

Tanaka J. and S. Yamaguchi 1996. Use of RAPD markers for the identification of parentage of tea cultivars. Bulletin of the National Institute of Vegetables, Ornamental Plants & Tea, Japan 9(B):31–36.

Tanksley, S.D., S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066.

Walton, M. F., F. A. Haskins and H. J. Gorz.1983. False Positive Results in the Vanillin-HCl Assay of Tannins in Sorghum Forage. Crop Science 23:197-200.

Wang Y.S., B.J. Barratt, D.G. Clayton, and J.A. Todd. 2005. Genome wide association study: theoretical and practical concerns. Genetics 6:109-119.

Waniska, R.D., L.F. Hugo, and L.W. Rooney. 1992. Practical methods to determine the presence of tannins in sorghum. J. Appl. Poultry Res. 1:122-128.

Waterman, P.G., and S. Mole. 1994. Analysis of phenolic plant metabolites. Blackwell Scientific Publications, Oxford, UK.

Wills. R.B.H and Ali, M.R. 1983. Effect of grain size on dehulling of sorghum. Cereal Chemistry 60:12-14

Wittkop, B., F. D. Lipsa, R. J. Snowdon, and W. Friedt. 2006 . Mapping of QTL for fibre and tannin compounds associated with meal quality in yellow-seeded oilseed rape. Pflanzenzüchtung und Genomanalyse 24:124-128

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

Churchill, G. A., and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Zeng, Z.-B., J. Liu, L.F. Stam, C.-H. Kao, J.M. Mercer, and C.C. Laurie. 1999. Genetic architecture of a morphological shape difference between two Drosophila species. Genetics 138:963–971.

Zhang, F., A. Gonzalez, M. Zhao, C.T. Payne and A. Lloyd. 2003. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130:4859-4869

Zhang, L.L., and Y.M. Lin. 2008. Tannins from Canarium album with potent antioxidant activity. Journal of Zhejiang University Science 9(5):407–415.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. The Plant Genome 1:5-20.