THE IMPROVEMENT OF GRAIN SORGHUM PRODUCTIVITY, BLACK PERICARP COLOR, AND PROTEIN DIGESTIBILITY

A Dissertation

by

BRIAN KEVIN PFEIFFER

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	William L. Rooney
Committee Members,	Seth C. Murray
	Joseph M. Awika
	John E. Mullet
Head of Department,	David D. Baltensperger

December 2017

Major Subject: Plant Breeding

ABSTRACT

Sorghum [*Sorghum bicolor* (L.) Moench] is an important cereal crop for the semiarid regions of the world. The improvement of sorghum grain production and sorghum grain quality are essential for the future of the crop. This study explores the genetic gain of hybrid sorghum breeding programs and investigates the nature of two important grain quality traits. A multi-environment trial was conducted that included hybrids ranging from the 1950s to today. Genetic yield gains in grain sorghum are increasing at approximately .008 t ha^{-1} annually and many other physiological traits have demonstrated changes as well. Also, a F_5 recombinant inbred line (RIL) population segregating for black and red pericarp color was evaluated in two Texas environments in the summer of 2017. Heritability and repeatability estimates were calculated and secondary plant metabolite biosynthesis genes were explored using RNAseq. The black pericarp trait appeared to be moderately heritable. Many of the same genes involved in 3-deoxyanthocyanidins (3-DOA) production in leaf tissue during fungal invasion appeared upregulated in black pericarp sorghum versus red pericarp sorghum. Finally, sorghum grain protein digestibility was investigated in a biparental mapping population and evaluated in two environments for two years. A near-infared spectrometry (NIR) calibration curve for *in-vitro* protein digestibility was developed to assist the phenotyping of this important quality trait. The protein digestibility trait was highly heritable and a significant genotype x environment effect was observed. Understanding key components of genetic gains in yield and other traits, as well as factors involved

with important quality traits will benefit sorghum production in the future.

DEDICATION

To my grandparents, Robert and Rose Alice Hughes, and Harold and Janet Pfeiffer. Each one inspired and continue to inspire me to achieve my goals.

ACKNOWLEDGEMENTS

Thank you to all my teachers since the very beginning of my educational career, including all of those from Tri-City Schools, Iowa State University, and of course Texas A&M University. I must extend appreciation to my committee chair, Dr. William L. Rooney, for the opportunity to conduct research in his program for both my M.S. and doctoral degrees. Dr. Rooney has had the tough job of guiding me through the entire process of graduate school. Without accepting me into his lab, his investment of time, and his guidance, my future career in plant breeding would not be realized. Dr. Rooney has been an excellent mentor and role-model to me.

I would also like to thank my other committee members, Dr. Seth Murray, Dr. Joseph Awika, Dr. John E. Mullet, for their guidance, insight, and support throughout this research. Thank you also to Dr. Brian W. Gardunia, special appointment on my graduate committee, and a tremendous mentor in my professional career.

Also, thanks to my fellow graduate students, past and present, in the Texas A&M AgriLife Research Sorghum Breeding Lab: Dr. Dustin Herb, Dr. Matthew Bartek, Kyle Burns, Dr. Geraldo Carvalho Júnior, Francisco Gomez, Luke Vacek, Melissa Ganotis, Ace Pugh, Dr. Lloyd Mbulwe, Bethany Andrews, David Horne, Ammani Kyanam, Lorin Harvey, and Jales Fonseca for their collaboration and camaraderie. I would especially like to acknowledge Dr. John Gill, whose leadership and experience allowed him to serve as a mentor to me.

Thanks also go to my friends and colleagues that I worked with on the Texas A&M Plant Breeding Symposium and the Soil & Crop Sciences Graduate Organization. I especially would like to recognize Laura Masor, Sean Carver, Laura Ann McLoud, Dustin Wilkerson, Ammani, and Francisco. I had a blast working hard, together to make our school a better place.

Conducting plant breeding research is a cooperative effort. The support I received from Delroy Collins, Dr. Leo Hoffmann Jr., Steve Labar, Vickie Horn, and the many hard-working undergraduate student workers should not go without mention.

Thank you to my friends outside of the Soil & Crop Sciences Department for making my time at Texas A&M University a great experience. Some of these individuals include Katie Snyder, Jessica Rodriguez, Eric Gasper, Mark Lenderman, Jana Lenderman, JJ Lenderman, and Travis Halm. These friends have encouraged, entertained, and supported me throughout the past two years. I greatly value their friendship, and I sincerely appreciate their belief in me.

Finally, a big thank you to my parents, Roger and Melissa, and brother, Ryan, for all their encouragement and support. I would never have achieved this without them. I am so grateful for everything they have given me, taught me, and sacrificed for me.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professors William L. Rooney (advisor), Joseph M. Awikia, and Seth C. Murray of the Department of Soil & Crop Sciences and Professor John E. Mullet of the Department of Biochemistry and Biophysics.

The experiments in Section 2 were performed in conjunction with Dr. Cleve Franks and Dr. Roger Monk of DuPont Pioneer. Dennis Pietsch and Dr. Ronnie Schnell also contributed field space, planting, and harvesting, etc. for multiple locations through the Texas A&M AgriLife Research Variety Testing Program.

Dr. Robert Klein, Dr. Patricia Klein, Dr. Nikhil Patil, Dr. Barbara Barbosa, and Elmer Roldan Salazar of Texas A&M University contributed to genetic resources utilized in Section 3 of this dissertation.

The experiments in Section 4 were performed in conjunction with Dr. Stephen Kresovich and Dr. Richard Boyles of Clemson University. The Clemson University team provided the South Carolina environments and coordinated the genotyping of the RIL population. Also, Dr. Scott Bean of the USDA-ARS in Manhattan, Kansas provided initial phenotypic data for in-vitro protein digestibility. Dr. Geraldo Carvalho Jr. and Dr. Leo Hoffmann Jr. of Texas A&M University assisted with the creation of NIR calibration curves for the high protein digestible trait.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All other work for the dissertation was completed independently by Brian Pfeiffer.

NOMENCLATURE

1-VR	Coefficient of determination for cross-validated values
4CL	4-coumarate:CoA ligase
*a**	Redness colorimeter value
ANS	Anthocyanidin synthase
*b**	Blueness colorimeter value
BLUP	Best linear unbiased predictor
C4H	Transcinnamate 4-monooxygenase
CHI	Chalcone isomerase
CHS	Chalcone synthase
DFR	Dihydroflavonol 4-reductase
EDGE	Extraction of Differential Gene Expression
F3H	Flavanone 3-hydroxylase
F3'H	Flavonoid 3'hydroxylase
FSII	Flavone synthase II
G x E	Genotype x environment interaction
HD	High digestibility trait
IVPD	*In-vitro* protein digestibility
*L**	Lightness colorimeter value
LD	Linkage disequilibrium
LS	Least squares

NIR	Near-infrared spectroscopy
PAL	Phenylalanine ammonia lyase
PDCAAS	Protein Digestibility Corrected Amino Acid Score
r^2	Coefficient of determination
RIL	Recombinant inbred line
SEC-V	Standard error for cross validation
SEL	Standard error for the laboratory
SC14	Environment located in South Carolina in the summer of 2014
St dev	Standard deviation
TX14	Environment located in Texas in the summer of 2014
TX15	Environment located in Texas in the summer of 2015
QTL	Quantitative trait loci

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

This dissertation is broken into three major research projects and follows two key themes involving grain sorghum research. The first theme is about what's changed over sixty years of hybrid sorghum breeding. The second theme is about improving grain sorghum for human nutritional value and grain quality.

The first research project, described in section 2, investigates the history and future state of hybrid sorghum breeding. In most major field crops, era studies have been conducted to determine the genetic progress that is being achieved. However, genetic gain in sorghum breeding efforts is currently unknown. Old and new hybrids were recreated to represent the entire span of the Texas A&M AgriLife Research Sorghum Breeding Program, and these hybrids were evaluated together in a multi-environment trial. The genetic yield gain over time was thus was calculated. Many physiological trait modifications also accompanied changes in yield. Overall, it appeared that sorghum had a slower rate of genetic yield gains than in other crops.

The second and third research projects focus more on grain quality and human nutritional traits. Section 3 describes my project on the transcriptome analysis of the black pericarp trait in sorghum. The black pericarp trait is of interest for food processors and consumers because of its unique color and high concentration of antioxidants. Despite the growing interest in black sorghum, breeding efforts have been difficult because the genetics of the trait appear complex. Since the black pericarp trait has such low frequency, a large F_2 population was grown, and individuals were selected based on

phenotype. Individuals were placed into black pericarp or red pericarp bulks. These lines were then selfed to the F_5 generation in which they were phenotyped in two environments and genotyped. From transcriptome analysis of a red and black pericarp line, many of the same genes involved in 3-DOA production in leaf tissue during fungal attack appear to be upregulated in the black pericarp sorghum line versus the red pericarp line.

The fourth section of the dissertation focuses on the genetic mapping of the high protein digestibility trait in grain sorghum. Compositionally, sorghum grain contains similar levels of starch and protein as other major cereal crops like wheat and corn (Bansal et al., 2008), but has diminished levels of digestible protein. Fortunately, some sorghum genotypes are known to be high in protein digestibility. A 287 individual RIL population between BTxArg-1-1 and the highly digestible line P850029 was created and evaluated in two environments for two seasons. Percent protein digestibility was phenotyped using a combination of lab-based assays and NIR predictions.

2. LONG-TERM SELECTION IN HYBRID SORGHUM BREEDING PROGRAMS

Estimating genetic gains in sorghum [*Sorghum bicolor* (L.) Moench] is necessary to review past and present research and to determine whether the current rates of improvement will meet future production demands. The current study was conducted to determine the rate of genetic gain in yield and associated traits over the commercial hybrid era using sorghum germplasm from the Texas A&M and DuPont Pioneer sorghum breeding programs. Sixty hybrids that represented a 50-year span of hybrid breeding and their respective parental lines from the Texas A&M AgriLife Research Sorghum Breeding Program were grown and evaluated in five environments across Texas in 2016. In addition, fourteen DuPont Pioneer hybrids that represented the same time span were evaluated in three Texas environments. In both programs, grain sorghum yields increased .008 t ha^{-1} annually over their respective time spans. Traits that increased over time included yield potential per plant, heterosis, test weight, panicle size, and grain number per panicle while leaf angle, days to maturity, plant height, and yield stability demonstrated little to no change. Overall, approximately 60% of total yield gains in U.S. sorghum production are attributed to genetic improvement through sorghum breeding. Compared to other major U.S. field crops, the rate of genetic gain in sorghum has been slower due to a combination of factors, which include continually shifting production environments, changing priorities in traits, reduced research investments (compared to other crops), and less than optimized heterotic groups.

2.1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a major crop in the semiarid regions of the tropics and subtropics that is commonly grown in stressful environments with reduced inputs (Monk et al., 2014). Despite those limitations, sorghum has demonstrated production gains over time due to enhanced farming practices and improved genetics and plant breeding. Defining the rate of yield gains in sorghum is important because the future potential of this crop has a profound impact on growers, land owners, input suppliers, scientists, investors, policy makers and others.

Sorghum has been grown in the United States as a commercial grain and forage crop for more than 150 years. Area planted to sorghum peaked in the United States in 1957 at 10.88 million ha and they have declined since that time (USDA-NASS). In 2016, Americans planted 2.71 million hectares to sorghum representing a 75% decrease over the past 59 years. When the USDA began keeping records on sorghum in the 1930s, national grain yields averaged less than 1 million t ha^{-1}. In 2016, sorghum growers had their best year in history, averaging 4.89 t ha^{-1}. This difference represents a 450% increase since 1929 or an addition of .048 t ha^{-1} on average each year (USDA-NASS).

Unlike maize (*Zea mays* L.), where national on-farm annual yield averages have increased steadily over time, rates of gain have been highly variable in sorghum. Like maize, sorghum hybrids were not always available to farmers; until the late 1950s, growers produced cultivars and yields improved from roughly .5 t ha^{-1} to 1.5 t ha^{-1} between 1929-1956. With the development of cytoplasmic male sterility systems, commercial sorghum hybrids became available in 1957; they were rapidly adopted

(Bennett et al., 1990). Production increased and yields jumped sharply (+ .121 t ha^{-1} annually) by nearly 80%. The commercialization of the crop increased investment in both the public and private sectors. By the 1960s there were 18 U.S. public universities with sorghum research programs and the private sector was also healthy (Monk et al., 2014). Between 1966-1975, yield gains slowed (+ .005 t ha^{-1} annually) as biotic stresses, such as greenbug [*Schizaphis graminum* (Rondani)], downy mildew (*Peronosclerospora sorghi*), and *Maize dwarf mosaic virus* became economic pests in sorghum growing regions. Breeding for tolerance to these pests became more important than yield potential per se (Bennett et al., 1990).

Productivity increased again between 1975 and 1984 (+ .033 t ha^{-1}) primarily because of partnerships developed during the previous ten years, but between 1985 and 1994 it stagnated again (+.003 t ha^{-1}) for various reasons (USDA-NASS). A major factor included government policy. In 1985, American agricultural policies (i.e., crop insurance, farm bill payments) pushed sorghum from high performing hectares to more marginal growing regions and reduced area planted to the crop. Sorghum was displaced by maize on the more favorable land (Monk et al., 2014).

By the mid-1990's, the reduction in planted area had the concomitant effect of reducing the research investment in grain sorghum breeding, resulting in the consolidation and reduction of breeding programs in both the public and the private sector. For example, the number of public sector programs with long-term sorghum breeding programs dropped to four (Texas A&M, Kansas State, Nebraska, and Purdue) and the number of private companies dropped as well (Monk et al., 2014). Sorghum

areas were dramatically reduced in Nebraska and Missouri, states that formerly grew sorghum extensively. Despite shrinking research expenditures, average yields have indicated modest gains, increasing + .016 t ha^{-1} annually between 1995 to 2016 (USDA-NASS).

In Texas, sorghum production occurs both irrigated and non-irrigated/rainfed environments with a greater proportion of the area being rainfed. From 1972 to 2016, irrigated hectares planted in Texas have fallen from 800,000 to 111,000. During the same period, rainfed hectares have dropped from 1.95 million to 658,000. (USDA-NASS). As an example of displacement, in the same period, corn hectares in Texas have risen from 304,000 to 1.17 million. The year 2016 marked the first time in history that more area in Texas was planted to corn than sorghum, with 1.17 million hectares of corn to 769,000 hectares of sorghum planted (USDA-NASS).

On-farm yield gains in sorghum are lower than other major U.S. field crops. Since 1970, U.S. sorghum yields improved 0.56% annually (USDA-NASS). In contrast, corn, with the highest rate of yield gains of any major field crops, increased in yield by 2.31% annually. Other crops with commercialized transgenic traits have increased at elevated rates, including cotton (2.13% annually) and soybean (1.83% annually). Wheat, a non-transgenic crop commonly grown on marginal lands, has improved by 1.17% annually (USDA-NASS).

National on-farm annual yield estimates provide useful information to crop performance trends, but these values are a product of both plant breeding and agronomic practices. Plant breeders need to assess the relative rates of gain to understand priorities

in a program. Such studies are presented for most agronomic crops (Miller and Kebede, 1984; Russel, 1991; Bubeck et al., 2006; Schwartz and Smith, 2008; Campbell et al., 2011; Haegele et al., 2013; Liu et al., 2013a). These era studies, evaluate genotypes from different eras under equal environment and management conditions are useful to establish the relative value of plant breeding effort in yield gains (Duvick, 2005a; 2005b; 1984; Duvick et al., 2004; Smith et al., 2004)

Duvick (2004) estimated the rate of genetic yield gains to be .077 t ha^{-1} in DuPont Pioneer's corn breeding program. A more recent estimate indicated an increase of .0876 t ha^{-1} yr^{-1} for DuPont Pioneer hybrids released from 1930 to 2011 (Smith et al., 2014). Genetic yield gain estimates in cotton suggest yields have improved by .0216 t ha^{-1} annually from 1981 – 2011 (Campbell et al., 2014). In soybeans, genetic yield gains have improved by .0231, .0228, and .0195 t ha^{-1} yr^{-1} in maturity group (MG) II, MG III, and MG IV, respectively (Specht et al., 2014).

In corn, in addition to grain yield improvement, other traits also changed over time. Some of these trait changes were intentional and expected (i.e., yield components) while other traits were intentionally not changed (i.e., plant height). Several traits unexpectedly changed over time due to indirect selection for selection of high yielding hybrids (i.e., acute leaf angle). Finally, some traits did not change despite the breeders' best intentions to alter them (i.e., yield potential per plant in corn) (Duvick, 2005).

Compared to other crops, there are far fewer studies examining the changes occurring due to long-term selection within sorghum breeding programs (Monk et al., 2014). Miller and Kebedee conducted a study that compared the then "new" hybrids of

1980 to hybrids of the 1950s. They found that yield increased 48% between the old and new material. The study also determined that the number of kernels per panicle, total plant weight, plant height, leaf area, and stay-green have improved over time in the breeding program (Miller and Kebede, 1984). However, the year 1980 represents the halfway point today in sorghum breeding for hybrids. A more recent study measured genetic gain in Advanta Seeds (Advanta Semillas) Grain Sorghum Breeding Program in Argentina which tested hybrids released from 1984 to 2014. The study found that grain yields improved + .0087 t ha^{-1} on average each year (Gizzi and Gambin, 2016).

Estimates of the percentage of yield improvement due to the contribution of genetics vary from experiment to experiment. In a recent study, Assefa and Staggenborg (2011) estimated the contribution of plant breeding to be as high as 60-65%. Miller and Kebede (1984) indicated the contribution of genetics was only 40%. Unger and Baumhardt (1999), focusing on yield under dryland conditions, measured the contribution of genetics at 46%.

For sorghum to remain a competitive crop for farmers, we must evaluate past yield gains to predict future gains from selection and assess the progress of past and present selection strategies. This study evaluates the effects of long-term selection in large, public and private sorghum breeding programs. The objectives of this study were to i) determine the genetic gain for grain yield; ii) measure changes in physiological traits that have accompanied changes in yield gains; iii) estimate changes in heterosis and hybrid and inbred performance; iv) estimate changes in yield stability; and v) determine the contribution of breeding and genetics to total yield gains.

2.2 Materials and Methods

2.2.1 Plant Material

Two distinct sets of era hybrids were developed for this study. First, hybrids representing different periods within the Texas A&M AgriLife Research Sorghum Breeding Program were produced from historically important inbred lines in the breeding program. In total, 60 hybrids were selected for planting (Table 1) and these represented release dates from 1959 to 2014. Of the hybrids, twenty were produced from parental lines of common release years (i.e., 1977 x 1976). The other 40 hybrids were derived combinations of older x newer lines (i.e., 1961 x 1999). In this study, year of hybrid release was determined by averaging the year of release of the two corresponding parental lines. In addition to the hybrids, the 20 inbred lines that were parents of the 60 historic hybrids were grown in this study for evaluation of heterosis. Second, a set of 14 hybrids from the DuPont Pioneer sorghum improvement program with release dates from 1961 to 2015 was also produced for evaluation (Table 2).

2.2.2 Experimental Design

Hybrid and inbred line evaluation was accomplished using seven diverse locations in Texas during the summer of 2016. The 60 Texas A&M hybrids and 20 Texas A&M inbreds were evaluated in Monte Alto, College Station, and Perryton, Texas. At Thrall and Gregory, Texas only subset of 20 Texas A&M hybrids with parental lines of a similar release date was planted. Finally, the 14 Pioneer hybrids were grown in Corpus Christi, College Station, and Halfway, Texas.

Table 1. Texas A&M Plant Material.

List of plant material tested in this experiment from the Texas A&M AgriLife Research sorghum breeding program. The information includes year of parental line release and publications describing the material.

Genotype	Type	Female Year	Male Year	Release Average	Environments	Reference
ATx3197/RTx7000	Hybrid	1961	1957	1959	5	
ATx3197/RTx2536	Hybrid	1961	1964	1962.5	5	
ATx378/RTx7000	Hybrid	1961	1957	1959	5	
ATx378/RTx2536	Hybrid	1961	1964	1962.5	5	
ATx623/RTAM428	Hybrid	1977	1974	1975.5	5	
ATx623/RTx430	Hybrid	1977	1976	1976.5	5	
ATx2752/RTAM428	Hybrid	1976	1974	1975	5	
ATx2752/RTx430	Hybrid	1976	1976	1966.5	5	
ATx626/RTx2783	Hybrid	1986	1982	1984	5	
ATx626/RTx435	Hybrid	1986	1984	1985	5	
ATx631/RTx2783	Hybrid	1985	1982	1983.5	5	
ATx631/RTx435	Hybrid	1985	1984	1984.5	5	
ATx2928/RTx436	Hybrid	1999	1992	1995.5	5	
ATx2928/RTx437	Hybrid	1999	1999	1999	5	
ATx645/RTx436	Hybrid	2001	1992	1996.5	5	
ATx645/RTx437	Hybrid	2001	1999	2000	5	
A05071/R08304	Hybrid	2005	2015	2010	5	
A05071/R07178	Hybrid	2005	2015	2010	5	
A07118/R08304	Hybrid	2007	2015	2011	5	
A07118/R07178	Hybrid	2007	2015	2011	5	
ATx626/R08304	Hybrid	1986	2015	2000.5	3	
A07118/RTx430	Hybrid	2007	1976	1991.5	3	
ATx2928/R07178	Hybrid	1999	2015	2007	3	
A07118/RTx436	Hybrid	2007	1992	1999.5	3	
A05071/RTx437	Hybrid	2005	1999	2002	3	
ATx626/RTx437	Hybrid	1986	1999	1992.5	3	
ATx378/RTx2783	Hybrid	1961	1982	1971.5	3	
ATx378/RTx2935	Hybrid	1961	2004	1982.5	3	
ATx2928/RTx435	Hybrid	1999	1984	1991.5	3	
ATx631/R07178	Hybrid	1985	2015	2000	3	
ATx626/RTx430	Hybrid	1986	1976	1981	3	
ATx626/RTx2536	Hybrid	1986	1964	1975	3	
ATx2752/RTx2783	Hybrid	1976	1982	1979	3	
ATx645/R08304	Hybrid	2001	2015	2008	3	
ATx645/RTx2536	Hybrid	2001	1964	1982.5	3	
ATx2752/R08304	Hybrid	1976	2015	1995.5	3	
ATx378/RTx437	Hybrid	1961	1999	1980	3	
ATx378/RTx430	Hybrid	1961	1976	1968.5	3	
ATx631/RTAM428	Hybrid	1985	1974	1979.5	3	
A05071/RTx2783	Hybrid	2005	1982	1993.5	3	
ATx623/R07178	Hybrid	1977	2015	1996	3	

Table 1 *Cont'd.*

Genotype	Type	Female Year	Male Year	Release Average	Environments	Reference
A07118/RTx435	Hybrid	2007	1984	1995.5	3	
ATx378/R08304	Hybrid	1961	2015	1988	3	
A05071/RTAM428	Hybrid	2005	1974	1989	3	
ATx2928/RTAM428	Hybrid	1999	1974	1986.5	3	
A07118/RTx2536	Hybrid	2007	1964	1985.5	3	
ATx3197/RTAM428	Hybrid	1961	1974	1967.5	3	
ATx2752/RTx2536	Hybrid	1976	1964	1970	3	
ATx2752/RTx437	Hybrid	1976	1999	1987.5	3	
ATx645/RTx430	Hybrid	2001	1976	1988.5	3	
ATx623/RTx435	Hybrid	1977	1984	1980.5	3	
A05071/RTx7000	Hybrid	2005	1957	1981	3	
ATx631/RTx7000	Hybrid	1985	1957	1971	3	
ATx3197/R07178	Hybrid	1961	2015	1988	3	
ATx2928/RTx7000	Hybrid	1999	1957	1978	3	
ATx645/RTx2783	Hybrid	2001	1982	1991.5	3	
ATx623/RTx436	Hybrid	1977	1992	1984.5	3	
ATx3197/RTx436	Hybrid	1961	1992	1976.5	3	
ATx631/RTx436	Hybrid	1985	1992	1988.5	3	
ATx623/RTx7000	Hybrid	1977	1957	1967	3	
R.Tx7000	Inbred		1957		3	(Peterson et al., 1984)
R.TAM428	Inbred		1974		3	(Peterson et al., 2009)
B.Tx2752	Inbred	1976			3	(Johnson et al., 1982)
R.Tx436	Inbred		1992		3	(Miller, 1992)
B.07118	Inbred	2007			3	
R.Tx435	Inbred		1984		3	(Miller, 1986)
B.Tx378	Inbred	1961			3	(Peterson et al., 2009)
R.Tx437	Inbred		1999		3	(Rooney et al., 2003)
R.08304	Inbred		2015		3	
B.Tx645	Inbred	2001			3	(Rosenow et al., 2002)
B.05071	Inbred	2005			3	
B.Tx3197	Inbred	1961			3	
R.07178	Inbred		2015		3	
B.Tx2928	Inbred	1999			3	
B.Tx623	Inbred	1977			3	(Peterson et al., 2009)
R.TX2536	Inbred		1964		3	(Duncan et al., 1991)
B.Tx631	Inbred	1985			3	(Miller, 1986)
R.Tx430	Inbred		1976		3	(Miller, 1984a)
B.Tx626	Inbred	1986			3	(Stephens and Karper, 1965)
R.Tx2783	Inbred		1982		3	(Peterson et al., 1984)

Table 2. DuPont Pioneer Plant Material

List of plant material tested in this experiment from the DuPont Pioneer sorghum breeding program. The information includes year of hybrid commercialization and whether the hybrids is currently commercially available in the United States.

Genotype	Release Year	Environments	Available
Pioneer 828	1966	3	
Pioneer 814	1968	3	
Pioneer 8416	1975	3	
Pioneer 8222	1982	3	
Pioneer 8313	1988	3	
Pioneer 8358	1989	3	
Pioneer 8282	1995	3	
Pioneer 8310	1993	3	
Pioneer 84G62	1998	3	X
Pioneer 85G50	2002	3	
Pioneer 83G15	2002	3	
Pioneer 83G19	2007	3	X
Pioneer 83P56	2014	3	X
Pioneer 83P73	2014	3	X

Table 3. Experimental Design

A summary of experimental design highlighting planting density, field replications, plot dimensions, harvest methodology, and irrigation procedures for each environment under observation

Material	Environment	Planting Density[a]	Reps[b]	Rows	Row Spacing[c]	Plot Length[c]	Irrigation
Texas A&M hybrids	College Station	197,680	3	2	0.762	6.7056	Limited
Texas A&M hybrids	Gregory	148,260	3	2	0.9652	9.144	Dry land
Texas A&M hybrids	Monte Alto	197,680	3	2	0.762	9.144	Full
Texas A&M hybrids	Perryton	148,260	3	2	0.762	9.144	Full
Texas A&M hybrids	Thrall	160,615	3	2	0.9652	9.144	Dry land
Pioneer hybrids	College Station	197,680	3	2	0.762	6.7056	Limited
Pioneer hybrids	Corpus Christi	148,260	4	1	0.9652	6.096	Dry land
Pioneer hybrids	Halfway	148,260	4	1	1.016	6.096	Limited
Texas A&M inbreds	College Station	135,905	3	2	0.762	6.7056	Limited
Texas A&M inbreds	Monte Alto	197,680	3	2	0.762	9.144	Full
Texas A&M inbreds	Perryton	148,260	3	2	0.762	9.144	Full

[a]seeds planted/hectare

[b]field replications

[c]meters

.

All trials were grown in a randomized complete block design, and all tests consisted of two-row plots with three replications, except for the Pioneer trials in Halfway and Corpus Christi where the experimental unit was a one-row plot with four replications. Agronomic practices varied but were standard for grain sorghum production in each area (Table 3).

2.2.3 Phenotyping

2.2.3.1 Yield and Test Weight

In most environments, grain yield was measured using a John Deere 3300 plot combine (Deere & Company, Moline, Illinois, USA) fitted with a Harvest Master GrainGage System (Juniper Systems, Logan, Utah, USA). Trials in Corpus Christi and Halfway were hand harvested and threshed using an ALMACO BT14 Belt Thresher (ALMACO, Nevada, Iowa, USA). Test weight ($kg\ m^{-3}$) was measured using a GrainGauge system (combine harvested plots) or a DICKEY-john mini GAC plus moisture tester (DICKEY-john, Minneapolis, Minnesota, USA).

2.2.3.2 Agronomic Traits

Leaf angle was measured in the field using a digital protractor on three plants per plot. On each plant, the top three leaf angles (flag leaf included) were measured and then averaged across all three plants to create a single leaf angle for each plot. Plant height (cm) on a plot basis was measured from the ground to the apex of the panicle. Days to anthesis was recorded in all environments (except Thrall) as the date when 50% of the plants in the plot were at mid-anthesis. The number of panicles per plot were counted

before harvest in College Station and Perryton and reported as the number of panicles per hectare.

2.2.3.3 Panicle and Seed Size

Grain samples were collected by hand harvesting five panicles per plot. The average panicle size (cm^2) per plot was determined by averaging panicle length and mid-panicle width of all five panicles. The panicles were then threshed using an ALMACO BT14 Belt Thresher, and 500 hundred seed weight was recorded by averaging the weight (g) of three packets of 500 seeds counted by a seed counter (The Old Mill Company, Savage, MD, USA). The number of grains per panicle was estimated by (average 500 seed weight * weight of 5 panicles) / 5 panicles.

2.2.3.4 Compositional Analysis

Compositional analysis of whole grain was performed using FOSS XDS MasterLab with the XDS Rapid Content module (FOSS North America, Eden Prairie, Minnesota, USA). The samples were scanned with wavelengths between 400 and 2500 nm, using ISIscan v.3.10.05933 software (FOSS North America). Near infrared reflectance (NIR) predictions for percent protein, starch, ash, fat, and fiber were based on NIR calibration curves developed by the Texas A&M AgriLife Research Sorghum Breeding Program. Biomass composition of these grain sorghum hybrids was estimated on dried, ground samples according to (Wolfrum et al., 2013) for percent protein, sucrose, lignin, glucan, and xylan.

2.2.4 Statistical Analysis

2.2.4.1 Basic Statistics and BLUPs

Basic statistics (mean, maximum, minimum, and standard error) for each trait were calculated using the tabulate platform in JMP Pro 13.1.0 (SAS Institute Inc., Cary, North Carolina, USA). BLUPs (best linear unbiased predictors) for each genotype and each trait measured were calculated in JMP using the Fit Model platform with the Standard Least Squares personality and restricted maximum likelihood (REML) method. An all random model was fitted as follows: response = genotype + environment + genotype*environment + replication within environment + error. Variance component estimates were also estimated using the Fit Model platform.

2.2.4.2 Genetic Gains and Heterosis

For each trait, the BLUPs for each hybrid were regressed against the year of hybrid release, and a linear model was fitted using R with the slope of the fitted line representing genetic gain over time. The coefficient of determination (r^2) denotes the proportion of the variance in hybrid release year that is predictable for each measured trait and was used to assess the effect of long-term selection on each trait. Relative heterosis was calculated using the genotypic BLUPs of each hybrid and their respective parental lines using the equation $\frac{(yield\ of\ hybrid) - (yield\ of\ midparent)}{yield\ of\ hybrid}\ x\ 100$ (Duvick, 2005).

2.2.4.3 Stability Analysis

Joint regression analysis was applied to determine the stability of yield BLUPs of hybrids with parental lines with common release years. Both regression coefficient (b_i) (Finlay and Wilkerson (1963)) and deviation parameter (δ_i^2) (Eberhart and Russell (1966)) were estimated using the plant breeding package in R (R Core Team, 2016; Rosyara, 2014.) When a genotype has $b_i = 1$, it has the "Type 2" or the "agronomic concept" of stability in which the genotype yield is consistent with the potential of the environment (Lin et al., 1986). The deviation parameter (δ_i^2) represents the "Type 3" concept of stability and measures "unpredictable irregularities in response to the environment provided by the deviation of the regression." (Lin et al., 1986). The closer δ_i^2 is to 0, the more stable the genotype is considered.

2.2.4.4 Genetic Contribution to Yield Gains

The relative contribution of genetics and plant breeding to improved yields was calculated according to summarized information in Russel (1991) and Duvick (2005). USDA-NASS data for annual on-farm yield estimates for the state of Texas were obtained from 1959 to 2016, the same time range of hybrid release in this genetic gain experiment. Both on-farm yield estimates and yield BLUPs from this genetic gain study were regressed by year, and the slopes were obtained. The relative genetic contribution was then obtained by the following equation: $\frac{slope\ of\ yield_{genetic\ gain\ experiment}}{slope\ of\ yield_{Texas\ on-farm\ estimate}}\ x\ 100$

Table 4. Basic Statistics

Maximum, minimum, mean, and standard error for grain yield, plant height, average leaf angle, number of panicles per hectare, days to flowering, panicle size, test weight, 500 seed weight, and grain number per panicle based on hybrids grown in three to five diverse environments in the state of Texas in 2016. Percent variance associated with environment, genotype (hybrid), genotype x environment (G x E) interaction, and replication nested within environment, and residual are described based on hybrids grown in three to five diverse environments in the state of Texas in 2016. Significance denoted if $P < 0.05$.

	Yield	Height	Leaf Angle	Number of Panicles	Days to Flowering	Panicle Size	Test Weight	500 Seed Weight	Grain Number
	t ha^{-1}	cm	^{o}	No. ha^{-1}	d	cm^{2}	kg m^{-3}	g	No. $panicle^{-1}$
Max	10.01	175	77	90,908	84	233	840	28.1	5401
Min	0.97	107	27	36,189	58	72	457	6.9	338
Mean	5.36	137	44	59,396	71	149	733	11.3	2197
Std. Err.	0.08	0.47	0.31	473	0.27	1.03	2.46	0.11	32.28
Environment	81.27	12.37	52.16	6.82	89.48	28.28	45.10	22.80	28.72
Genotype	2.21*	40.06*	8.14*	22.99*	2.48*	14.26*	5.30*	12.34*	19.68*
G x E	3.39*	14.48	0.92	26.33*	2.60*	5.75*	6.90*	1.43	10.57*
Rep(Env.)	0.04	2.01	9.58	0.00	0.12	7.45	0.18	10.14	2.36
Residual	13.09	28.08	29.20	43.87	5.33	44.27	42.52	53.29	38.66
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

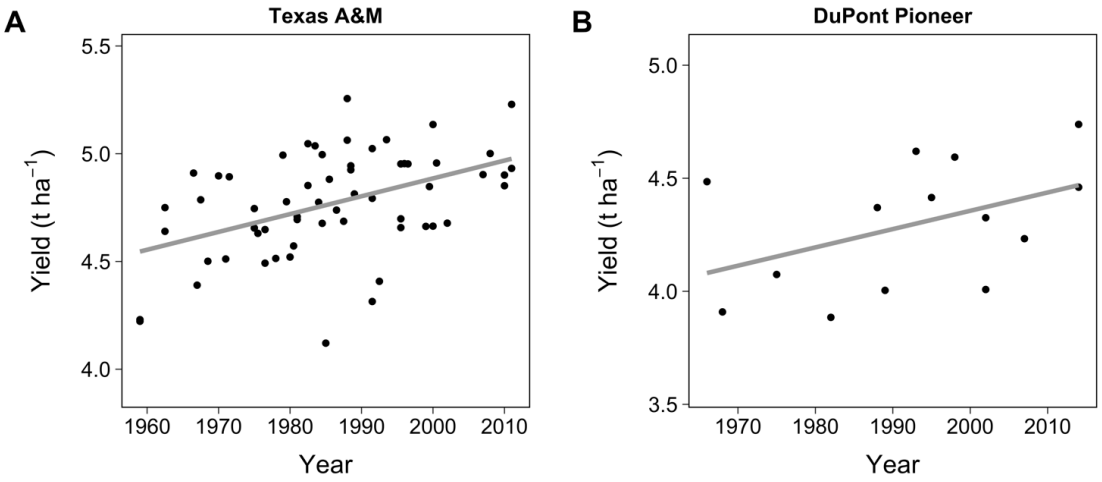

Figure 1. Genetic Yield Gain

A) Grain yield per Texas A&M hybrid regressed on the average parental line release year for each hybrid. Best Linear Unbiased Predictors (BLUPS) of Texas A&M hybrid grain yield based on hybrids grown in three to five diverse environments across the state of Texas in 2016. Grain yield is increasing + .008 t ha^{-1} annually and the $r^2 = 0.21$. B) Grain yield of 14 DuPont Pioneer sorghum hybrids regressed on year of commercial release. BLUPs of DuPont Pioneer material based on hybrids grown in three diverse environments across the state of Texas in 2016. Grain yield is increasing + .008 t ha^{-1} and the $r^2 = 0.20$.

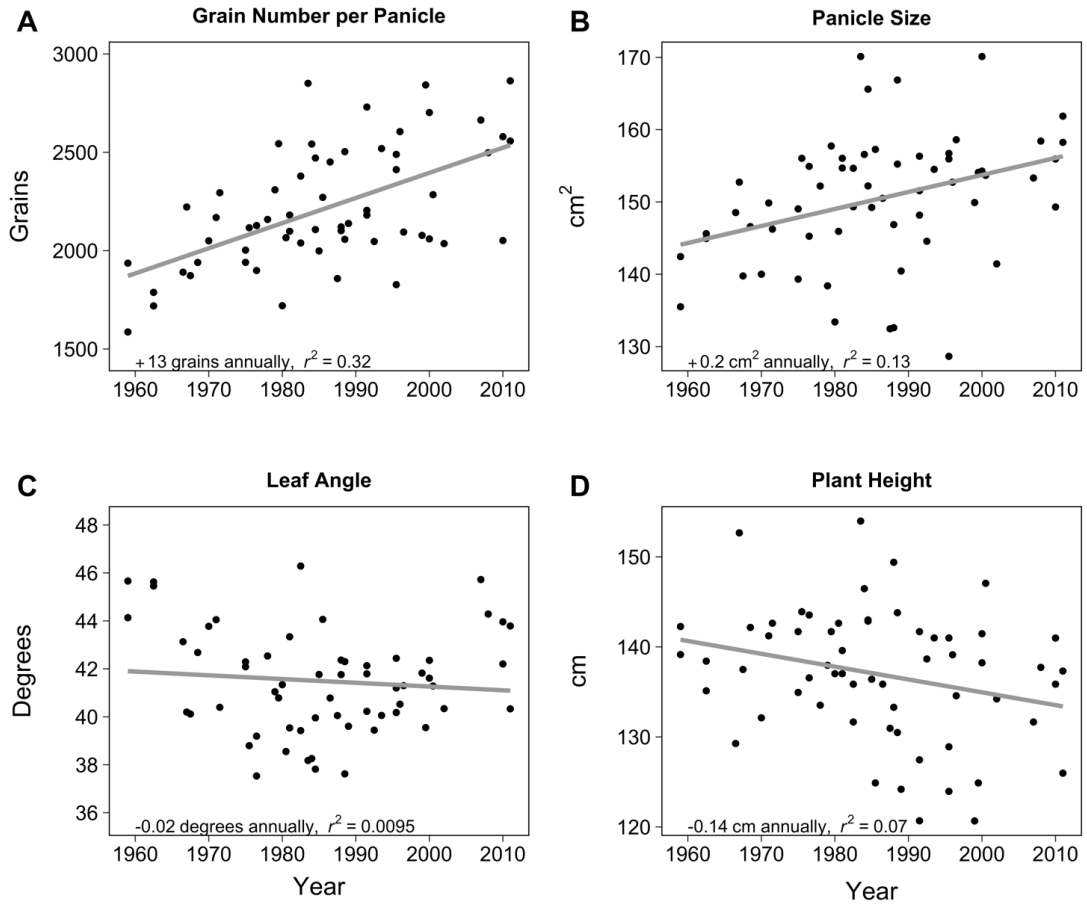

Figure 2. Genetic Gain of Physiological Traits

A) Grain number per panicle per hybrid, B) panicle size, measured in cm^2, per hybrid, C) average leaf angle per hybrid, and D) plant height, measured in cm, per hybrid regressed on the average parental line release year for each hybrid. Best Linear Unbiased Predictors (BLUPS) of the respective traits were based on hybrids grown in three to five diverse environments across the state of Texas in 2016.

2.3 Results

2.3.1. Genetic Yield Gains

In the Texas A&M sorghum germplasm, ANOVA revealed that genotype, environment, and genotype x environment were significant sources of variation for grain yield (Table 4). In the DuPont Pioneer trials, a very large environmental effect was observed for this trait (88% of variation) and genotype per se was not a significant source of variation. In both cases, the cause of yield differentials varied across environments, ranging from excess moisture in some locales to dry conditions in others. Overall, this range in yield reflects the variation often associated with sorghum production environments in Texas. Regression analysis indicates that yields improved by .008 t ha^{-1} annually within the Texas A&M sorghum germplasm (Fig. 1). The oldest hybrids, ATx3197/RTx7000 and ATx378/RTx7000 (1959 release), had yield BLUPs of 4.22 and 4.23 t ha^{-1}, respectively. The newest hybrids, A07118/R07178 (2011 release) and A07118/R08304 (2011 release), had yield BLUPs of 4.93 and 5.23 t ha^{-1}, respectively. The highest BLUP for grain yield was 5.26 t ha^{-1} for hybrid ATx378/R08304 (Table 1). This was a hybrid with an average parental release year of 1988; however, the hybrid consists of a current, unreleased pollinator line (R08304). The lowest BLUP for grain yield was a hybrid with an average parental release year of 1985, ATx626/RTx435, which yielded 4.12 t ha^{-1} (Table 1). Grain yields in Pioneer sorghum germplasm also increased by .008 t ha^{-1} yr^{-1} (Fig. 1). The lowest yielding hybrid in the study was Pioneer 8222, a hybrid released in 1982 with a yield of 3.88 t ha^{-1} while the highest yielding hybrid was Pioneer 83P73, released in 2014, at 4.74 t ha^{-1}.

Table 5. Summary of Genetic Gains

Genetic gains and coefficient of determination (r^2) for other traits measured in this experiment. Genetic progress was analyzed by calculating the slope of the linear model between BLUPs for the respective trait regressed on the average parental line release year for each hybrid. Best Linear Unbiased Predictors (BLUPS) of the respective traits were based on trials grown in three to five diverse environments in the state of Texas in 2016.

	Texas A&M		**DuPont Pioneer**	
Trait	Slope	**r^2**	**Slope**	**r^2**
Grain yield	+ .008 t ha^{-1} annually	.21	+ .008 t ha^{-1} annually	0.20
Grain number per panicle	+ 13 grains annually	.32		
Leaf angle	- .02 degrees annually	.0095	+ .08 degrees annually	.17
Plant height	- .14 cm annually	.07	- .24 cm annually	.02
Test weight	+ 0.28 kg m^3 annually	.16		
Panicle size	+ .24 cm^2 annually	.13	+ .51 cm^2 annually	.32
Number of panicles	- .015 panicles annually	.00079		
Days to flowering	+ .02 days annually	.08	+ .01 days annually	.11
500 seed weight	- .04 g annually	.37		
Biomass Protein%	- .00005 % annually	.002		
Biomass Sucrose%	- .043 % annually	.19		
Biomass Lignin%	+ .0056 %annually	.025		
Biomass Glucan%	+ .0087 % annually	.032		
Biomass Xylan%	+ .0031 % annually	.006		
Grain Protein%	- .0083 % annually	.08	- .0004 % annually	.0002
Grain Starch%	+ .0146 % annually	.13	+ .0176 % annually	.34

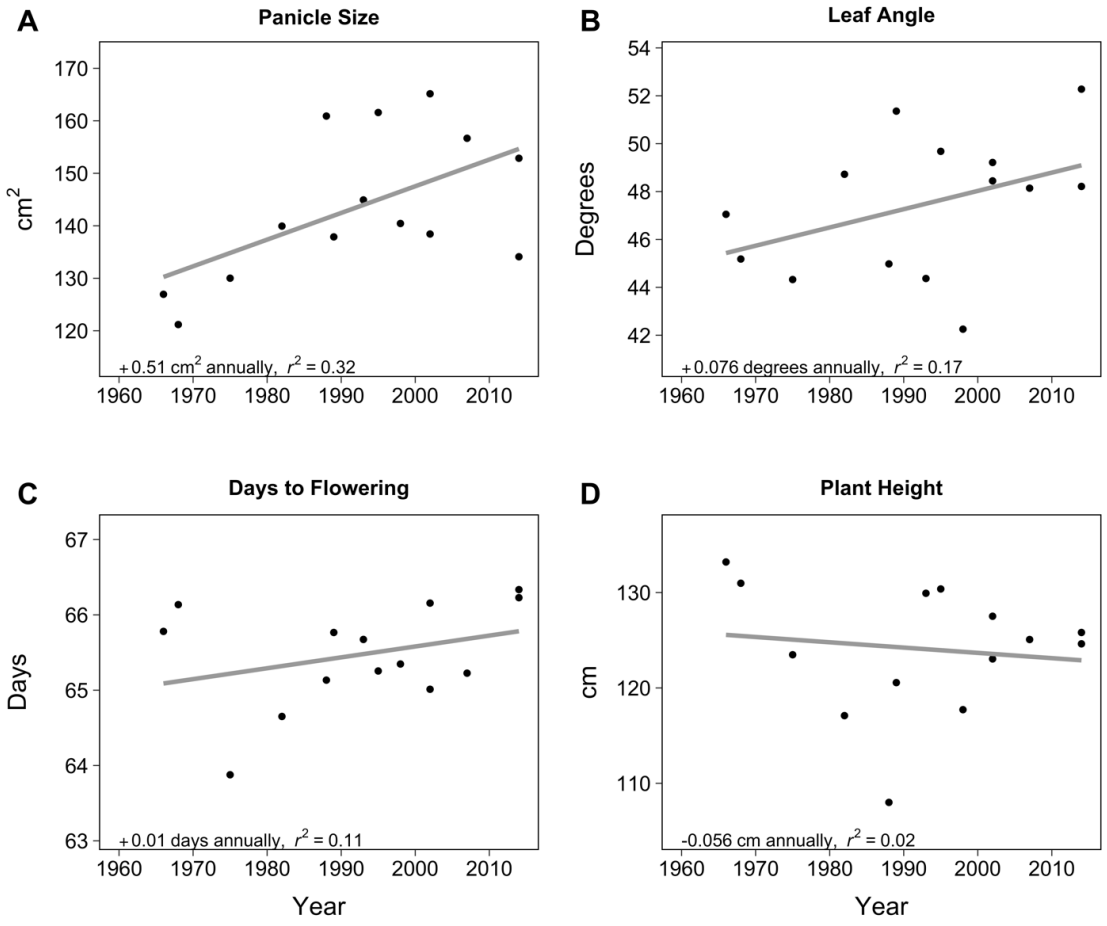

Figure 3. Genetic Gains of DuPont Pioneer Material

A) Panicle size, measured in cm^2, per hybrid, B) average leaf angle per hybrid, C) days until flowering per hybrid, and D) plant height, measured in cm, per hybrid regressed on the average parental line release year for each hybrid. Best Linear Unbiased Predictors (BLUPS) of the respective traits were based on DuPont Pioneer hybrids grown in three diverse environments across the state of Texas in 2016.

2.3.2 Genetic Changes in Other Traits

As expected, traits other than grain yield have changed under long-term selection. Grain number per panicle increased dramatically over the past 50-60 years (Fig. 2) going from a low of 1,586 on a 1959 hybrid (ATx3197/RTx7000) to a high of 2,863 on a 2011 hybrid (A07118/R07178). Through long-term selection, 13 grains have been added to a panicle on average each year in the Texas A&M breeding program ($r^2 = .32$) (Fig. 2).

Similar to grain number per panicle, panicle size has increased linearly by .24 cm^2 annually in the Texas A&M material (Fig. 2; Table 5). The oldest hybrids, ATx3197/RTx7000 and ATx378/RTx7000 (1959 release years), had panicle sizes of 135.51 and 142.44 cm^2, respectively. The newest hybrids, A07118/R07178 (2011 release) and A07118/R08304 (2011 release), had panicle sizes of 158.24 and 161.86 cm^2, respectively. The relative dimensions of the panicle have also increased by .5 cm^2 on average each year in the Pioneer material (Table 5).

Other traits have shown little to no change over time. For example, leaf angle increased (more upright leaf architecture) at a rate of 0.02 degrees annually. Given the low r^2, the trait is essentially unchanged in the Texas A&M breeding program (Fig. 2, Table 5). Leaf angles appear to have become more obtuse (flatter leaf architecture) over time in the DuPont Pioneer breeding program. The average leaf angle has decreased .076 degrees annually ($r^2 = .17$) (Table 5, Fig. 3).

According to linear regression, plant height has decreased by .14 cm annually in the Texas A&M breeding program, but given the low r^2 (0.07), it too is considered unchanged (Fig. 2). Likewise, in the DuPont Pioneer germplasm, the slope of the line

indicates plant height decreased by .24 cm annually, but the r^2 was very low (Table 5, Fig. 3). Given the long-established and positive relationship between increased plant height and yield (Rooney, 2004), there is a clear preference to maintain existing plant height within the agronomic production system.

The number of panicles per hectare is another trait that has had a limited change under long-term selection (Table 5). The number of panicles decreased by 19 panicles per hectare annually, but with a very low r^2 (0.007). Days to flowering is not changing dramatically over time in the Texas A&M breeding program (Table 5). The relationship between release year and flowering data only had an $r^2 = .08$, but the slope of the line indicated that flowering date had increased 1 day every 50 years, or .02 days on average each year. No change was observed in the DuPont Pioneer hybrids as well.

Conversely, seed weight decreased dramatically from long-term selection (Table 5). The weight of 500 seeds decreased by .04 grams on average each year in the Texas A&M hybrids. The r^2 for this trait was higher than all other traits measured at .37. Test weight has shown moderate increases between 1959 and today (Table 5). According to the slope of the line, test weight has improved by 0.28 kg m^{-3} annually ($r^2 = .16$) in the Texas A&M hybrids.

2.3.3 Genetic Gains of Biomass Composition

Percent sucrose in the biomass of grain sorghum has decreased over time due to long-term selection (Fig. 4). According to NIR estimates, sucrose decreased .048% annually in the Texas A&M sorghum breeding program with an r^2 of .19 (Table 5).

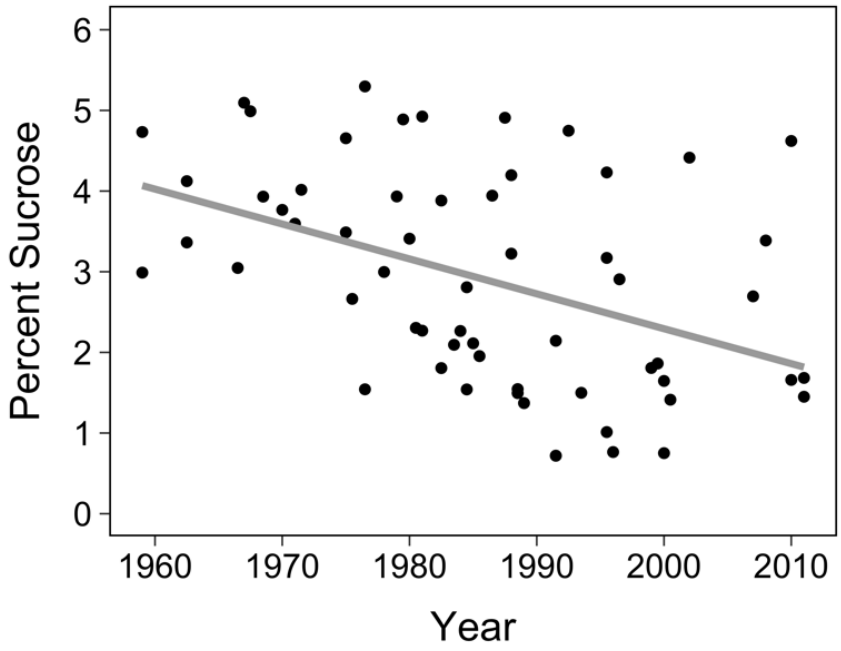

Figure 4. Genetic Gain of Percent Sucrose

Percent sucrose per hybrid regressed on the average parental line release year for each hybrid. Best Linear Unbiased Predictors (BLUPS) of percent sucrose were based on hybrids grown in 3 replications in College Station, Texas in 2016. Percent sucrose is decreasing – 0.043 % annually and the $r^2 = 0.19$.

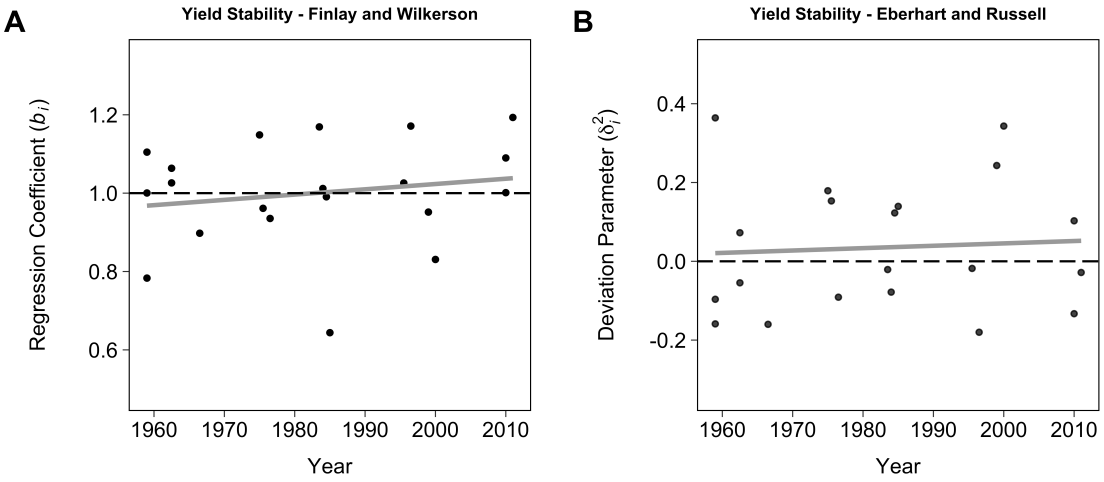

Figure 5. Yield Stability

(A) Finlay and Wilkerson's (1963) regression coefficient (b_i) and (B) the deviation parameter (δ_i^2) of Eberhart and Russel (1966) regressed on the average parental line release year for each hybrid. Results are based on twenty hybrids grown in five environments across Texas in 2016.

Other NIR estimated compositional traits had little to no response due to long-term selection (Table 5). Percent protein, a measurement that is important for forage quality, only had a r^2 of .002. Percent lignin, a trait that may influence root lodging, had a low r^2 of only .025 and increased by.58% annually. Percent glucan, a trait related to percent cellulose, only had an r^2 of .032 and increased by .87% annually. Percent xylan, a trait related to hemicellulose production, had a low r^2 of .0063 and increased by .31% annually (Table 5).

2.3.4 Genetic Gains of Grain Composition

Percent protein in the kernel trended lower over time from long-term selection. In the panel of Texas A&M old and new hybrids, protein decreased .08% every decade of breeding (Table 5). This represented a negative 1.03 percent change of percent protein each decade. However, the r^2 was small at only .08. Percent protein in the grain of DuPont Pioneer hybrids showed no change over time. Percent starch in the sorghum kernel has increased over time in the Texas A&M breeding program. The slope of the line indicated a 0.14% starch increase every decade of breeding or a 0.21% change in percent starch each decade. The r2 of the relationship was .13 (Table 5). The DuPont Pioneer program has increased starch content 0.18% per decade of selection. Other grain composition traits did not have any significant shifts in content.

2.3.5 Stability Analysis

With regard to grain yield, there was no discernable relationship between Finlay and Wilkerson's (1963) regression coefficient (b_i) and the year of release for a hybrid (Fig. 4). If $b_i = 1$ is considered the agronomic definition of stability, the newest hybrids

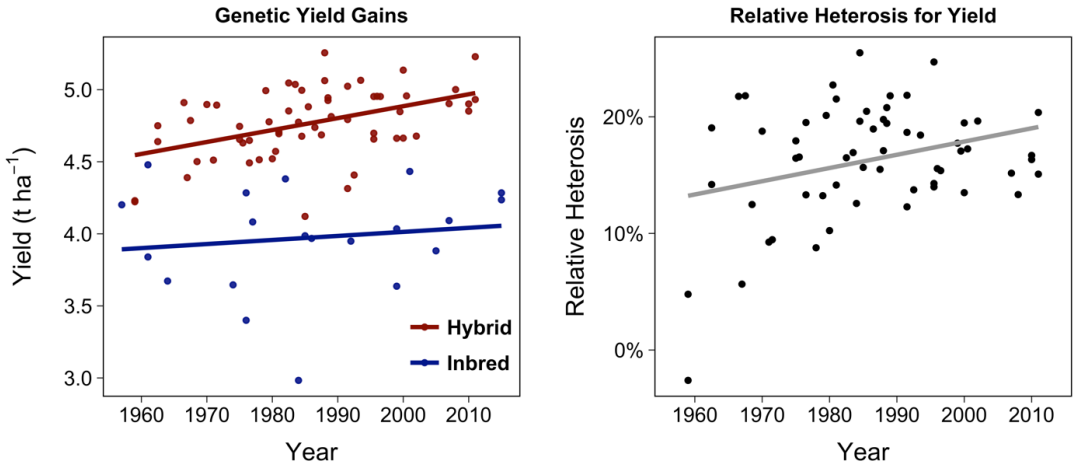

Figure 6. Genetic Gain of Heterosis

A) Genetic yield gains for hybrids and their respective inbred parents regressed against year of release. Hybrid yields have improved .008 t ha^{-1} annually. Inbred lines have improved .0028 t/ha annually. B) Relative heterosis for grain yield regressed on the hybrid average year of parental line release. Relative heterosis for grain yield has improved 0.11% annually. Best Linear Unbiased Predictors (BLUPS) of hybrid and inbred yields were based on trials in three to five environments across the state of Texas in 2016.

are not any more stable than older hybrids. The same results were observed when applying the deviation parameter (δ_i^2) of Eberhart and Russel (1966) (Fig. 5).

Ideally, the best hybrids should demonstrate high yield and high yield stability. In the current studies, there was not one hybrid that had b_i values within one standard deviation of yield stability (b_i = .86 – 1.14) and had yields greater than one standard deviation in BLUP mean yield (> 4.99 t ha^{-1}). Hybrids A07118/R08304 and ATx631/RTx2783, which have an average parental line release years of 2011 and 1983.5, respectively, had yield BLUPs greater than one standard deviation from the BLUP mean. However, those two hybrids had a stability greater than one standard deviation from the stability mean.

2.3.6 Heterosis

Genetic yield gain is increasing at a faster rate in sorghum hybrids than the inbred parents. In the Texas A&M sorghum breeding program, hybrid yield has increased at .008 t ha^{-1} annually. Yield in the parental lines that make up those hybrids increased at the rate of .003 t ha^{-1} annually. This disparity has caused relative heterosis for yield to increase over time where relative heterosis increases averaged 0.11% per year (Fig. 6).

2.3.7 Contribution of Breeding to Yield Gains

The slope of the line for genetic yield gain within the Texas A&M AgriLife Research Sorghum Breeding Program was .008 t ha^{-1} (Fig. 7). This slope represents the change in yield due to plant breeding or genetics. Over the same time (1959 – 2016), the

.

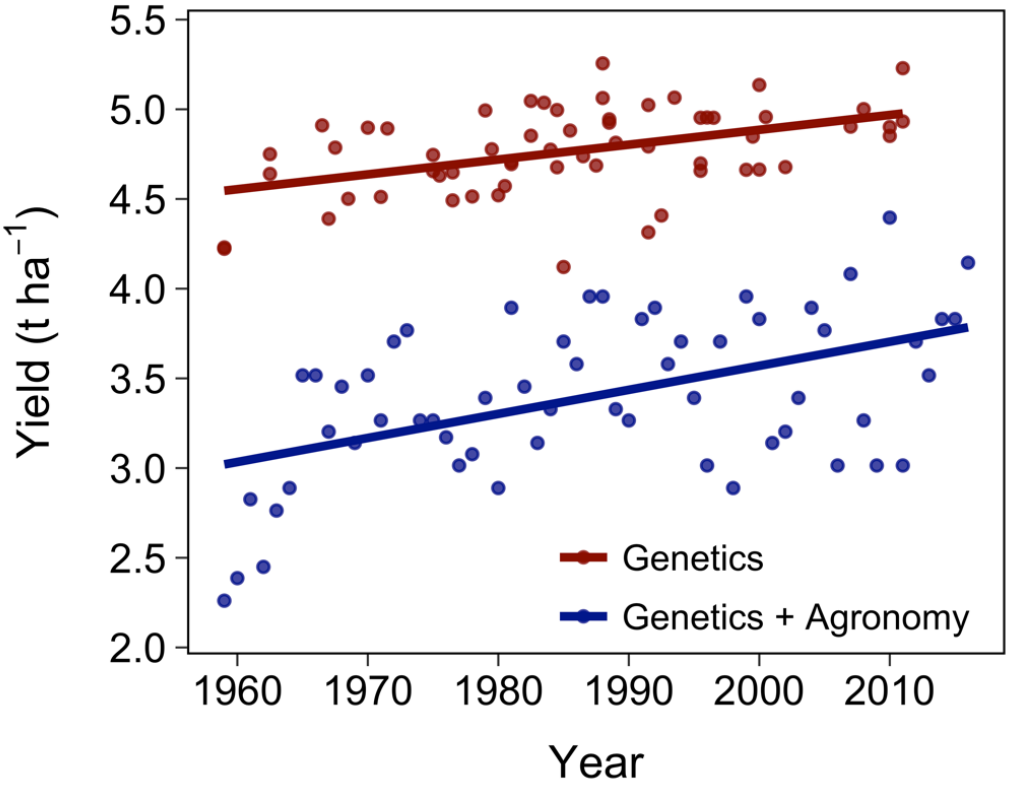

Figure 7. Contribution of Genetics to Yield Gains

Experimental results of sorghum grain yield per hybrid regressed on year of release (red) compared to USDA-NASS estimated on-farm sorghum grain yield annual estimate in Texas over the same time (blue). The experimental results represent changes in genetics over time. The USDA results represent changes in genetics, farming practices, and climate.

slope of the USDA-estimated Texas on-farm yield was .013 t ha^{-1} (Fig. 7). This value effectively represents the changes in yield due to both improved genetics and agronomy practices. Therefore, the relative contribution of plant breeding to total yield gains is 62%.

2.4 Discussion

2.4.1 Genetic Gains in Grain Yield of Hybrids

In both sets of germplasm, genetic yield gain was the same at .008 t ha^{-1} annually. Gizzi and Gambin (2016), published a study on the genetic yield gains of Advanta Semillas grain sorghum program in Argentina and reported a similar rate of genetic gain, increasing by .0087 t ha^{-1} annually. In all three of these estimates of genetic yield gain, different numbers of hybrids were tested, testing environments were different, and the germplasm used to create the hybrids were from different programs.

Given the number of differences in the trials, the consistency in the rate of gain reported implies that sorghum breeding programs are progressing at similar albeit slow rates. Exactly why there were few differences between the rates of gain could trace to the limited genetic diversity within sorghum breeding programs. In fact, much of the original germplasm in existing hybrids has its origins in the Texas A&M AgriLife Research sorghum program and the Sorghum Conversion program, which was a collaborative research effort between Texas A&M AgriLife Research and USDA-ARS.

Genetic yield gain in sorghum is significantly slower than other major field crops. The observation is especially small compared to DuPont Pioneer's corn breeding program (.0876 t ha^{-1} yr^{-1}) (Smith et al., 2014). Despite yielding less than sorghum

overall, cotton (.0216 t ha-1) and soybeans (~.022 t ha-1) both demonstrate faster rates of genetic yield gain (Campbell et al., 2014; Specht et al., 2014).

2.4.2 Improvement in Heterosis for Grain Yield

Relative heterosis is improving from long-term selection. This is in contrast to maize, in which inbred genetic yield gains have outpaced hybrid genetic yield gains, translating into no improvement in relative heterosis (Duvick, 2005). The different results could be related to the method of pollination. Corn is a naturally cross-pollinated species. At the beginning of hybrid breeding, the corn hybrids were already "elite" while inbred line performance was poor. Therefore, the inbreds have improved at a faster rate than the hybrids. The opposite is true in sorghum, which is a naturally self-pollinating species. At the beginning of the hybrid sorghum era, the inbred lines were already "elite" because cultivars were previously used in production. The hybrids have made many improvements since the late 1950s, leading to an increasing rate of heterosis in sorghum.

2.4.3 No Change in Yield Stability

Yield stability has not shown any discernible trend with year of hybrid release. High yield stability has been an important trait for breeders since the beginning of sorghum breeding and remains important today. It is possible that this study has overemphasized favorable environments. In the five environments in which the Texas A&M material was observed, three were irrigated and one of the two rainfed environments (Gregory) actually had excess soil moisture during the 2016 growing season. Admittedly, five environments is a limited number of environments to assess stability so inferences presented herein are subject to further evaluation.

2.4.4 Large Contribution of Genetics to Total Yield Gains

The estimate of nearly 62% as the relative contribution of genetics and plant breeding to total yield gains was similar to a recent estimate by Assefa and Staggenborg (2011). The estimate derived from this study was higher than other estimates in sorghum, but those were the focus of dryland production (Unger and Baumhardt, 1999) or described early eras of hybrid production (Miller and Kebede, 1984).

2.4.5 Other Physiological Changes

2.4.5.1 Traits Demonstrating Improvement from Indirect Selection

In addition to grain yield, many other physiological traits have changed from 55 years of selecting high yielding hybrids. While not under direct selection, the size of the panicles has become larger in both sets of hybrids. Not only have panicles become larger, but the panicles have more seed. Improvements in both grain number per panicle and panicle size relate to an increasing "yield potential per plant." Unlike in maize, where most of the genetic yield gains have come from higher planting populations (Duvick et al., 2004), sorghum's major yield increases appear to be of improvement on a per-plant basis. Although these traits are under intentional selection *per se*, an improvement in total yield potential on a per plant basis could be considered improvement from indirect selection by breeders.

Test weight also showed substantial improvement from long-term selection. The importance of test weight in selection varies between sorghum improvement program; the strong linear improvement hints that either through direct or indirect selection it is important in the selection of high-yielding hybrids.

Without a doubt, the sucrose content (%) in the biomass of grain sorghum hybrids has never been a selection criterion. The noteworthy decline in this trait from long-term selection may indicate that breeders have indirectly selected for more efficient utilization of the available assimilate. Sucrose is comprised of both glucose and fructose and is a major sugar transporter in plants (Lemoine, 2000). Sucrose is transported from site of synthesis to sync, i.e., starch production in grain. Less percent sucrose in the biomass at maturity of newer hybrids hints at the possibility of greater utilization of sugars for higher yielding hybrids.

Grain composition is another trait of variable importance within sorghum breeding programs, and the small changes observed appear to confirm this observation. Environmental conditions often influence grain composition even more than genetics (Beta and Corke, 2001) making systematic selection difficult in a crop commonly grown in stressful environments. Because the production of starch requires less energy than the production of protein (Duvick, 2005), breeders may have indirectly selected for lower percent protein and higher percent starch by selecting for high yielding hybrids. Duvick reported the same general trend has occurred in maize. In that study, protein decreased by 0.3% decade^{-1} and starch increased 0.3% decade^{-1} (Duvick, 1997). Like grain yield, the grain composition of sorghum is changing at a slower pace than in maize. Given the relationship between starch content and grain yield, the difference in the grain composition rate of change might be explained by the different rate of yield gains between the two crops.

2.4.5.2 Traits Demonstrating No Change from Intentionally Holding Traits Constant

In both the Texas A&M and DuPont Pioneer hybrid trials, days to flowering increased very slightly from long-term selection indicating that this is a trait that breeders have kept constant over many years of breeding. Plant height dropped slightly in both the Texas A&M and DuPont Pioneer studies. While the slope of the lines is indeed negative, plant height has been held under specific thresholds due to grower preferences. The market demand to maintain existing height in grain sorghum is likely a major reason why genetic yield gains are limited. In fact, it is intriguing that minor yield increases have been accompanied by minor height decreases considering the strong relationship between plant height and grain yield (Rooney, 2004).

2.4.5.3 Traits Demonstrating No Change

In the Texas A&M AgriLife Research hybrids, there was no relationship between leaf angle and year of hybrid release. In the DuPont Pioneer hybrid trial, leaf angles became less acute from long-term selection. These results are surprising due to the strong indirect selection for more acute leaf angles in maize breeding (Russel, 1991; Duvick et al., 2004). Because sorghum is much shorter than maize and there is less competition for light interception; in taller sorghums, leaf angle has been associated with greater biomass yields (Truong et al., 2016). Also, sorghum will tiller if there are sufficient light and moisture are available which is a completely different system than in corn. Overall, the lack of change in leaf angles suggests that improved light interception occurs by flattening the leaf architecture. This observation indicates a management

concern; if more acute leaf angles are not strongly associated with higher yields, it implies that the crop, as planted, is not capturing light as effectively as possible. If so, then it indicates that the row spacings should be narrowed to improve both competition between plants and light capture. Studies on row spacing over the years have consistency confirmed increased yields with narrower row spacing (Myers and Foale, 1981).

Related to light interception, the number of panicles per area showed no change regarding long-term selection. As seeding rates and populations were constant, this indicates that tillering rates have not changed over time.

2.4.6 Possible Reasons for Genetic Yield Gains

There are several explanations for why sorghum has a lower rate of genetic yield gain than other major field crops. First, high yielding hybrids in sorghum are valued only if they possess genetic tolerances to abiotic and biotic stresses. The relative importance of these breeding targets is because sorghum is usually planted on the most marginal lands with fewer inputs. Thus, the demands of multiple trait breeding detracts from and may place a genetic "drag" on yield increases.

Due to the logistics of commercial seed production, sorghum does not have highly optimized heterotic groups (Menz et al., 2004). Heterotic patterns are essentially divided on the criterion of whether inbred lines maintain or restore (Rooney, 2004). Since previous efforts have produced inconclusive results regarding the identification of heterotic groups, breeders have kept B- and R- inbred line development programs separated for practical reasons. In diversity studies, sorghum lines cluster based on working group and not cytoplasmic male sterility factors (Menz et al., 1996). Genetic

yield gains might have improved more quickly if heterotic groups in sorghum were based solely on genetic diversity (Menz et al., 2004; Gabriel, 2006). Therefore, heterotic groups are not as optimized as they are in corn where no biological restrictions exist on the use of a line as a seed or pollinator parent.

Sorghum inbred line development requires an extra step because of it has a perfect flower. Seed parents must be male-sterilized for hybrid seed production and in many cases, testcrossing. New B-lines are sterilized using a backcrossing procedure where the B-line is crossed to a standard A-line that serves as the source of male sterile cytoplasm. The sterilization process takes a multiple backcrosses, and this burden limits the number of new seed parents in hybrids (Rooney, 2004). A previous study found a lower diversity among B-lines because the development of new A/B lines is less efficient compared to R-line development (Menz et al., 1996). These factors have reduced the rate of development of new female germplasm and thus slowed genetic progress.

The number of United States universities working on sorghum research has plummeted over the last 40 years. There are currently only three U.S. universities with consistent and long-term grain sorghum breeding programs (Purdue University, Kansas State University, and Texas A&M University). Other start-up programs have formed recently (Iowa State University, Clemson University, USDA-ARS in Lubbock, TX, University of Florida, and the University of Illinois), but these programs are new and time is required to establish a productive program. Much of the public sorghum research has shifted from investigating grain sorghum to investigating sorghum as a cellulosic

ethanol feedstock. Also, more research than ever focuses on optimizing new genomics and phenotyping platforms which are accessory to improving germplasm.

The investment in research and development has also reduced in the private sector. As an example, the number of companies entering hybrids in the Texas A&M AgriLife Extension Variety Testing performance trials in Weslaco (irrigated) has dropped from 30 in 1985 to five in 2016. Similar trends exist in all locations where the tests are produced. While not all companies participate in these variety testing trials, the trends indicate reduced effort in breeding and development. Combine that with the reality that many of the companies that enter these trials simply retail hybrids and do not have active breeding programs, it means that the number of companies breeding sorghum is even less than the number of companies selling sorghum and participating in these tests. All of this means a reduced research and development infrastructure. Over the timespan included in this study, the target environments for sorghum production have shifted dramatically. Initially, a substantial proportion of sorghum production occurred under irrigation. Today, irrigated sorghum represents a small proportion of total production due to increasing water restrictions. Sorghum has been displaced from productive dryland hectares as corn and soybean production expanded from the traditional Corn Belt. Thus, the average productivity potential of a sorghum hectare has dropped. Consequently, the yield gains look more impressive given the reduction in yield potential of the target environment.

Finally, it appears that sorghum growers have placed phenotypic constraints on hybrid sorghums that may restrict genetic gain. Plant height and maturity have a

substantial relationship with grain yield, with later maturity and tall height associated with improved grain yield (Casady, 1965; Graham and Lessman, 1966; Dalton, 1969). Thus, the absence of changes in plant height imply that either grower preferences or marketing demands a sorghum hybrid of a specified height. Changing growers' perspectives, and thus plant breeders' perspectives, regarding acceptable plant height represents “low-hanging fruit” for potential sources of yield increases. The lack of adoption of taller hybrids, or even hybrids becoming shorter over time as this study indicated, has contributed to sorghum’s slow rate of genetic yield gain.

2.4.7 Promises for Future Gains from Sorghum Breeding

Although sorghum genetic yield gains have been modest at best, there are emerging technologies that could significantly impact yield gains in a positive manner. Doubled haploids have revolutionized the corn breeding pipeline, delivering genetically pure inbred lines in only two generations. Several research projects are actively pursuing this technology in sorghum. If successful, this new technology could greatly increase the efficiency of sorghum breeding programs. High-Throughput Phenotyping is already making an impact by collecting agronomic information like plant height faster than manual measurement. This technology shows promise in the future by seeing more than what the breeder’s eye can see via advanced sensors and methods. Genomic Selection is already having a major impact in the private industry in major field crops. As this technology advances, implementing genomic selection at key points in sorghum breeding programs could have a major impact on gains from selection.

2.5 Conclusions

In many crops, era studies compare hybrids from different generations to assess genetic progress. Findings from this study indicate the rate of genetic yield gain in sorghum is slower than in other major field crops. Relative heterosis continues to rise in sorghum, but improvements in yield stability did not change with long-term selection. Overall yield gains reflect both improved genetics and management practices. Some physiological traits have changed due to a presumed association with increased yield while others have not. Despite a plurality of reasons for why the rate of genetic yield gains is slower in sorghum, emerging technologies in plant breeding should lead to future increases.

3. GENE EXPRESSION AND HERITABILITY ANALYSIS OF THE BLACK PERICARP TRAIT

3.1 Introduction

3.1.1 Overview

Sorghum [*Sorghum bicolor* (L.) Moench] typically has a red, yellow, or white pericarp color (Rooney, 2000). However, there are also rare sources of germplasm that contain a black pericarp within breeding programs (Gous, 1989; Dykes et al., 2009). The source of black sorghum used in the Texas A&M AgriLife Research Sorghum Breeding Program was an accession from Sudan (Rooney et al., 2013b). Since the introduction of this black Sudanese accession, the Texas A&M AgriLife Research Sorghum Breeding Program has developed both pollinator and seed parents with a black pericarp. These inbred lines are used together to create a temperately-adapted black grain sorghum hybrid known as Onyx that is grown commercially for utilization in the health food market (Rooney et al., 2013a; b).

Sorghum with a black pericarp has high-levels of antioxidants and phenolic compounds (Gous, 1989; Dykes et al., 2013). The main compounds of interest are 3-deoxyanthocyanidins (3-DOAs). The 3-DOAs in black sorghum are a unique type of anthocyanin (Clifford, 2000; Awika et al., 2004). Unlike the anthocyanin pigments that are found in many other plants, sorghum's 3-DOAs lack a hydroxyl group at the 3' carbon position increasing the stability of molecules (Sweeny and Iacobucci, 1981; Mazza and Brouillard, 1987). Black sorghum has a much greater 3-DOA content than red, yellow and white pericarp types (Awika, 2000; Awika et al., 2004). The 3-DOAs

found in black sorghum contribute to black sorghum's great antioxidant values, which are higher than blueberries on a per weight basis (Awika et al., 2003, 2005; Wu and Prior, 2005; Dykes and Rooney, 2006). Research has demonstrated that eating foods rich in antioxidants can be beneficial to decreasing risk for some of our most widespread chronic diseases, such as heart disease (Kushi et al., 1999; Awika and Rooney, 2004), high blood pressure (Tsuda et al., 2003), and even cancer (Chen et al., 1993; Shih et al., 2007). Since these highly concentrated phenolic compounds are found in sorghum, a shelf-stable commodity with easy processing (Kushi et al., 1999; Anderson, 2003), there is increasing interest from the food industry to utilize black sorghum as a nutraceutical additive, natural colorant, and natural food preservative (Awika et al., 2005; Dykes and Rooney, 2006; Dykes et al., 2013).

3.1.2 Genes Involved with Pericarp Color

Red, yellow, and white pericarp colors are controlled by relatively simple genetic factors. The *R* and *Y* genes interact epistatically to produce red *(R_Y_*), white (*R_yy or rryy*), or yellow (*rrY_*) colors (Graham, 1916; Vinall and Cron, 1921; Rooney, 2000). Other known genes influence the color of sorghum seeds; for example, the intensifier (*I*) gene promotes a more intense pericarp color (Ayyangar et al., 1933; Rooney, 2000). The mesocarp thickness (*Z*) gene also influences shades of pericarp color as more anthocyanins may accumulate in a thicker, larger mesocarp layer (Ayyangar et al., 1934; Rooney, 2000). The B_1 and B_2 genes interact epistatically to control the presence and absence of condensed tannins in a testa layer (Laubscher, 1945; Stephens, 1946; Rooney,

2000) while the spreader (*S*) genes allows the condensed tannins to "spread" from the testa layer to the mesocarp and epicarp layers (Vinall and Cron, 1921; Rooney, 2000).

Despite these known genetic factors, black pericarp sorghum's genetic inheritance cannot be completely explained based on any combination of these genes or traits. Using the known two-gene pericarp color model of *R* and *Y* genes, black sorghum was determined to be genetically red (*R_Y_*) from test crossing to other pericarp colors (Rooney et al., 2013b) and thus, it has been proposed that additional genes must exist to control this phenotype (Rooney et al., 2013b). While the trait is heritable (Rooney et al., 2013b; a), the frequency of the phenotype typically observed in an F_2 population (approximately one black pericarp plant out of 1,000 F_2's) suggests that multiple recessive genes influence the trait's expression.

3.1.3 Previous Research on Black Sorghum Trait

Pfeiffer and Rooney (2016) performed a generation means analysis to understand further the complex genetics involved with the trait. The experiment showed that multiple genes with additive, dominance, and epistatic interactions were controlling the black pericarp color and associated prominent levels of beneficial health compounds. Using several different methods to estimate the number of genes controlling the black pericarp trait, Pfeiffer and Rooney (2016) concluded that the number of genes ranged from two to twelve. The number varied based on which estimation equation and which measurements were used. The observed estimates of two-to-four genes controlling 3-DOA concentration were marginally lower than expected considering observing 1:1000 black F_2 phenotypes suggests five segregating genes (Pfeiffer and Rooney, 2016). In the

same experiment, the trait appeared to be moderate to highly heritable with broad-sense heritability estimates ranging from .42 to 1, depending on which measurement of color or phenol composition was used (Pfeiffer and Rooney, 2016).

In addition to complicated genetic inheritance, the trait is also influenced by environmental factors. The black pericarp and associated healthy compounds are not expressed under all environmental conditions (Dykes et al., 2009). Panicles of Tx3362 that are partially shaded have a red pericarp phenotype, whereas unshaded panicles are black (Dykes et al., 2009; Rooney et al., 2013b; Pfeiffer and Rooney, 2015). Therefore, sunlight exposure is essential for the development of the black color (Dykes et al., 2009; Pfeiffer and Rooney, 2015). Plants grown in off-season nursery in Puerto Rico do not exhibit pericarp colors as dark as observed under long-day, summer environments in Texas (unpublished data). At present, it is unclear as to the feature of temperate-zone sunlight (e.g., light quality, photoperiod duration) that regulates this difference in expression of the black pericarp phenotype between temperate and tropical environments.

Pfeiffer and Rooney (2015), conducted a series of experiments to explore the effects of the timing of light exposure on the production of 3-DOAs and other compositional factors in addition to the black color itself. Pollination bags were placed on panicles to shade the panicle from sunlight for different time intervals on black pericarp line Tx3362. The authors concluded that the longer the panicle was shaded, the lighter the color of the panicle and the lower the concentration of beneficial phenolic

compounds. In some cases, long-term shading eliminated the black color resulting in red grain (Pfeiffer and Rooney, 2015).

3.1.4 Biosynthesis of Phenolic Compounds

Phenolics, which include the 3-DOAs, are synthesized in plants from the phenylpropanoid pathway that synthesize a wide range of essential metabolites with important implications for food quality and human health (Lo and Nicholson, 1998). The phenylpropanoid pathway follows a series of metabolic branch points that can result in the biosynthesis of lignins or flavonoids, with the later divided into various classes including anthocyanins, proanthocyanidins, flavones, flavonols, flavanones, flavan-4-ols and isoflavonoids (Petti et al., 2014). Each group of flavonoids serves a specific function in plants that is dependent on particular developments or environmental cues. The synthesis of all flavonoids share a series of early biosynthetic steps catalyzed by Phenylalanine ammonia lyase (PAL), trans-cinnamate 4-monooxygenase (C4H), 4-coumarate:CoA ligase (4CL), and chalcone synthase (CHS) (Mizuno et al., 2012; Liu et al., 2013b). The chalcones are then isomerized into naringenin by chalcone isomerase (CHI), which is a common precursor to all flavonoids (Stich and Forkmann, 1988; Boddu et al., 2006; Mizuno et al., 2012; Ibraheem et al., 2015).

From the flavonoid precursor naringenin, specific branch biosynthetic enzymes are involved in the production of the various classes of flavonoids, and the regulation of these enzymes/genes (plus early biosynthetic steps) largely determines the flavonoid composition in specific tissues and cell types (Gonzalez et al., 2009; Heppel et al., 2013; Jun et al., 2015). In the 3-DOA synthesis branch pathway, the flavanones naringenin and

eriodictyol are converted to the flavon-4-ols apiforol and luteoforol via the enzyme Flavanone 4-reductase (FNR) with the production of luteroforol also requiring the action of a flavonoid 3'-hydroxylase (Mizuno et al., 2012). The final step, converting the flavon-4-ols to 3-DOAs is controlled by an unidentified enzyme that has been proposed to involve an anthocyanidin synthase (Mizuno et al., 2012). The same research group later reported that the proposed final enzyme was in fact acting as the FNR enzyme (Mizuno et al., 2016). In the flavones biosynthesis branch pathway, flavone synthase II (FNSII) converts naringenin and eriodictyol into the flavones, apigenin and luteolin, which can be condensed to produce the pholobaphenes in sorghum and maize (Mizuno et al., 2012, 2016; Kawahigashi et al., 2016). In both the 3-DOA and flavones pathway, flavonoid 3'-hydroxylase (F3'H) introduces a hydroxyl group at the 3' position of ring B of naringenin, eventually leading to the synthesis of luteolinidin or luteolin (Shih et al., 2006; Mizuno et al., 2012, 2016; Sharma et al., 2012; Kawahigashi et al., 2016).

In the anthocyanidin synthesis pathway, a flavonoid 3-hydroxylase (F3H) converts naringenin to dihydrokaempferol while a flavonoid 3'-hydroxylase (F3'H) converts dihydrokaempferol into dihydroquercetin (Mizuno et al., 2012; Liu et al., 2013b; Tohge et al., 2017). These dihydroflavanols are converted to the flavan-3,4-diols (leucopelargonidin, leucocyanidin) via NADPH- dependent reduction of the C-4 carbonyl group by the action of a specific dihydroflavonol 4-reductase (DFR). The anthocyanidins pelargonidin and cyanidin are synthesized via the enzyme anthocyanidin synthase (ANS) by removing a hydroxyl group, and finally the unstable anthocyanidins

are then converted to the stable anthocyanin by a flavonol 3-O glucosyltransferase (3GT) (Liu et al., 2010; Mizuno et al., 2012; Poloni and Schirawski, 2014).

The regulation of flavonoid biosynthesis represents an intricate regulatory grid rather than a simplistic linear enzymatic pathway, and this regulation is both species dependent and varies depending on the tissue and developmental stage of the plant (for review see Tohge et al., 2017).The flavonoid pathway is under tight developmental regulation with multiple environmental cues including light, hormones and biotic stressed have been shown to control specific enzyme within the complex biosynthetic pathway (for review, see Liu et al., 2013). Each branch of the pathway leading to different classes of flavonoids is under separate control with the result being a complex array of colored and uncolored flavonoids accumulation. The regulation of the pathway is largely at the transcriptional level, and the regulation resides at the level of biosynthetic genes and regulation of the transcription factors regulators themselves. As a review of the complex species-specific regulation of flavonoid biosynthesis is beyond the scope of this dissertation, the focus will be on the current state of literature related to this pathway in sorghum.

3.1.5 Light-Dependent Gene Expression

Some genes in the biosynthesis pathway demonstrate light-induced expression. In mesocotyl tissue, F3H, DFR, and ANS genes were over expressed in the presents of light (Liu et al., 2010). Another study has shown that early in the pathway, PAL and CHS are induced by light as well as fungal infection (Lo and Nicholson, 1998). Whether

light induces expression of genes appears to depend on where the location of the tissue within the plant.

3.1.6 Phenolics in Tan-Colored Plants

While plants with a pigmented secondary plant color accumulated 3-DOA in their leaves, plants with a tan secondary plant color are not able to produce these important compounds in their leaf tissue (Mizuno et al., 2016). Alternatively, tan-colored sorghum's produce higher levels of flavones than their pigmented counterparts (Dykes et al., 2009, 2011). Sorghum with a tan plant color do not have a functional FNR gene, also known as the *P* gene, thus produce lower levels of flavon-4-ols, and unable to produce apigenidin and luteolinidin (Kawahigashi et al., 2016). Instead, FNSII appears to be activated in tan plants to produce greater levels of flavones (Mizuno et al., 2016). The F3H enzyme was not expressed in the tan injury response in leaves, suggesting that the anthocyanidin pathway was not activated (Mizuno et al., 2016).

3.1.7 Production of 3-DOA from Fungal Attack

In sorghum, the most common anthocyanidin types are the 3-deoxyanthocyanidins, primarily luteolinidin (LUT) and apigeninidin (AP), and their glycosylated derivatives, known as 3-deoxyanthocyanins (Dykes et al., 2009, 2013) The 3-DOA compounds are often produced in sorghum leaves in response to biotic stresses and can be considered phytoalexins (Lo et al., 1999). When sorghum leaf cells are under fungal attack, 3-DOAs are produced as a defense mechanism (Snyder et al., 1990). Previous studies have reported PAL, C4H, 4CL, CHS, CHI, F3H, DFR, FNSII,

and ANS are upregulated genes in infected tissue compared to non-infected tissues (Liu et al., 2010; Mizuno et al., 2012, 2016).

3.1.8 Yellow Pericarp Gene

In synthesis of 3-DOAs in sorghum requires the presence of the *Y1* yellow pericarp gene that functions as a MYB transcription factor. Previous research has shown *Y1* to be an essential gene for 3-DOA production in sorghum grain as it is required for the transcription of FNR/DFR (Boddu et al., 2005). However, plants without the *Y1* gene still express 3-DOAs in the leaf (Mizuno et al., 2012). Thus, it has been proposed that regulatory factors other than Y1 (e.g. additional MYB transcription factors) control the accumulation of 3-DOAs in leaves in response to pathogen attack.

3.1.9 Production of 3-DOA in Various Tissue

Based on gene expression studies, Mizuno et al. 2012 found that 3-DOAs, not anthocyanidins, accumulated in leaf tissues after infection with fungal pathogens. Although previous reports suggest that transcription factors likely differ between leaf and pericarp 3-DOA expression, the expression of these genes in the secondary plant metabolite biosynthesis pathways are likely to differ between plants with different pericarp colors.

3.1.10 Objectives

This project aims to summarize studies related to the inheritance of black pericarp phenotype and the effect of panicle shading on this trait. Additionally, this chapter provides an initial characterization of gene expression in pericarp of black and red-seeded sorghum to yield insight into the dynamic changes in transcription of the

flavonoid pathway genes during the development of black pericarp. Although the black pericarp trait in grain sorghum has been investigated using classical quantitative genetics approaches (Pfeiffer and Rooney, 2016), a molecular examination of global transcriptional activity during the development of this unique phenotype of pericarp has not previously been characterized. The objective of this research is to gain further insight into the genetic inheritance, environmental perturbation, and actual genes that control the black pericarp phenotype which may aid breeding sorghum with enhanced 3-DOA accumulation for human consumption.

3.2 Materials and Methods

3.2.1 Plant Material

The plant material for the project was created in the 2012 Weslaco, Texas fall nursery. In June of 2012, a parental cross between Tx378 and Tx3362 was made to create the F_1 seed in College Station, Texas. The F_1 seed was then self-pollinated to create segregating F_2 individuals in fall 2012 at Weslaco, Texas. B.Tx378 is a red pericarp inbred line released in 1965 by the Texas Agricultural Experiment Station (TAES) (Stephens and Karper, 1965). Tx3362 is an inbred line with a black pericarp that was released in 2013 by the Texas A&M AgriLife Research Sorghum Breeding Program (Rooney et al., 2013b). It was developed from a cross between Shawaya #2 and RTx430, a common restorer line with good combining abilities (Miller, 1984b).

3.2.2 Population Development

Approximately 300 plots of the Tx378/Tx3362 F_2 seed was planted in College Station in the summer of 2015. From these plots, 360 randomly-selected individual plants were self-pollinated at flowering with a typical pollinating bag. Because the pollination bag blocked sunlight to the developing grain, this material was unable to be phenotyped at the F_2 generation. All other individuals planted in the 2015 College Station environment were not bagged and pollinated naturally. After grain maturation, these approximately 15,000 remaining $F_{2:3}$ individual plants were phenotyped by visual examination of pericarp color. Sixty-nine plants with the blackest phenotype observed were selected. An additional 50 plants were selected for their red phenotype. The number of F_2 plants that were evaluated was necessary to find the desired black phenotype, but moving every plant forward would be too large of a research project. Therefore, the black and red pools and the random population were created to enrich the information from a small number of individuals.

The 69 black selections varied in their intensity of black color. To determine the relative darkness of these 69 individuals, a panel of sorghum scientists individually rated the panicles on a 1-5 scale (1 = black, 5 = red). Those panicles with an average score of < 1.5 were considered "true blacks."

All F_3 seeds were then replanted in College Station in the summer of 2016. Ten plants from each F_3 plot were phenotyped, using a colorimeter. Several of the remaining plants in the plot were self-pollinated to advance to future generations. Based on the phenotyping, only 45 of the blackest individuals were advanced to future generations.

Also, 30 of the reddest individuals and an additional 82 F_3s from the randomly selected F_2 pool were advanced to future generations. All other lines were dropped from the experiment, except the parents of the population.

The F_4 seeds were planted and selfed in Weslaco in the Fall of 2016 to create F_5 seed. Since sunlight levels are known to affect the expression of the black pericarp trait, this environment could not be accurately phenotyped as it was an off-season nursery.

The F_5 seeds were then grown in Weslaco and College Station in the summer of 2017 for phenotyping by the colorimeter. In Weslaco, the entries were grown in a randomized complete block design with two replications, and only open-pollinated panicles were phenotyped. In College Station, the entries were not replicated; both open-pollinated and bagged/self-pollinated plants were phenotyped.

In all environments, plots were irrigated when necessary to maximize growth and minimize potential drought stress. The agronomic practices used were standard (e.g., fertilization, tillage, pest control) for grain sorghum production in the area. In College Station, all plots were 5.5 meters in length with rows spaced 0.76 meters apart; in Weslaco, all plots were 5.18 meters in length with rows spaced 1.02 meters apart.

3.2.3 Threshing and Phenotyping

All panicles were hand harvested and then threshed on using an ALMACO BT14 Belt Thresher (ALMACO, Nevada, Iowa, USA). Five open pollinated panicles were harvested from each RIL in the F_5 generation in both environments. In College Station, five additional self-pollinated/bagged panicles were also harvested from each RIL. The color of the grain was quantitatively measured using a CR-410 Colorimeter (Konica

Minolta Sensing Americas, Inc., Ramsey, NJ). Each data output from the colorimeter was an average of three successive measurements. Additionally, each sample was measured three times from three different views and averaged to create a single data point. Therefore, samples were measured nine times in total to increase the accuracy of the data. Measurements are expressed in accordance with the *Comission Internationale de l'Eclaorage* (CIE) (*Comission Internationale de l'Eclaorage*, 2004) as *L*a*b** color space values. *L** is a lightness value (0 = black, 100 = white); *a** indicates green or red (-*a** = greenness, +*a** = redness); *b** indicates blue or yellow (-*b** = blueness, +*b** = yellowness).

3.2.4 Index Selection

Since *L** and *a** values both play a large role in determining the black color, an index rating was calculated to weight the two metrics equally. All *L** data points were summed together, and all *a** data points were summed together. Division of the grand sum of *L** divided by *a** equals the value that each *a** data point should be multiplied by to bring parity to both values. For each data point, the *L** value and the adjusted *a** value were summed together to create an index value. The smallest index values were considered the darkest (blackest) samples. The adjustment factor for *a** values was calculated independently for each environment.

3.2.5 Basic Statistics and ANOVA

Analysis of variance was calculated using an all random model of X = environment + genotype + genotype x environment + replication[environment] + error, in which X = *L**, *a**, or *b** colorimeter values. ANOVA was calculated in JMP Pro

13.1.0 (SAS Institute, Cary, NC) using the Fit Model platform. The REML (Restricted Maximum Likelihood) estimation method and the Standard Least Squares personality were applied. Means, standard deviations, minimums, and maximums were calculated in JMP using the Tabulate platform.

3.2.6 Heritability Estimates

Heritability estimates were calculated using the 82 randomly segregating individuals. Repeatability estimates were also calculated using all RILs in the experiment (black, red, and randomly segregating populations). Heritability and repeatability estimates were calculated on a plot basis using the formula, $h^2 = \frac{\sigma_g^2}{\sigma_{err.}^2+\sigma_{gxe}^2+\sigma_g^2}$, in which σ_g^2 is the genetic variance, $\sigma_{err.}^2$ is the error variance, and σ_{gxe}^2 is the genetic x environment variance (Fehr, 1987). Entry mean heritability and repeatability were calculated by the formula, $h^2 = \frac{\sigma_g^2}{\frac{\sigma_{err.}^2}{n_{rep}\,x\,n_{loc}}+\frac{\sigma_{gxe}^2}{n_{loc}}+\sigma_g^2}$, in which n_{rep} is the number of replications and n_{loc} is the number of locations (Fehr, 1987).

3.2.7 Tissue Harvest for RNAseq

To obtain tissue for RNA-seq, panicles were harvested from F_3 plants in the summer of 2016 in College Station. The tissue was harvested early July approximately 20 days post-anthesis. The two RILs that were utilized were RIL_CS562, the darkest black RIL in the population, and RIL_CS448, which expressed a typical red color. Pericarp tissue was excised from the developing seed minutes after field harvest and immediately frozen in liquid nitrogen, and stored at -80° Celsius until processing.

3.2.8 Sequencing & Calculation of Differential Expression

Pericarp tissue was pulverized in liquid nitrogen with a mortar and pestle. Approximately 100 mg of pulverized tissue was used for each RNA extraction, which was performed according to the recommendations of the PureLink Plant RNA Reagent (Thermo Fisher®) manual. The samples were subsequently treated with the Turbo DNA-free Kit (Ambion®) to remove residual DNA contamination. RNA integrity was assessed by 1.0% agarose gel electrophoresis. The samples were quantified at 260nm with a DeNovix spectrophotometer (DeNovix Inc., Wilmington, DE, USA). Samples that exhibited high levels of RNA integrity and purity were used for TruSeq library preparations. Libraries were constructed for each sample using a cDNA Synthesis kit (Illumina Inc., San Diego, CA, USA) following the manufacturer's instructions. One lane of paired-end (2 × 150 bp) sequencing of the cDNA libraries was performed on the Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA). Library preparation and sequencing was performed by AgriLife Genomics and Bioinformatics Services (Texas A&M University). Sequence cluster identification, quality pre-filtering, base calling and uncertainty assessment were done in real time using Illumina's HCS 2.2.58 and RTA 1.18.64 software with default parameter settings.

Illumina reads were imported into the CLC Genomics Workbench version 8.5.1 (Qiagen, Valencia, CA, USA), trimmed to 64 bp, and mapped to the *Sorghum bicolor* BTx623 reference genome (Sbicolor_255 v3.1, www.phytozome.jgi.doe.gov). Pathway analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG)

database and a single enrichment analysis was also perfomed using the tool available on the AgriGo website (http://bioinfo.cau.edu.cn/agriGO/analysis.php).

3.2.9 Analysis of Expression of Flavanoid Biosynthesis-Related Genes

Expression of genes previously reported (papers) to be involved in 3-DOA, anthocyanin, and flavones biosynthesis pathways were examined in the red and black pericarp RILs. The EDGE analysis (Extraction of Differential Gene Expression) tool was utilized to determine differential expression. Differentially expressed genes were defined as having a log2 fold change with a false discovery rate adjusted *p*-value <0.05.

3.3 Results

3.3.1 Black deterioration

From the 69 selected black panicles in our F_2 plots, only nine F_2 individuals were determined to be very black from visual score ratings. The very black definition was defined as having an average visual score rating of less than 1.5 (Table 6). Only eight of nine very black panicles were advanced into the F_4 generation. There was not sufficient F_3 seed for one of the darkest panicles (RIL_CS508). Two more RILs were not advanced to the F_5 because the phenotypes in the F_3 were not consistently black (RIL_CS558 and RIL_ CS564) (Table 6).

3.3.2 Basic Statistics

In the F_5 generation, the environment source of variation played a small role in the experiment. This is demonstrated by a low percentage of total variance in the analysis of variance table (Table 7). However, environment was statistically significant for *a**, *b** according to the Wald *p*-value. Genotype, genotype x environment, and

Table 6. Phenotypic Comparisons between 2015 and 2016.

A comparison of F_2 and F_3 RIL lines grown in 2015 College Station and 2016 College Station respectively. The *L**, *a**, and visual score ratings are provided to demonstrate how phenotypes shifted. The "Result" column indicates the status of the RIL line for planting in the F_4 generation.

	F_2 (2015)			F_3 (2016)			
	L*	a*	Mean Visual Score	L*	a*	Visual Score	Result
RIL_CS508	30.5	7.45	1	32.033333	9.763333	1	Seed not sufficient
RIL_CS538	30.48	9.29	1.214285714	31.466667	10.403333	2	Advanced to F_5
RIL_CS553	31.07	9.58	1.142857143	31.08	13.196667	3	Reddish, still advanced to F_5
RIL_CS556	32.22	9.82	1.214285714	30.896667	10.38	2	Advanced to F_5
RIL_CS558	34.82	10.93	1	32.57	10.63	3	Advanced to F_3
RIL_CS562	30.19	8.34	1	27.33	4.866667	1	Advanced to F_5
RIL_CS563	26.78	5.02	1	31.006667	13.866667	2	Advanced to F_5
RIL_CS564	32.47	9.81	1	30.58	9.95	2	Advanced to F_3
RIL_CS567	33.49	10.22	1.285714286	30.246667	9.813333	2	Advanced to F_5

Table 7. Analysis of Variance of F_5 Generation

Analysis of variance of colorimeter values on F_5 RILs grown in College Station and Weslaco, TX in 2016.

	*L**		*a**		*b**	
	Pct. of Total	Wald p-value	Pct .of Total	Wald p-value	Pct. of Total	Wald p-value
Environment	5.70	0.5032	0.00	<.0001*	0.00	<.0001*
Genotype	68.86	<.0001*	30.71	<.0001*	44.70	<.0001*
G x E	7.01	<.0001*	26.78	<.0001*	12.43	<.0001*
Rep[Env.]	0.34	0.5078	4.91	<.0001*	5.50	<.0001*
Residual	18.09		37.59		37.37	
Total	100.00		100.00		100.00	

*p<0.05

Table 8. Parental Phenotypes in Each Environment

Comparison of Tx3362 (black parent) and Tx378 (red parent) in each F_5 testing environment.

			*L**	*a**	*b**
Tx3362	College Station	Mean	30.20	8.25	6.59
		St. Dev	1.13	1.81	1.68
		Range	28.31 - 31.54	6.38 - 11.31	5.08 - 9.81
	Weslaco	Mean	27.45	5.81	5.52
		St. Dev	0.59	0.77	1.03
		Range	25.64 - 28.51	4.07 - 7.49	3.36 - 7.75
Tx378	College Station	Mean	44.76	16.37	23.86
		St. Dev	0.83	0.80	1.16
		Range	43.24 - 45.74	14.31 - 17.16	21.33 - 25.44
	Weslaco	Mean	43.10	13.66	21.45
		St. Dev	1.06	0.87	1.02
		Range	40.68 - 45.07	11.77 - 15.03	19.42 - 23.18

Table 9. Phenotypic comparisons of pools.

The *L**, *a**, and *b** values of different phenotypic pools and the two parents of the population, grown in College Station and Weslaco in 2016.

	*L**		*a**		*b**	
	Mean	Range	Mean	Range	Mean	Range
Tx3362	28.04	25.64 - 31.54	6.34	4.07 - 11.31	5.75	3.36 - 9.81
10 Blackest	29.91	24.39 - 35.57	7.37	3.49 - 12.85	7.10	2.0 - 13.78
Black Pool	33.50	24.39 - 44.51	10.15	3.49 - 16.46	11.50	2.0 - 22.47
Red Pool	38.93	32.17 - 48.94	13.23	7.76 - 19.52	17.88	9.60 - 24.39
Tx378	43.51	40.68 - 45.74	14.34	11.77 - 17.16	22.05	19.42 - 25.44
Random	36.84	28.7 - 50.32	11.90	5.14 - 17.62	15.28	5.98 - 25.65

Table 10. College Station and Weslaco Comparison

Comparison of index ratings and index rating rankings for each RIL line in the F_5 generation. (NA=no data)

RIL Number	College Station Index Rating	College Station Rank	Weslaco Index Rating	Weslaco Rank	Bulk Type
RIL_CS562	42.08261	1	42.15524	1	Black
Black Parent	49.97693625	2	46.94266833	2	
RIL_CS566	55.11670625	3	52.5396	4	Black
RIL_CS543	56.260186	4	53.11364	6	Black
RIL_CS563	57.317885	5	47.23824	3	Black
RIL_CS561	58.75601357	6	54.02344	7	Black
RIL_CS568	58.787681	7	54.71764	8	Black
RIL_CS111	59.9315075	8	55.05892	9	Random
RIL_CS556	60.574658	9	60.94712	18	Black
RIL_CS112	61.662467	10	60.30824	15	Random
RIL_CS553	61.90844999	11	59.31668	10	Black
RIL_CS557	62.170467	12	NA	NA	Black
RIL_CS345	62.511679	13	63.60352	29	Random
RIL_CS501	64.99328445	14	59.7362	12	Black
RIL_CS248	65.034732	15	61.82572	24	Random
RIL_CS504	65.161924	16	63.90252	30	Black
RIL_CS164	65.20212875	17	67.27572	50	Random
RIL_CS173	65.368549	18	65.94176	40	Random
RIL_CS512	65.646653	19	64.9499	34	Black
RIL_CS567	65.911058	20	52.98684	5	Black
RIL_CS522	65.924205	21	64.75232	33	Black
RIL_CS036	66.714508	22	66.21044	41	Random
RIL_CS519	66.97436	23	61.50908	21	Black
RIL_CS560	67.018512	24	NA	NA	Black
RIL_CS014	67.527566	25	67.703	52	Random

Table 11. Analysis of Covered and Uncovered Effects

Analysis of variance for the difference in *L** value between covered and uncovered panicles in all genotypes of F_5 plants grown in College Station in 2017. Also, the Student t's least square means contrast between covered and uncovered panicles of all F_5 genotypes grown in College Station in 2017.

Source	Mean Square	Prob. > F
Genotype	5.19	0.24
Error	0.48	
Treatment	**Connecting Letters***	**Least Sq. Means**
Covered	A	40.79
Uncovered	B	37.23

*Student t's LS Means Contrast

Table 12. Differences between O.P. and bagged phenotypes.

The difference in mean *L** values between open pollinated and bagged/self-pollinated panicles of the same RILs.

RIL Number	L* Diff
RIL_CS562	-11.93
RIL_CS117	-11.85
RIL_CS534	-9.41
RIL_CS142	-9.29
RIL_CS561	-8.89
RIL_CS294	-8.63
RIL_CS126	-7.98
RIL_CS296	-7.50
RIL_CS036	-7.14
RIL_CS248	-7.13

Table 13. Heritability and Repeatability Estimates.

The heritability and repeatability estimates on both a plot and entry mean basis for *L**, *a**, and *b** values.

	Heritability		Repeatability	
	Plot Basis	Entry Mean	Plot Basis	Entry Mean
*L**	0.56	0.72	0.73	0.85
*a**	0.30	0.46	0.32	0.49
*b**	0.57	0.72	0.50	0.67

replication (nested within environment), were all significant at $p < .05$ for all three colorimeter values (*L**, *a**, and *b**) (Table 7).

Although environment appears to have a small effect on plant color, a closer examination of the red and black parents suggests otherwise. *L** and *a** values for the black parent, Tx3362, demonstrate obvious differences in mean rating and range between College Station and Weslaco environment (Table 8). Tx3362 was less black (lower *L** and *a**) in College Station than in Weslaco. The difference between the two environments is less noticeable in the red parent, Tx378. On average, the red parent was darker in Weslaco according to *L** and *a** values (Table 8).

The lines under study were divided into different categories, and the phenotypic differences can be compared. The ten darkest panicles had *L**, *a**, and *b** values like the black parent, Tx3362 (Table 9). The black pool has a higher mean and higher maximum

values than the ten blackest and Tx3362. The red parent, Tx378, appears to reside in the middle of the *L** and *a** range of the red pool. The random pool does not comprise a range equivalent to the black parent. However, the random pool range does encompass a good range of values of the other categories (Table 9).

3.3.3 F_5 Phenotypes

The two F_5 environments agreed for approximately the ten darkest phenotypes. Both College Station and Weslaco phenotypes had similar values and rankings for these top 10 (Table 10). After the top ten darkest plots, the rankings appear less uniform in the two environments. One of the remaining members of the darkest nine F_2s (Table 6) failed to make the top 25 (RIL_CS538). Additionally, RIL_C563 was labeled as black in the F_2, red in the F_3, became fifth blackest RIL in College Station in the F_5 (Table 6, 10).

3.3.4 Open pollinated versus bagged panicles

The RILs evaluated at the F_5 generation in College Station did not demonstrate a differential response in phenotype between open-pollinated panicles and bagged panicles (Table 11). This is confirmed by analysis of variance in which the difference in *L** value between covered and uncovered panicles was not significant. There was a difference in mean seed color between bagged and un-bagged panicles. This was confirmed in this study by an LS Means Contrast between the covered and uncovered panicles of the entire College Station F_5 location. However, in this study, some RILs had a larger magnitude difference in seed color between bagged and un-bagged panicles than other RILs (Table 12). RIL_CS562 and RIL_CS117 had the largest differences in *L** value between the open-pollinated and bagged plants. Other individuals, including the red

parent Tx378, showed small differences in *L** value between their open-pollinated and bagged plants.

3.3.5 Heritability

The *L** value had the highest heritability among the other color values, *a** and *b** (Table 13). All color values were less heritable when calculated on a plot-basis versus an entry-mean basis. Most repeatability values were higher than the heritability values, which only included data from the selected black and red RIL pools. The one exception to that trend was the *b** value. Overall, the heritability of the pericarp color appears to be moderately heritable (Table 13).

3.3.6 Transcriptome Analysis of Black and Red Pericarp

In the gene expression comparison between a red pericarp (RIL_CS448) and a black pericarp RIL (RIL_CS562), there were 8,333 genes that were differentially expressed (log2 fold change with a false discovery rate adjusted *p*-value <0.05) (Table 14). There were 4,260 genes in total that were significantly down-regulated in the black pericarp (versus red pericarp) while 4,073 genes were up-regulated (Table 14).

In this preliminary study of the expression of the flavonoid biosynthesis pathway in black pericarp, we investigated 53 genes that have been reported by Liu et al. (2010) and Mizuno et al. (2012, 2016) as encoding structural enzymes in the flavonoid biosynthesis pathway of sorghum. These structural genes include early biosynthetic gene that are shared by the various flavanoid classes, and key branch pathway genes that are critical to the synthesis of specific classes of flavanoids (Fig. 8). Of these 53 flavanoid biosynthetic genes examined, 26 were differentially expressed in black versus

Table 14. Transcripts and Loci Differentially Expressed.

The number of transcripts and loci that are differentially regulated in black pericarp tissue in relation to red pericarp tissue

		No. annotated (%)	No. of novel (%)
Up	Transcripts	2,800 (17%)	1,558 (9%)
	Loci	2,256 (13%)	1,230 (7%)
Down	Transcripts	2,719 (16%)	2,093 (13%)
	Loci	2,348 (14%)	1,719 (10%)

red pericarp. Differential expressed genes were observed in early biosynthetic genes and also for genes that encode enzymes critical for specific flavanoid classes. The following sections examine these differentially expressed flavanoid pathway genes beginning with early biosynthetic genes and concluding with genes that regulate the accumulation of specific flavanoids.

3.3.6.1 Early Biosynthetic Genes

The sorghum reference genome of genotype BTx623 harbors eight genes eoncoding PAL (Mizuno et al., 2012) which represents the first enzyme in the phenylpropanoid pathway. Three of the eight annotated PAL genes were differentially expressed in black pericarp when contrasted to red pericarp tissue, including one of the

tandemly duplicated PAL genes on chromosome four (Sobic.004G220300_1, upregulated 10-fold). Perhaps the most noteworthy differentially expressed PAL gene was Sobic.006G148800.1, which was up-regulated 27-fold in the black pericarp when compared to red pericarp tissue (Table 15, Fig. 8.) In examining addition gene families of other early biosynthetic pathway genes, a similar situation was observed with particular family members showing moderate enhanced expression in black pericarp while other gene family members displayed similar expression levels in black and red pericarp (Table 15). In particular, two of three C4H gene family members and four of the eleven 4CL genes were moderately (~3-to-6-fold) upregulated, but the remaining family members were not differentially expressed (Table 15, Fig. 8).

The genome of sorghum contains 15 annotated CHS genes dispersed across chromosomes 2, 5, 6, and 7, including nine tandemly repeated genes on chromosome five. In total five CHS gene family members were upregulated; two adjacent CHS genes (Sobic.005G136200.1 and Sobic.005G136300.1) were up-regulated 97- and 107-fold while three additional CHS genes were up-regulated between 37- and 87-fold in black pericarp (Table 16, Fig. 8). The final early biosynthetic gene leading to the production of the flavanone naringenin is CHI, which represents a family of three genes dispersed across the genome of sorghum. Two of these three CHI gene family members were significantly up-regulated in the sorghum line with black pericarp (Table 16, Fig. 8).

Table 15. Expression of the Early Biosynthetic Pathway Genes

Gene Categories	Transcript Identifier	FDR *p*-Value	Fold Change	Reference
Phenylalanine ammonia lyase (PAL)				
	Sobic.001G160500.1	1.00	1.00	Mizuno et al., 2012
	Sobic.004G220300.1	0.00**	10.04	Mizuno et al., 2012
	Sobic.004G220400.1	0.01*	1.51	Mizuno et al., 2012
	Sobic.004G220500.1	1.00	-1.03	Mizuno et al., 2012
	Sobic.004G220600.1	0.58	1.15	Mizuno et al., 2012
	Sobic.004G220700.1	0.12	1.34	Mizuno et al., 2012
	Sobic.006G148800.1	0.00**	27.54	Mizuno et al., 2012
	Sobic.006G148900.1	0.00**	2.22	Mizuno et al., 2012
Transcinnamate 4-monooxygenase (C4H)				
	Sobic.002G126600.1	0.00**	3.20	Mizuno et al., 2012
	Sobic.003G337400.1	0.00**	6.13	Mizuno et al., 2012
	Sobic.004G141200.1	0.62	1.08	Mizuno et al., 2012
4-coumarate:CoA ligase (4CL)				
	Sobic.001G436300.1	0.00**	-2.15	Mizuno et al., 2012
	Sobic.001G465600.1	0.91	-1.10	Mizuno et al., 2012
	Sobic.001G516600.1	0.00**	2.14	Mizuno et al., 2012
	Sobic.002G079000.1	0.99	1.00	Mizuno et al., 2012
	Sobic.004G062500.1	0.04*	1.49	Mizuno et al., 2012
	Sobic.004G272700.1	0.01*	2.23	Mizuno et al., 2012
	Sobic.006G272600.1	0.00**	6.65	Mizuno et al., 2012
	Sobic.007G029300.1	0.00**	3.21	Mizuno et al., 2012
	Sobic.007G089900.1	0.19	-1.74	Mizuno et al., 2012
	Sobic.007G145600.1	0.00**	1.98	Mizuno et al., 2012
	Sobic.010G214900.1	0.01*	1.63	Mizuno et al., 2012

*p-value < .05
**p-value < .01

Table 16. Expression of Additional Early Biosynthetic Pathway Genes

Gene Categories	Transcript Identifier	FDR *p*-Value	Fold Change	Reference
Chalcone synthase (CHS)				
	Sobic.005G136200.1	0.00**	107.11	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G136300.1	0.00**	96.59	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G136450.1	0.08	8.41	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G136600.1	0.22	2.91	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G136800.1	0.51	4.00	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G137000.1	0.13	-1.64	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G137100.1	0.00**	37.13	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G137200.1	0.00**	81.78	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.005G137300.1	0.00**	83.26	H. Liu et al., 2010 & Mizuno et al., 2012
	Sobic.002G115700.1	0.00**	1.79	Mizuno et al., 2012
	Sobic.002G372700.1	0.25	-1.19	Mizuno et al., 2012
	Sobic.005G102100.1	1.00	1.00	Mizuno et al., 2012
	Sobic.005G107800.2	1.00	1.00	Mizuno et al., 2012
	Sobic.006G237000.1	0.42	-1.15	Mizuno et al., 2012
	Sobic.007G058900.1	1.00	1.15	Mizuno et al., 2012
Chalcone isomerase (CHI)				
	Sobic.004G109000.1	0.06	-1.75	Mizuno et al., 2012
	Sobic.008G030100.1	0.00**	8.96	Mizuno et al., 2012
	Sobic.001G035600.1	0.00**	9.36	H. Liu et al., 2010

*p-value < .05
**p-value < .01

Table 17. Expression of Late Biosynthetic Genes

Gene Categories	Transcript Identifier	FDR *p*-Value	Fold Change	Reference
Flavonoid 3'hydroxylase (F3'H)				
	Sobic.004G200800.1	0.00**	47.52	Mizuno et al., 2016
	Sobic.004G200900.1	0.00**	18.30	Mizuno et al., 2012
	Sobic.004G201100.1	0.00**	35.90	Mizuno et al., 2012
	Sobic.009G162500.1	1.00	-1.11	Mizuno et al., 2012
Flavanone 4-reductase (FNR)				
	Sobic.006G226800.1	0.00**	31.46	Mizuno et al. 2016
Flavone Synthase II (FSII)				
	Sobic.002G000400.1	0.00**	17.69	Mizuno et al., 2016
Flavanone 3-hydroxylase (F3H)				
	Sobic.006G253900.1	1.00	-1.45	H. Liu et al., 2010
Dihydroflavonol 4-reductase (DFR)				
	Sobic.003G230900.1	0.29	-5.54	H. Liu et al., 2010
	Sobic.003G231000.1	0.00**	-21.80	H. Liu et al., 2010
	Sobic.004G050200.1	0.00**	34.39	H. Liu et al., 2010
	Sobic.009G043800.1	0.00**	2291.49	H. Liu et al., 2010
Anthocyanidin synthase (ANS)				
	Sobic.004G000700.1	0.63	2.53	H. Liu et al., 2010

*p-value < .05
**p-value < .01

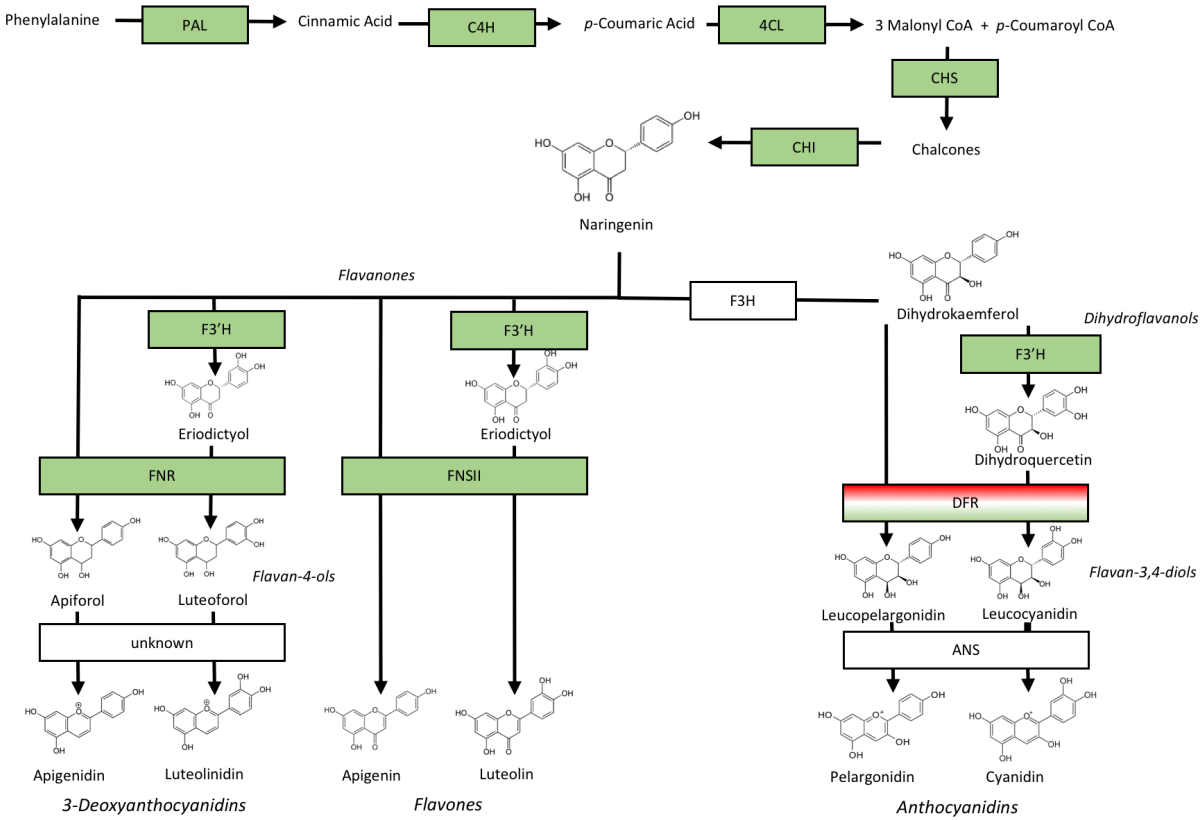

Figure 8. 3-DOA Biochemical Pathway

Biochemical pathway for 3-Deoxyanthocyanidins. Genes highlighted in green denote up-regulated expression in black pericarp sorghum. Green and red together denote a mixture of up- and down-regulation.

It should also be noted that several of the gene family members of early biosynthetic genes were moderately down-regulated (2-fold reduction) in black pericarp; nevertheless, the vast majority of differentially expressed genes were up-regulated in black pericarp. This observation is consistent with the elevated flavanoid levels (flavones and 3-DOAs) within the pericarp of black sorghum compared to flavanoid levels in red- or white-seeded sorghums. In addition, a comparison was made between differentially expressed early biosynthetic genes observed in the present with those early biosynthetic gene family members reported to be differentially expressed by pathogen attack in sorghum leaves (Mizuno et al. 2012). In examination of the CHS gene family, five CHS gene family members induced in sorghum leaves by pathogen attack (reference) were also up-regulated in black pericarp. The exceptions were two CHS genes (Sobic.007G058900_1, Sobic.005G137000.1) that were induced in infected sorghum leaves but not up-regulated black pericarp (Table 11). Similarly, the two CHI family members reported by Mizuno et al. (2012) to be induced in sorghum leaves by pathogens were also moderately up-regulated in black pericarp. By contrast, Mizuno et al. did not detect any members of PAL, C4H, or 4CL gene families to be induced in sorghum leaves, which is in contrast the present results for these genes families in black pericarp.

3.3.6.2 Late Biosynthetic Genes

From the flavonoid precursor naringenin, there are various branches in the pathway to create 3-DOAs, flavones, or anthocyanidins (Fig. 8). The F3'H gene is responsible for converting naringenin to eriodictyol, and thus the synthesis of

luteolinidin (3-DOA) and luteolin (flavone). Three tandemly repeated F3'H gene family members were upregulated between 18- and 47-fold in black sorghum pericarp (Table 17, Fig. 8) while an additional F3'H gene residing on chromosome 9 was not differentially expressed. It should be noted that only one of the three tandemly F3'H genes on chromosome 4 (Sobic004G200800_1) was up-regulated (~6 fold) in response to pathogen attack indicating that the regulation of the flavonoid biosynthetic pathway in pericarp and leaves has unique features.

The DFR consist of four putative paralogs and an additional flavanone reductase (FNR) identified as a committed step in 3-DOA biosynthesis in sorghum leaves under pathogen attack (Table 17, Fig. 8). The flavonoid reductase gene (Sobic.006226800_1) was also up-regulated 31-fold in black versus red pericarp. A DFR gene family member (Sobic.009G043800_1), which was not expressed in sorghum leaves (Liu et al., 2010), was up-regulated nearly 2,300-fold in black pericarp versus red tissue suggesting that this DFR may be critical to the accumulation of 3-DOAs in black pericarp. The fact that the substrate specificity of DFR recombinant proteins *in vitro* has been shown to include dihydroflavonols and flavanones lends credence to this hypothesis. It should be noted that an additional DFR on chromosome 4 was also up-regulated 34-fold in black pericarp while one additional DFR was actually down-regulated 21 fold in black pericarp. Based on this wide range of differential expression displayed by DFR genes, additional genetic and biochemical studies are warranted before conclusions can be drawn as to the exact gene family member catalyzing the conversion of flavanones to flavan-4-ols during 3-DOA biosynthesis in pericarp.

While the final step in the 3-DOA synthesis pathway likely involves an anthocyanidin synthase, the exact gene has not been identified. One gene that has been implicated in this final enzymatic step is Sobic.006G226800.1 (Mizuno et al., 2012). This gene was upregulated in black sorghum versus red sorghum by 31-fold and was significant. At this point, the identity of the final enzyme and gene in 3-DOA biosynthesis remains unknown.

In the flavones pathway, a flavone synthase (FNSII) is a critical branch step in flavone biosynthesis, and this gene family consists of four family members. Of these four genes, Mizuno et al. (2012) demonstrated that Sobic.002G000400_1 was the paralog responsible for flavone biosynthesis in tan-colored leaves of sorghum. In black pericarp, this FNSII gene was also upregulated 18-fold in the black versus red pericarp, (Table 17, Fig. 8), which is consistent with the elevated levels of flavones in black pericarp tissues (Dykes et al., 2009)

Within the anthocyanin branch of the flavonoid biosynthetic pathway reside two enzymes, F3H and ANS, which are specific to anthocyanin biosynthesis (Fig. 8). Additionally, Liu et al. (2010) identified a DFR (Sobic.003G230900.1) that is specific to light-induced anthocyanin biosynthesis in sorghum leaves. In black pericarp, the anthocyanin-specific DFR (Sobic.003G230900.1) was not up-regulated nor were two other genes (F3H and ANS) that are involved in anthocyanin production in epicotyls (Table 17, Fig. 8).

3.4 Discussion

3.4.1 Changes in Phenotype by Generation

Several of the blackest nine F_2 individuals shifted phenotypes throughout the creation of RILs. There are several explanations for why this might occur. Since sunlight is involved in the expression of this trait, we cannot phenotype self-pollinated panicles. Because there is still segregation in the F_3 to F_5 generations within plot, plants that are not representative of our phenotype may have been advanced. The loss of black color from the F_2 to F_3 generations may also be explained by outcrossing. Pollen from neighboring red panicles may have partially pollinated the selected, black seeded F_2s. This would not have been noticed until re-phenotyping in the F_3 generation. A third scenario is that unknown modifier genes continue to segregate for the black pericarp trait, eventually being selfed to a dominant red form in some lines.

3.4.2 Changes in Phenotype by Environment

In the F_5 generation, there were phenotypic differences between the two environments. The darkest RILs and the black parent had a consistent ranking in Weslaco and College Station. However, approximately the tenth to twenty-fifth darkest RILs in College Station were not in agreement with the Weslaco results. These changes possibly suggest multiple genes are involved in the black pericarp trait and they were differentially expressed in the two environments. Although location was not a large source of variation, environment may be influencing the expression of the very darkest panicles. Previously research has indicated genotype, environment, and genotype x environment effects for L^*, a^*, and b^* values (Pfeiffer and Rooney, 2016). Even black

sorghum also grown in College Station 2017, but not a part of this experiment, appeared less black than typical seasons. No clear reasoning can be ascertained for why College Station performed differently in that season. The differences are likely attributed to a combination of weather trends and a slightly earlier planting date than typical, which could impact the quality or quantity of light imparted on the panicle during the development of the phenotype. Further studies examining the influence of light (e.g., light quality, fluence, daylength) on the expression of the black pericarp phenotype are required before the variability on trait expression can be fully understood.

3.4.3 Heritability

The heritability values in this experiment are substantially higher than the previous reported values in Pfeiffer and Rooney, 2016. The estimates in the previous publication were calculated on a single plant basis, and these estimates have been calculated on a plot and entry-mean basis. Also, the previous reported values were calculated in a generation means analysis experiment and these values are from a segregating RIL population. Nevertheless, these current estimations confirm previous reports that this trait is moderate to highly heritable and explain in part why success has been achieved in creating parental lines and commercial hybrids with this trait.

3.4.4 Open-Pollenated vs. Bagged Phenotypes

Previous research has indicated that there is a clear effect of light on the expression of the black pericarp trait and it was re-confirmed in this study. Bagged panicles which were shaded from direct sunlight in College Station 2016 were less dark than open-pollinated plants within the same plot. Previous anecdotal (W.L. Rooney,

personal communication) evidence suggests that pericarp color is suspended differently in various genotypes from bagging; however, evidence of this phenomenon was not observed in this study. The differences in phenotypic expression from non-exposure to sunlight can potentially be genetically mapped in future research. This population does not appear to be a worthwhile target for such an endeavor. Further characterization of the effect of shading on different genotypes is warranted. In addition, determining what parameters of light are conditioning full expression of the black phenotype will permit selection of the proper environment/location to conduct the study.

3.4.5 Gene Expression

The biosynthesis of the 3-DOAs, the compounds believed responsible for the black color in pericarp tissue, has been previous explored in sorghum from the perspective of phytoalexin accumulation in leave in response to fungal attack (Liu et al., 2010; Mizuno et al., 2012, 2016). Various studies have examined the expression of early and late flavanoid biosynthetic genes, and these studies have demonstrated that specific genes (and specific gene family members) within the flavanoid pathway are differentially expressed during the hypersensitive response. During the accumulation of 3-DOAs in leaf lesions, early biosynthetic genes (e.g., CHS, CHI) are up-regulated while anthocyanin branch pathway-specific genes are repressed, and these changes occur in conjunction with the upregulation of 3-DOA branch pathway genes. It has been proposed that this repression of genes in the anthocyanin branch pathway in conjunction with the up-regulation of expression of key early biosynthetic genes a 3-DOA-specific branch pathway genes leads to the enhanced accumulation of 3-DOAs in leaf lesions. In

the present study, differential expression of flavanoid biosynthetic pathway genes in black pericarp was also observed; however, genes in the anthocyanin branch and the flavone branch pathway were not repressed in black pericarp. In addition, the induction of 3-DOA accumulation in leaves in independent of light while a clear requirement for light in 3-DOA accumulation in pericarp exists. Thus, the accumulation of 3-DOAs in black pericarp does not appear to be a consequence of light-independent redirecting the precursor naringenin to 3-DOAs at the expense of anthocyanins as proposed for phytoalexin accumulation (Lo and Nicholson, 1998). In fact, Dykes et al. (2009) reported that the predominant flavanoid with red and black pericarp were the 3-DOAs along with flavones while anthocyanins were not reported for sorghum pericarp. Thus, the present results support studies that indicate that the control of 3-DOA production in leaves and pericarp are under different regulatory mechanisms, likely by tissue-specific transcription factors (Ferreyra et al., 2010).

In sorghum leaves, the pathogen-induced accumulation of 3-DOAs occurs independent of the *yellow seed 1* (*Y1)* gene that, in part, controls the accumulation of flavanoids in sorghum pericarp. *Y1* gene is a member of the R2R3 MYB transcription factor family that are instrumental in controlling flavanoid accumulation in numerous plants species and tissues. R2R3-MYB, basic helix-loop -helix (bHLH) and WD40 repeats (WDRs), comprise a regulatory protein complex (designated as MBW complex) regulating multiple flavonoid metabolisms. In the present study, it was determined that the expression of the *Y1* gene in black pericarp was similar to that of red pericarp indicating that the black pericarp phenotype is not simply the result of over-expression

of this MYB transcription factor. As the black phenotype is controlled by a series of recessive genes, it is conceivable that loss-of-function mutations in transcription factors that repress the 3-DOA pathway results in elevated levels of 3-DOAs and thus, black pericarp phenotype in select sorghum genotypes. An elucidation of the genes conditioning this polygenic recessive trait are well beyond the present dissertation, but this study provides a foundation for detailed molecular, physiological, and genetic studies that are required to better understand the expression of this trait.

The up-regulation of flavone and 3-DOA biosynthesis is consistent with results of Dykes et al. (2009), that stated black sorghum had higher 3-DOA levels than red sorghum, and higher flavone levels than most red sorghum. Mizuno et al., 2012 proposed that Sobic.006G226800.1 is the final enzyme in the 3-DOA synthesis pathway. This gene was up-regulated 31-fold in the black sorghum line in comparison to the red line. The Sobic.006G226800.1 gene was listed in the results as an FNR gene as Mizuno et al. 2012 and 2016 provide conflicting reports of its role in 3-DOA synthesis. Nevertheless, our results give further evidence of this genes' involvement in the production of 3-DOAs regardless of its function. The most intriguing observation in the present transcriptome analysis was the differentially regulation of sorghum gene Sobic.009G043800.1, a DFR gene, up-regulated nearly 2300-fold in black versus red pericarp. Since 3-DOAs are the primary flavonoid present in black sorghum, we hypothesized that a gene specific to the 3-DOA pathway such as a flavonone-4-reductase would likely be gene up-regulated rather than a class of gene more notably involved in anthocyanin biosynthesis. However, the classification of Sobic.009G043800.1 gene as

encoding a dihydroflavonol-4 reductase, is based on existing knowledge of sequence homology rather than biochemical analysis of substrate specificity of the enzyme. Thus, *in vivo*, this enzyme may function as a key enzyme within the 3-DOA biosynthetic pathway of pericarp. Indeed, Liu et al. (2010) did not detect expression of gene Sobic.009G043800.1 in sorghum leaves, either in control tissues or in response to fungal attack. The strong pericarp-specific expression of Sobic.009G043800.1 and its up-regulation in black versus red pericarp seems to suggest it may be a candidate for the 3-DOA pathway, but further biochemical and genetic studies are warranted before any conclusion can be made of this gene (or other genes) and their involvement in 3-DOA biosynthesis in black pericarp.

3.4.6 Number of Genes

Previous research from Pfeiffer and Rooney (2016) suggested between two and twelve genes are involved in the black pericarp trait in sorghum. This transcriptome analysis experiment suggests at least 26 loci of nine gene categories are associated with the trait. The two lines under investigation are divergent for the pericarp color trait, but they are not isogenic lines. Differences in gene regulation between the two lines could be related to other unrelated biochemical processes.

3.5 Conclusions

In conclusion, the black pericarp trait showed different expression in two F_5 environments and was found to be moderate to highly heritable. This study was the first of its kind to explore 3-DOA biosynthesis gene expression patterns in the pericarp of sorghum grains. The results indicated that 3-DOA production in the pericarp is

controlled by many of the same genes involved in leaf 3-DOA production. Future research should utilize the phenotypic information collected in this study to map chromosomal regions associated with the black pericarp trait using a modified bulk-sergeant analysis approach. Gene expression patterns could be clarified using bagged and un-bagged panicles of the black parental line, Tx3362, instead of the approach taken in this study using separate black and red recombinant inbred lines. Since the expression of the trait is so easily controlled by sunlight, sampling pericarp tissue from bagged and unbagged panicles of the same genotype could help target the essential genes involved in the biochemical pathway.

4. INVESTIGATION OF THE HIGH-PROTEIN DIGESTIBILITY PHENOTYPE IN GRAIN SORGHUM

4.1 Introduction

4.1.1 Overview

Grain sorghum [*Sorghum bicolor* (L.) Moench] has several unique production advantages compared to other cereals. Sorghum is relatively successful in hot and dry environments and performs well when grown under low-input cultural practices (Doggett, 1988). In the Western world, sorghum is mainly utilized as animal feed and forage and recently in the production of biofuels (Sarath et al., 2008; Bean et al., 2011). However, about one-third of total sorghum production is used as a food grain, generally in regions with limited food supplies (Awika and Rooney, 2004; Bean et al., 2011). Approximately half of the worldwide production of sorghum is in semi-arid regions of Africa and Asia where sorghum has traditionally been grown as a staple crop for animal feed and for human consumption (Chamba et al., 2005).

Compositionally, sorghum grain contains similar levels of starch and protein as other major cereal crops like wheat and corn (Bansal et al., 2008), but it has diminished levels of digestible protein. The protein digestibility of wheat, maize, and rice are approximately 85% digestible where sorghum is only approximately 60% (Mertz et al., 1984). The lower level of digestible protein is a major limiting factor in the use of sorghum for human consumption, animal feed, ethanol production and other industrial uses (Tesso et al., 2008; Bean et al., 2011).

There are nine essential amino acids that cannot be synthesized by the human body. The most limiting of these indispensable amino acids in cereal grain diets is lysine (Young et al., 1998). Compared to other cereal grains, sorghum has a very low lysine content with just 35-90% of the other cereals (Henley et al., 2010). The sorghum inbred line RTx436 contains 0.18 g lysine/100 g flour compared to white maize which is reported to have 0.30 g lysine/100 g flour and pearl millet which has 0.48 g lysine/100 g flour (Henley et al., 2010). Quality Protein Maize contains lysine levels of 38-40 mg/g protein, compared to just 21.8 mg lysine/g protein in Tx436 and 29.6 mg lysine/g protein in the P521-*opaque* sorghum line (Henley et al., 2010). The lower levels of lysine are mostly because the major protein storage proteins of sorghum are essentially lysine free (Belton et al., 2006).

As ethanol production from grain sorghum has risen significantly in recent years, high protein digestibility may also have significant economic benefits for that end-usage (Wu et al., 2010). A more highly digestible sorghum would lower energy and enzymatic inputs to digest grain proteins and thus lower the cost of production. Easier digestion will decrease fermentation times and increase the capacity of an ethanol plant. The improved protein digestibility will likely also increase the availability of starch and would increase the amount of ethanol that could be produced from the same quantity of grain (Wu et al., 2010).

4.1.2 Factors Affecting Protein Digestibility

The causes of the lower digestibility in sorghum grain appear to be multifactorial (Duodu et al., 2002). As reported by Ejeta et al. (1987), higher lysine content is

correlated with higher protein digestibility rates (Tesso et al., 2008). Additionally, as the germ and pericarp fractions of grains become less of the total proportion of the caryopsis, protein digestibility levels increase (Duodu et al., 2002, 2003).

Grain sorghum is classed into two groups; those with a pigmented testa and those without a pigmented testa. The presence or absence of the layer in the pericarp is related to the concentrations of condensed tannins present therein. Grain with a pigmented testa are very high in condensed tannins and these tannins are useful for protection against birds and other pests but reduces protein digestibility as well. However, sorghums without a pigment testa have low levels of these tanins compounds and thus the associated anti-nutritional properties (Serna-Saldivar and Rooney, 1995). Tannin molecules can bind up to twelve times its weight of protein, so even sorghum with relatively low tannin levels could theoretically bind much of the protein in the grain (Butler et al., 1984).

The influence of tannins is confounded because there are significant differences in protein digestibility among high-tannin sorghums, as well as low-tannin sorghums (Elkin et al., 1996). Sorghum is also rich in flavonoids and other phenolic compounds which binds to proteins by their hydroxyl groups. Unlike tannins, there is no evidence that other phenols reduce protein digestibility (Serna-Saldivar and Rooney, 1995). Phytate is the phosphorus storage molecules in seed. Phytate has great affinity to proteins and through binding may also lower the protein digestibility in sorghum (Knuckles et al., 1985).

Other factors are potentially involved with protein digestibility include the

interactions between proteins and dietary fiber. Many proteins bind to the cell walls, which are not digestible by humans (Bach Knudsen and Munck, 1985; Duodu et al., 2003). Kafarin proteins in the endosperm are also highly associated with the starch granules. The kafarin-starch connections could lower protein digestibility when cooking as the starch becomes gelatinized (Duodu et al., 2003). Another theory is that protein degrading enzymes are not able to digest the proteins in sorghum because sorghum proteins are in a non-aqueous environment. Enzymes that degrade sorghum proteins operate in aqueous environments but sorghum kafarins are very hydrophobic (more hydrophobic than maize zein proteins). Therefore, sorghum proteins are less available to proteases (Duodu et al., 2003).

Depending on the individual characteristics of each sorghum line or hybrid, different factors will have more or less relative impact on protein digestibility. However, the most important factor likely is the proteins themselves. Protein-protein interactions, or protein crosslinking, appears to have the greatest effect on the digestibility of sorghum proteins (Duodu et al., 2003).

4.1.3 Sorghum Seed Storage Proteins

There are many seed storage proteins in cereal grains. The seed storage proteins in sorghum are categorized based on solubility (Landry and Moureaux, 1970; Jambunathan et al., 1975; Bansal et al., 2008). Albumins are classified as water soluble seed storage proteins; globulins are the class of seed storage proteins that are salt soluble; glutelins are the proteins that are soluble in dilute alkali (Hamaker and Bugusu, 2003; Belton et al., 2006; Bansal et al., 2008). The kafirins, which are also referred to as

the prolamins, are the largest group of seed storage proteins and are characterized as being alcohol soluble (Aboubacar et al., 2001).

Kafarins comprise of 68-73% of protein in whole grain flours and 77-82% of the protein in the endosperm (Belton et al., 2006). Kafarins contain high levels of the amino acids proline and glutamine (Bansal et al., 2008). The kafarin proteins accumulate in the lumen of the endoplasmic reticulum and form protein bodies (Hamaker and Bugusu, 2003). Because kafarins are overwhelmingly the largest seed storage protein class in sorghum, they primarily control protein quality (Chamba et al., 2005) and therefore have been the main focus of research.

The extracted kafarins have a wide range of molecular mass between 15,000 and almost 30,000 (Belton et al., 2006). The nomenclature of different types of kafarins follow the same classification system as maize zein proteins. Sorghum kafarins can be divided into four types: α-, β-, γ-, and δ- kafarin.

The α-kafarins comprise of 80-84% of total kafarin in the hard (corneous, vitreous) endosperm (Elkin et al 1996) and 66-71% of total kafarin in soft (floury, opaque) endosperm (Watterson et al., 1993; Bansal et al., 2008). They have low levels of cysteine (Hamaker and Bugusu, 2003). Approximately 20 genes have been identified that control the α-kafarins production in sorghum (Belton et al., 2006).

The β-kafarins fraction accounts for 7-8% of total kafirins in hard endosperm tissue and 10-13% in soft endosperm (Watterson et al., 1993; Bansal et al., 2008). It is believed that β-kafarins are under the control of a single gene (Chamba et al., 2005)

which corresponds to a single gene for maize β-zein. The β-kafarins contain a high level of cysteine (Belton et al., 2006).

The γ-kafirins are unique as they are soluble in water as reduced subunits, however these kafirins are insoluble in water in their native state due to disulfide bonding (Bansal et al., 2008). The γ-kafirins comprise of 9-12% and 19-21% of the total kafirin composition in hard and soft endosperm types, respectively (Watterson et al., 1993; Belton et al., 2006; Bansal et al., 2008). Only one gene is responsible for γ-kafirin production (Bansal et al., 2008). Like the β-kafirins, the γ-kafirins are also cysteine rich (Hamaker and Bugusu, 2003). Ironically, the γ-kafirins are the most hydrophobic of all the kafirins that are observed at detectable levels, despite γ-kafirins being the only water-soluble kafarin (as reduced subunits) (Belton et al., 2006).

The lesser-known δ-kafarin are believed to comprise <1% of total kafirin content (Laidlaw et al., 2010). Only one gene has been identified (Izquierdo and Godwin, 2005) to control this type. The δ- kafarins are characterized by a very high methionine content (Belton et al., 2006), but a lack of cysteine (Wong et al., 2009).

4.1.4 Kafarin Protein Body Conformation

The α-kafarins are the most digestible type of kafarin proteins and are located in the interior of the protein bodies in wildtype cultivars (Shull et al., 1992; Oria et al., 2000; Wong et al., 2009). The γ- and β-kafirins are less digestible (Oria et al., 1995a; b) and in the wild-type lines, they enclose the α-kafarins within the protein body by disulfide bonding (Shull et al., 1992). Other material in the endosperm forms a coating

around the protein body adhering the protein bodies to the starch matrix (Hamaker and Bugusu, 2003). About 80% of endosperm protein is located within these protein bodies (Hamaker and Bugusu, 2003). Even though α-kafarins are the most digestible kafarin type in isolation, the γ- and β-kafirins are degraded first due to their location on the perimeter of the protein bodies (Mazhar and Chandrashekar, 1993). Consequently, the α-kafarins are only digested after γ- and β-kafirin digestion (Oria et al., 1995a).

In mutant lines of sorghum that have highly digestible (HD) protein, a structural conformation leads to the quick digestion of α-kafarins prior to the digestion of γ- and β-kafirins (Weaver et al., 1998; Oria et al., 2000). Oria et al. (2000) reported the HD mutant lines had irregularly shaped protein bodies with deep invaginations into the center of the body. Specifically, the low γ-kafarins accumulated at the base of the invaginations, altering the protein body shape, increasing the protein bodies' surface area, and allowing easier access of protease enzymes to the highly-digestible α-kafarins (Oria et al., 2000). The irregular protein body structure increased protein digestibility approximately 2.5 times after 30 minutes of digestion (Weaver and Cleveland, 2006).

3.1.5 Breeding History of HD Trait

After the discovery of the opaque-2 and floury-2 traits in maize and high lysine barley, Singh and Axtell (1973) scanned the sorghum germplasm collection to identify a high-lysine trait (hl). Unfortunately this trait was found in sorghums with soft, floury endosperm and thus negative qualities for end-use (Henley et al., 2010). Several years later, a high-lysine sorghum was created from a normal, low-tannin variety with hard endosperm and high quality agronomic traits, P-721-N, via chemical mutagenesis

(Guiragossian et al., 1978; Tesso et al., 2006). Unfortunately, the mutant high lysine line, named P721-*opaque*, demonstrated no significant benefit in human feeding trials over normal sorghums despite a 60% increase in lysine content (Guiragossian et al., 1978). This breakthrough led researchers to focus not on the creation of high protein lines, but rather highly digestible protein lines (Henley et al., 2010).

The high-lysine mutant, P721-O was crossed to lines with superior agronomic qualities and wild-type protein levels and several new lines were developed with notably improved protein digestibility. The new lines also have good food-grade qualities and a hard endosperm (Weaver et al., 1998). One of the lines that was created from crossing P721-O with an elite-line is P850029, which showed remarkably higher cooked and uncooked protein digestibility than typical sorghum cultivars (Weaver et al., 1998). The difference between P852009 and the wild-type was a change in protein body conformation in which increased surface area aided enzyme accessibility to the quickly digestible α-kafarins (Oria et al., 2000). However, the benefit of the HD trait was accompanied with low-quality, soft, floury, endosperm grain (Bean et al., 2006).

Subsequent efforts to create a highly digestible, hard endosperm line have been mildly successful. A mutant identified in F_6 lines created from crosses between P721-*opaque* and hard endosperm lines demonstrated high digestibility, with intermediate hardness and a partial vitreous phenotype (Tesso et al., 2006). This suggests that with further determination, creation of an HD/hard endosperm line is possible (Tesso et al., 2006).

Today, transgenic approaches are being utilized to create highly digestible sorghum cultivars. The African Biofortified Sorghum (ABS) project is funded by the Bill and Melinda Gates Foundation under the Grand Challenges in Global Health Initiative (Henley et al., 2010). The goal of the project is to provide better nutrition to young children in sub-Saharan Africa. The project measures the value of protein using a formula that accounts for true protein digestibility and amino acid score, called "low Protein Digestibility Corrected Amino Acid Score" (PDCAAS). The transgenic biofortified sorghum claims to have double the PDCAAS of normal sorghum lines (Henley et al., 2010). The project accomplishes HD improvement through RNA interference (RNAi) technology by suppressing various combinations of kafarin tyes. The RNAi improved lines show irregularly shaped protein bodies similar to non-transgenic HD lines previous discussed (Henley et al., 2010). Although the potential impact of this project is large, it remains to be seen if the world will benefit from these genetically modified cultivars due to political challenges.

4.1.6 Relationship to Grain Hardness

Watterson concluded that the soft, sorghum lines with large fractions of opaque endosperm have less protein content due to lower levels of kafarins (Watterson et al., 1993). The hard endosperm types contain more kafarin molecules, because a smaller fraction of the seeds are soft and floury (Watterson et al., 1993) and the hard endosperm areas have greater cross-linking due to the higher levels of kafirins. This creates a denser endosperm matrix that is harder or more corneous (Shull et al., 1990; Ioerger et al., 2007). Conversely, the soft, floury endosperm is less tightly packed and since protein

bodies are not in as close of contact, less cross-linking occurs (Shull et al., 1990; Ioerger et al., 2007). The γ-kafirins, which are most responsible for crosslinking, remain in high quantities in the soft endosperm fractions. Therefore, the difference is not an absence of the γ-kafirins (Ioerger et al., 2007). One hypothesis is that the γ-kafirins must cross-link to other γ-kafirins in the floury fractions (Ioerger et al., 2007). What is clear is that there is a strong correlation between protein cross-linking and endosperm texture. Grain hardness, therefore, may be a function of kafirin composition (Mazhar and Chandrashekar, 1993).

4.1.7 Environmental Effects on HD and Endosperm Texture

The expression of endosperm texture, grain hardness, and protein digestibility traits appear to be heavily influenced by environment. Tesso et al. (2008) highlighted that sorghum lines which normally have vitreous endosperms when grown in temperate environments reverted to a floury textured endosperm when grown in winter nursery at Puerto Rico. When Puerto Rican-grown seeds were planted again in Indiana, they returned to their normal, hard endosperm phenotype (Tesso et al., 2008).

When genotypes were grown in similar locations over multiple years, significant differences were found in sorghum protein digestibility (Weaver et al., 1998). Likewise it is well known that protein levels are also influenced by environment. Differences in farming practices, such as nitrogen fertilizer application, can greatly impact protein level in sorghum grain as well (Bean et al., 2011).

4.1.8 Allelic Nature of HD Trait

Protein digestibility has a unique allelic nature due to kafarin proteins being expressed in the endosperm tissue. This region of the caryopsis is triploid with two copies of the genome of the maternal parent. Therefore, endosperm genes have four potential genotypes at a locus, compared to only three in diploid tissues. According to Tesso (2008), it is difficult to perceive differences between three versus two copies of an allele at a locus with regards to the highly digestible sorghum allele.

4.1.9 Yields of HD Sorghum

Multiple studies have reported that highly digestible sorghum lines yield as high or higher than wild type lines in the same experiment. Jampala (2012) demonstrated that was no yield penalty for lines with the HD trait compared to wildtype material with the same background. In hybrid material, Tesso (2008) observed that HD hybrids were among the highest yielding. Additionally, there was comparable values of protein digestibility when a HD seed parent was crossed by either a wild-type or HD pollinator line (Tesso et al., 2008). The nonsignificant difference of pollinator lines gives breeders even more options for selecting R lines, ensuring good yield and agronomic abilities.

4.1.10 Genetic Study of HD Trait

Winn et. al (2009) conducted a genetic mapping study of the HD trait using an F_4 population crated from the parental cross of the highly digestible inbred line, P850029, and the wild-type inbred line, Sureño. From conducting in-vitro protein digestibility assays on the parental lines and 377 inbred line progeny, a continuous range of digestibility/turbity was observed (Winn et al., 2009). Therefore, the trait was not likely

to be under the control of a single gene. The gradient expression of digestibility phenotype suggests that multiple genes and genetic intersections are controlling this trait, both in positive and negative actions. Transgressive segregants were also observed, especially on the lower-digestibility end of the spectrum (Winn et al., 2009).

A bulked segregant analysis approach was used in which the most digestible lines' DNA were bulked together in one pool and the least digestible lines were pooled together in a separate pool (Winn et al., 2009). Using SSR markers, two significant quantitative trait loci (QTL) were identified on chromosome 1 (Winn et al., 2009). The first QTL referred to as "locus 1" controlled about 29% of total variation and had additive and dominance effects that acted negatively on digestibility, i.e. the presence of locus 1 contributed to a low digestibility (Winn et al., 2009).

The second QTL, known as "locus 2," is located 20 cM away on chromosome 1 (Winn et al., 2009). This locus acted favorably to protein digestibility, i.e. favorable alleles at this chromosomal region increase sorghum protein digestibility. However, it had slightly less control on the trait than locus 1, with 18% correlation (Winn et al., 2009).

Despite these efforts, the two QTL discovered are not highly predictive of protein digestibility in sorghum. Through analysis of genotypic linkage groups, the inbred lines containing the Sureño (low digestibility, *AA*) parental genotype had comparable average digestibility of lines containing the P850029 (HD, *BB*) parental genotype (Winn et al., 2009). Inbred lines with Sureno's genotype at locus 1 and P850029's genotype at locus 2 (*AB)* performed the same as either parental genotype (*AA* and *BB*) (Winn et al., 2009).

Only the progeny with the recombinant genotype that included the P850029 linkage block at locus 1 and Sureno's linkage block at locus 2 (*BA*) demonstrated significantly different phenotypic averages with considerably lower digestibility (Winn et al., 2009).

4.1.11 Objectives

Despite extensive research about sorghum kafarins and protein digestibility from a cereal chemistry and genetics laboratory perspective, relatively few field studies have been conducted to elucidate the nature of this highly impactful trait. One of the main challenges to large field experimentation is the lack of a high-throughput system to phenotype the trait. Therefore, a predictive model using near-infrared spectroscopy based on wet chemistry results will be created to estimate protein digestibility in a rapid, but accurate and precise manner. Additionally, a large recombinant inbred line (RIL) population segregating for protein digestibility will be evaluated in two diverse growing regions for two seasons each in order to discern the level of variation caused not by genetics, but rather the environment. The same population will be used to determine the heritability of the protein digestibility trait. Finally, using a large RIL population and genetic analysis of genotyping-by-sequencing data, the study will elucidate other genomic regions involved with protein digestibility trait in order to conduct marker-assisted selection.

4.2 Materials and Methods

4.2.1 Plant Material

A 287-individual bi-parental recombinant inbred line population was utilized for this study. The seed parent, BTxArg-1, is a waxy, low amylose inbred line. It has a white

pericarp, a tan secondary plant color, and is a popular breeding line for food-grade sorghum (Miller et al., 1992.) BTxArg-1 is characterized as wild-type for protein digestibility. The pollinator parental line, P852009 is a highly digestible, high lysine inbred line (Weaver 1998; Jampala 2012). P850029 has a white pericarp and a pigmented secondary plant color. The populations were grown and phenotyped in the $F_{4:5}$ generation.

4.2.2 Experimental Design

The RILs were grown in two locations, Florence, South Carolina and College Station, Texas. The Texas location was grown for two years, in 2014 and 2015. The South Carolina location was evaluated in 2014 only. These environments are hereafter referred to as SC14, TX14, and TX15. In each of the three environments, two replications were grown in a randomized complete block design. Individual plots in SC14 consisted of 6.1 m single row plots with a row spacing of 0.965 m. In TX14 and TX15, the plots were 5.5 meters long with row spacing of 0.76 meters. In TX14 and SC14, each plot consisted of one row; in TX15, the experiment was grown as two-row plots. The plots were irrigated when necessary to maximize growth and minimize potential drought stress. The agronomic practices used were standard (eg., fertilization, tillage, pest control) for grain sorghum production in the area. About ten panicles from each plot were hand harvested, hand threshed, and cleaned using a Wintersteiger LD180 (Wintersteigher Ag; Reid, Austria.)

4.2.3 Wet Chemistry

Samples were ground into flour using UDY Cyclotec mill (UDY Corporation, Fort Collins, CO, USA) equipped with a 0.1-mm screen. Then, *in vitro* protein digestibility experiments were conducted on uncooked flour samples by the Dr. Scott Bean laboratory (USDA, Manhattan, KS, CS14 dataset, rep 2) and by Brian Pfeiffer in the Texas A&M Cereal Quality laboratory (partial CS14 dataset) according to Mertz et al. (1984). Exactly 200 mg of flour were weighed and placed into test tubes. All samples were run as duplicates (two flour samples from the same source, tested at the same time.) Replicates (multiple samples tested at different times) were also conducted on 21 of the samples. Then, 35 mL of porcine pepsin solution (Sigma P-700 activity; 890 U/mg protein; 1.5 g of pepsin/L in 0.1 *M* KH_2PO_4; pH 2.0) was added to the test tube with the flour. The test tubes were then placed in a shaking warm water bath at 37 °C. After two hours, digestion was stopped via the addition of 2 mL of 2*N* NaOH and centrifugation (20 minutes, 4,900 × g, 4 °C). After centrifugation, the supernatant was discarded, and remnant material in the test tube is washed two times with 20 mL of buffer (0.1*M* KH_2PO_4; pH 7.0) and centrifugation (20 minutes, 4,900 × g, 4 °C). After the two washes, the samples were spread into aluminum pans and air dried for 12-24 hours. The dried pepsin-digested samples, as well as the undigested (untreated) flour, were sent to the Texas A&M Soil Testing Laboratory for analysis via Nitrogen Analyzer. Finally, protein digestibility was calculated using the following equation:

$$Percent\ Digestibility = \frac{(\%\ flour\ protein) - (\%\ digested\ flour\ protein)}{\%\ flour\ protein} \times 100$$

. The standard error of the laboratory (SEL) was calculated on both a best case and worst case scenario. The best-case values were calculated only with the duplicate data points tested with the *in-vitro* protein digestibility protocol at the same time. The worst-case values were calculated with replicate data, run in the assay at different times, and from different flour sources.

4.2.4 NIR Model Creation

All samples from one rep of the TX14 environment were evaluated via pepsin *in vitro* protein digestibility assay. These samples were also scanned as whole grain using a FOSS XDS NIR spectrometer (Foss North America, Eden Praire, MN, USA. Each grain sample was packed (2.54 cm deep) into a rectangular-shaped sample cup (15.24 cm x 3.81 cm x 5.08 cm, length x width, depth) and was scanned twice. Collection of the spectra data was done using the ISIscan software (Version 3.10.05933) set at 32 readings per scan and stored at 2 nm intervals. The average of two scans was used for the equation development. The calibration model was developed using the WinISI software (version 4.0.03770). The WinISI software uses a modified partial least-squares (mPLS) method to create a calibration equation for each of the variables separately. In the development process, several equations were tested which varied in derivative math treatment. Quality parameters of each equation will be considered [coefficient of determination (r^2), standard error in cross-validation (SECV), and coefficient of determination for cross-validation (1-VR) values)]. The equation with the best results (highest r^2 and 1-VR; lowest SECV values) was employed. Using this NIR model, the first rep of TX14, and both reps of TX15 and SC14 were scanned using the same

procedure and percent protein, percent digested protein, and *in*-vitro protein digestibility (IVPD) was predicted.

4.2.5 Genotyping

The genotyping was conducted according to Boyles et al. (2016). DNA from the tissue of two-week-old seedlings F_5 plants was genotyped. All 278 individuals from the BTxARG-1 x P850029 population were genotyped as well as 200 individuals from a RIL population between Tx642 x BTxARG-1, as well as the three parents. These samples were genotyped at the Cornell University Genomic Diversity Facility. Individual DNAs were extracted according to the CTAB protocol (Mace et al., 2003) and digested with the restriction enzyme ApeKI. Digested DNA fragments of 96 individuals barcoded and pooled for sequencing. Five 96-plex GBS libraries were single-end sequenced using an Illumina HiSeq 2500 to obtain 64-bp reads (excluding adaptor sequences). Reads were aligned, called, and imputed with the TASSEL 5.0 GBS pipeline (Glaubitz et al., 2014). *Sorghum bicolor v3.1* reference genome (https://phytozome.jgi.doe.gov/) was used for the alignment of reads. To impute missing genotypes, the TASSEL plugin FSFHap (Swarts et al., 2014) was used, and each population and chromosome was imputed independently. The "cluster" algorithm was used to infer haplotypes. The sites were also filtered when the correlation (r) was < 0.4 with neighboring sites and/or missing > 0.9 across individuals. All other parameters were maintained at their default values. Following imputation, individual sites with minor allele frequency < 0.05 were removed. After these processes, there was 49,617 genome-wide SNPs for the BTxArg-1/P850029 population.

4.2.6 Genetic Map Construction

Genetic map construction was conducted in R/qtl (Broman et al., 2003) and described in detail by Boyles et al. (2016). Genotypes were converted to ABH allele format. SNPs were placed into recombination bins according to Huang et al. (2009). In total, there were 4,154 recombination breakpoints in the population, including 3,337 hom/hom breakpoint and 777 hom/het breakpoints. The recombination bins were treated as individual markers for genetic map construction. In R/qtl, genetic maps for BTx642 and P850029 were converted to cross type "RIL by selfing" (Broman et al., 2003). Heterozygous sites across the data were treated as missing. Bin markers with MAF < 0.05 and markers with severe segregation distortion ($P < 10^{-20}$) were removed. In total, there were 4,149 bin makers.

4.2.7 Statistics

The analysis of variance was calculated in JMP Pro 13.1.0 (SAS Institute Inc., Cary, North Carolina, USA). The Fit Model platform with the Standard Least Squares personality and restricted maximum likelihood (REML) method. An all random model was fitted as follows: response = genotype + environment + genotype*environment + replication within environment + error. Basic statistics (mean, median, min, max, etc.) was calculated on the Tabulate platform of JMP.

The SEL (standard error of the laboratory) was calculated on a best and worst case scenario. The best case calculated only the standard error of duplicate data points which received very similar treatment. The worst case calculated the standard error of genotypes that were run in replicate, defined by the same grain source, but ground to

flour and assayed as different laboratory events. All other statistics related to the NIR was generated by WinISI software.

4.2.8 Heritability Estimates

Heritability estimates were calculated using the 289 randomly segregating RILs grown in TX14, TX15, and SC14. Heritability estimates were calculated on a plot basis using the formula, $h^2 = \frac{\sigma_g^2}{\sigma_{err.}^2+\sigma_{gxe}^2+\sigma_g^2}$, in which $\sigma_g^2$is the genetic variance, $\sigma_{err.}^2$is the error variance, and $\sigma_{gxe}^2$is the genetic x environment variance (Fehr, 1987). Entry mean heritability and repeatability were calculated by the formula, $h^2 = \frac{\sigma_g^2}{\frac{\sigma_{err.}^2}{n_{rep}\,x\,n_{loc}}+\frac{\sigma_{gxe}^2}{n_{loc}}+\sigma_g^2}$, in which $n_{rep}$is the number of replications (two) and n_{loc} is the number of unique environments (three) (Fehr, 1987).

4.3 Results

4.3.1 Basic Statistics

Based strictly on wet chemistry data, the RIL population surveyed (including the parents) had a mean IVPD of 57% (Table 18, Fig. 8). The HD parent, P850029 (70.1%) had an IVPD% ten percentage points higher than the low HD parent, TxARG (60.3%). This difference was significant according to at least squares (LS) means contrast (p = 0.0002). Despite having a much greater protein percent in the undigested flour, P852009 had less protein in the digested flour (4.063 %). However, the different between P850029 and BTxArg-1 (4.382 %) in the digested flour was not significantly different (p = 0.6083) according to the LS Means Contrast.

Table 18. Basic Statistics for Protein Traits

Basic statistics for the F_4 population of TxArg-1 and P850029 grown in College Station, Texas in the summer of 2014.

	Undigested Flour Pct. Protein	Digested Flour Pct. Protein	*In-Vitro* Protein Digestibility
Max	16.97	9.79	77.10
Median	12.18	5.15	57.20
Min	9.21	2.73	37.20
Mean	12.44	5.20	57.88
St. Dev	1.54	1.24	7.57
TxArg-1	11.04	4.38	60.30
P850029	13.64	4.06	70.10

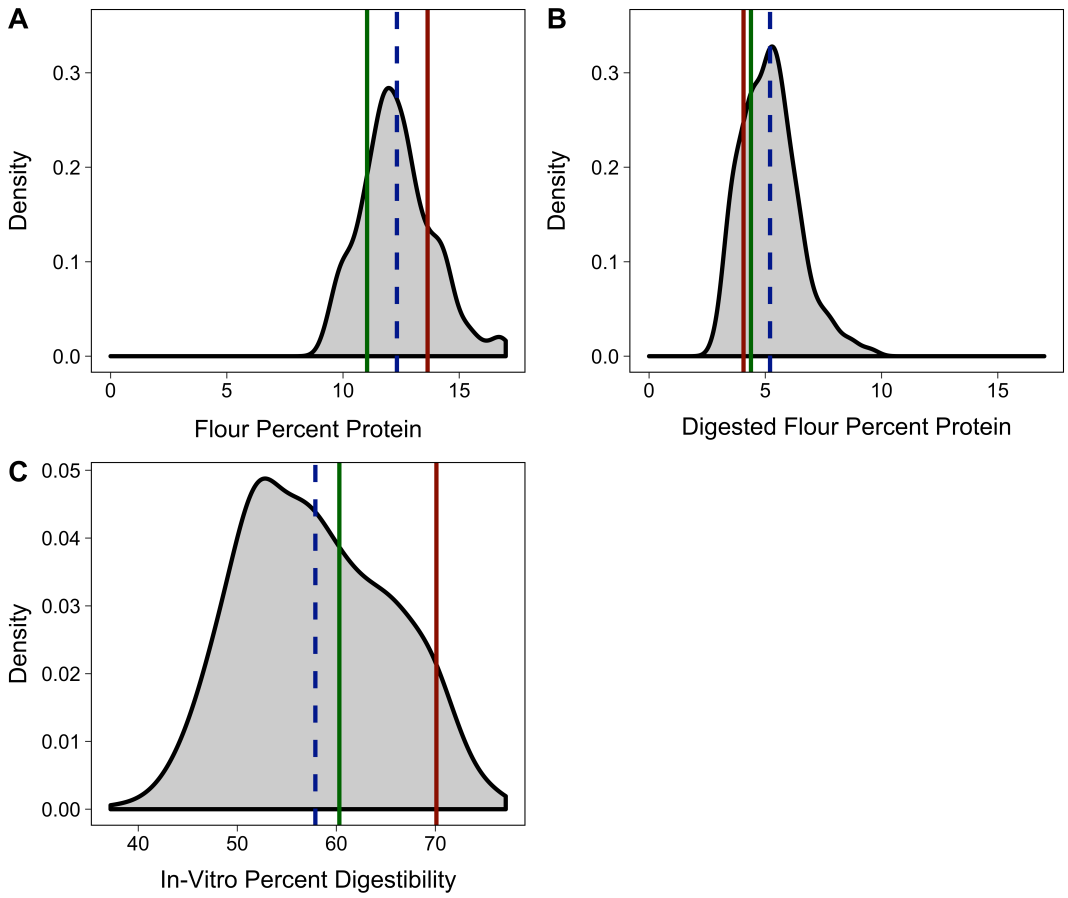

Figure 9. Distribution of Protein Traits

Distribution of three protein traits from phenotyping F_4 population grown in College Station, Texas in the summer of 2014. A) Distribution of flour percent protein; B) Distribution of digested flour percent protein; C) Distribution of in-vitro percent digestibility. The green line represents the value of the non-HD parent, TxARG. The red line represents the value of the HD parent, P850029. The dotted blue line represents the value of the population mean.

Table 19. NIR Calibration Curve Statistics

Statistics on the NIR calibration curve based on wet chemistry from the College Station 2014 environment. Included in the table is the coefficient of determination (r^2), standard error in cross-validation (SECV), and coefficient of determination for cross-validation (1-VR) values. The standard error of the laboratory (SEL) of the wet chemistry results is also listed.

	Undigested Flour Protein Percent	Digested Flour Protein Percent	In-Vitro Pct. Digestibility
SEL (best case – worst case)	0.139 – 1.06	0.223 – 0.8	1.783 – 6.24
r^2	0.344	0.469	0.364
SECv	1.163	0.884	6.579
1-VR	0.295	0.364	0.168
No. in Training Set	251	250	255
No. of Wavelengths	1034	1034	1034

The population was normally distributed (Fig. 8). The mean of the population for undigested flour percent protein (12.44%) sits in between the two parents (TxARG = 11.04%, P850029 = 13.64%) (Table 18, Fig. 8). There are transgressive sergegants on both ends of the spectrum. For digested protein percentages, both parents are skewed to one side of the bell curve. The two parents did not separate well for the digested protein trait, although there appears to be a good amount of variation. For the IVPD trait, both parents again fall on the same side of the graph. Despite this issue, the population did show a good range of values, from 37% to 77% IVPD (Table 18, Fig. 8).

4.3.2 NIR Calibration Curve

The calibration curve with the best fit was an MLPS model. SEC and SECV errors are high compared to the size of the parameters under measurement. The coefficients of correlation were low and poor compared to other NIR estimated traits used in the Texas A&M AgriLife Research Sorghum Breeding Program (Table 19). However, the standard error of the laboratory worst-case values (SEL) was similar to the standard error of the NIR predicted values (SEP).

Using a cross-validation set, the NIR training model predicted the undigested protein percent, digested protein percent, and *in-vitro* protein digestibility values of fifty randomly selected RILs. The IVPD predictions had much higher standard errors and lower r^2 values than undigested and digested percent protein traits (Table 20). Despite large amounts of error in all traits, the slopes were close to 1.

Table 20. NIR Cross-Validation Statistics.

	Undigested Flour Protein	Digested Flour Protein	*In-Vitro* Protein Digestibility
Training Set			
No. of Samples	203	201	206
Mean	12.12	5.08	57.84
SD	1.46	1.15	1.64
SEC	1.17	0.85	6.29
r^2	0.37	0.44	0.26
SECV	1.23	0.89	6.44
1-VR	0.29	0.40	0.22
Cross-Validation Set			
No. of Samples	50	50	50
Slope	1.07	0.71	1.04
Intercept	-0.79	1.46	-0.59
Bias	0.09	0.03	1.63
SEC	1.06	0.85	5.64
SEP	1.04	0.86	5.76
SEP(C)	1.05	0.87	5.58
r^2	0.38	0.25	0.24

4.3.3 Basic Statistics of Predicted Data

The statistics of the NIR predicted dataset had a narrower distribution regarding the IVPD trait. The maximum NIR predicted IVPD value was 67 and the minimum just 49 (Table 21). However, the NIR predicted protein percent and digested protein percent ranges are similar to the wet chemistry distribution (Table 18). There also seems to be large differences in the distribution of IVPD whether the trait is predicted outright by the NIR (*In-Vitro* Protein Digestibility (NIR)) versus calculated in a formula based on the NIR predicted digested and undigested flour protein levels (*In-Vitro* Protein Digestibility (Calculated)) (Table 21). The two parents appear to have mean predictions different than performed under wet chemistry, especially the high digestibility parent P850029 (Table 18 and Table 21). The differences may not be due only to NIR error as environmental effect could also be influencing the values.

4.3.4 Analysis of Variance of NIR Predictions

Undigested protein percent, digested protein percent, and IVPD values all have significant genotype and genotype x environment effects (Table 22). Environment and rep (nested within the environment) are non-significant sources of variation for all traits. Environment was the largest total variance for the undigested flour protein percent trait but it was not significant ($P = 0.3678$). Genotype had the largest total variance of the remaining traits and while G x E was significant for all traits measured, the variance attributed to this trait was minimal. (Table 22).

Table 21. Basic Statistics of NIR Predicted Results

	Protein Pct.	Digested Protein Pct.	*In-Vitro* Protein Digestibility (NIR)	*In-Vitro*Protein Digestibility (Calculated)
Max	16.66	8.29	66.91	75.04
Median	12.08	5.19	57.29	56.81
Min	8.37	2.29	49.59	34.66
Mean	12.05	5.23	57.57	56.71
Std Dev	1.24	1.05	3.48	6.93
BTxArg-1 Mean	10.12	4.76	57.59	53.46
BTxArg-1 St Dev	1.42	1.12	0.74	4.30
P850029 Mean	12.35	4.34	62.07	64.91
P850029 St Dev	1.15	0.60	1.35	2.46

Table 22. Analysis of Variance for NIR Predicted Protein Traits.

Analysis of variance for three protein traits, including IVPD calculated from NIR data and IVPD predicted outright by the NIR, from phenotypic observations of F_4 plants grown in CS14, SC14, and CS15.

	Protein Percent		Digested Protein Percent		*In-vitro* Protein Digestibility (NIR)		*In-Vitro* Protein Digestibility (Calculated)	
Source of Variation	Pct. Variance	Wald p-value	Pct. Variance	Wald p-value	Pct. Variance	Wald p-value	Pct. Variance	Wald p-value
Environment	41.04	0.3678	36.67	0.3393	6.03	0.328	19.51	0.3203
Genotype	26.56	<.0001*	39.81	<.0001*	76.49	<.0001*	58.44	<.0001*
G x E	4.36	<.0001*	3.08	0.0002*	2.63	0.0001*	3.24	0.0002*
Rep[Env]	8.67	0.2243	3.28	0.2291	0.20	0.3315	0.16	0.3905
Residual	19.36		17.16		14.66		18.65	
Total	100.00		100.00		100.00		100.00	

Table 23. Highest Standard Deviations of IVPD%

Ten entries with the highest standard deviations for F_4 RILs tested in College Station in 2014. Data was generated via lab *in-vitro* pepsin digestibility assay (Mertz et al., 1984) by the USDA-ARS in Manhattan, KS.

RIL Number	In-Vitro Pct. Digestibility	Std. Dev of In-Vitro Pct. Digestibility
CS76-274	59.69	9.39
CS76-217	55.83	7.11
CS76-266	55.46	6.80
CS66-113	70.17	5.66
CS76-285	60.29	5.19
CS76-216	64.57	4.87
CS76-261	52.94	4.87
CS66-044	58.49	4.51
CS66-164	46.79	4.44
CS76-247	44.02	4.39

Table 24. Heritability Estimates of Protein Traits

Heritability estimates on a plot and entry-mean basis for three protein traits, including *in-vitro* protein digestibility (NIR predicted), for 289 recombinant inbred lines grown in three environments (TX14, TX15, SC14).

	Heritability	
	Plot Basis	Entry Mean
Protein Percent	0.50	0.85
Digested Protein Percent	0.65	0.91
In-Vitro Protein Digestibility	0.81	0.96

4.3.5 High Error Rate

There was a significant amount of variance associated with the pepsin assay as indicated by the standard deviation of the IVPD% trait (Table 23). The standard deviation was greater than 2.0 for 88 of the 289 lines tested indicating that the precision and consistency of this assay is low.

4.3.6 Heritability Estimates

The three protein traits were moderately heritable on a plot basis and highly heritable on an entry-mean basis (Table 24). *In-vitro* protein digestibility was estimated to be 96 percent heritable on the entry-mean basis. The traits were always more heritable calculated on an entry-mean basis compared to calculating heritability on a plot basis (Table 24).

4.3.7 Genetic Map

The genetic map for this population has been published previously by Boyles et al. (2016). The mapping population pairwise LD average fell below $r^2 = 0.2$ after 5.7 Mb. The extent of LD varied both within and across chromosomes which will lead to a variable mapping resolution that is dependent on QTL position. The average intermarker distance was ≤0.5 cM for all ten sorghum chromosomes. Segregation distortion, marker deviation from the expected 1:1 Mendelian ratio, was present at various genomic regions. All chromosomes except 1, 3, 8, and 10 contained distorted regions. Two of these distorted genomic locations were near known height loci, *Dw3* and *Dw1*, both of which were segregating in this population.

4.4 Discussion

Sorghum is a major cereal group across the world. Protein digestibility has significant considerations in sorghum's utilization for human nutrition, as a feedstock for animal agriculture, and grain-ethanol production. Based on the distribution of the IVPD trait in both wet chemistry and NIR predicted results, we can hypothesize this economic trait is quantitative. Previous crop improvement efforts approached this trait in a binary manner, categorizing genotypes either high or low. The reason for the previous dogma is likely due to poor phenotyping methodologies in which researchers have been unable to measure the trait with high precision or accuracy.

4.4.1 Narrow Genetic Base Evaluated

This study utilized a biparental RIL population using one high digestible parent, P850029, and TxArg-1 which was intermediate with regards to IVPD. For our NIR calibration curve, as well as the results of our experiment, to be broadly applicable a greater genetic diversity should be explored for the HD trait. Many lines known to possess the HD trait should be investigated to determine if there are genetic and perhaps compositionally different mechanisms to achieve the HD trait.

4.4.2 Genotype x Environment Effect

Previous studies reported an environmental effect for protein related traits. Although an environmental effect was not found for any of the protein traits measured, the genotype x environment effect was significant. This result indicates that protein digestibility phenotypes are not widely applicable across locations and perhaps across years.

4.4.3 Heritability

The IVPD trait was estimated as highly heritable in this experiment. Despite the presence of a G x E effect, a high heritability estimate suggests selection for this trait will be very effective. The trait may be selected in early breeding generations as opposed to the end of the breeding pipeline. Since the heritability is so high, the importance of replicated testing for selection is reduced.

4.4.4 Problems with the IVPD Assay

The best existing tool to analyze this trait in grain sorghum has been the *in-vitro* protein digestibility assay described by Mertz, et al. (1984), and it is the basis for the NIR training set in this project. Unfortunately, this assay is tedious and error-prone and this explains problems in the data. The project uses a tiny amount of ground flour for each sample. The nature of measuring small quantities (200 mg) may create inconsistencies from sample to sample. Also, 200 mg may be too small of a sample size to represent the whole of the variation of one RIL. Additionally, the mills may not grind every sample to uniform flour particles. Samples with large flour particles may appear less digestible than highly ground samples with smaller particles and each genotype might grind to different particle size due to genetic differences.

It is also difficult to keep other laboratory processes consistent during the experiment. The protocol requires the use of several buffers. The quantity of buffer required for a study of this size would necessitate the recreating and replenishing of the buffer multiple times. Even slight differences in the buffers' chemical makeup or pH could affect the results of the experiment. Other sources of error in the laboratory

include the water bath keeping a consistent temperature through every run of the entire experiment. The digested protein samples are dried in an oven after the water bath. The length of time left to air dry could potentially affect the quality of the protein remaining in the digested samples. The protocol might also be improved by utilizing a variety of pepsins instead of just one. Using an array of different enzymes to digest the protein in the flour could allow the study to more closely resemble the digestion biochemistry of humans, livestock, or commercial factories.

By far the biggest issue with the *in-vitro* protein digestibility is the time involved in phenotyping a project of this magnitude. At maximum productivity, one laboratory technician can run 12 to 18 samples in a single ten-hour day. If using replicates in the process, that is only six to nine genotypes assayed in one day. Therefore, weeks pass in-between the analysis of the first sample to the last. The process is too daunting to perform for an entire bi-parental population in multiple environments. It is too time-consuming to be used in a breeding program where selection for many important, economic traits are required.

4.4.5 The Need for a New Phenotyping System

From the results of this experiment, the wet chemistry IVPD data had a greater magnitude of means squares for the replicate source of variation than the genotype source of variation. The replicate effect was higher than the genotypic effect and this indicates the error of the lab results exceeds the natural differences between the genotypes in this population. These issues confirm the need for improved phenotyping methodologies for the high digestibility trait. The NIR curve is a first step because they

are faster and at least in this case as efficient as the laboratory assay. However, the NIR calibration curve is only as good as the data that it is created from. The high error of the wet chemistry information means there will be at minimum and equally high error of the NIR predicted information. At the time of writing, a new method of estimating digestibility is in development including the evaluation of protein body structure using scanning electron microscopy (Joseph Awika, personal communication). However, this protocol has some of the same limitation as the pepsin assay, namely the specialized training and significant time required to perform evaluations.

4.4.6 Future Research

Genetic mapping of this trait has not commenced as of writing. Future work should utilize the phenotypic information produced by the NIR predictions and the previously created genetic map to identify chromosomal regions associated with the high digestibility trait.

Addition genotypes should be added to the NIR calibration curve to strengthen its predictive ability across all germplasm in the Texas A&M AgriLife Research Sorghum Breeding Program. Many samples from this RIL population could be assayed to reduce error. Finally, the lines from this dataset should be investigated using the new electron scanning microscopy technique. The information from this innovative process could serve as an excellent check on the validity of the results from this study.

5. CONCLUSIONS

This dissertation has explored the various topics related to the improvement of production and grain quality in grain sorghum. The second section explored genetic gain within hybrid breeding programs and concluded that genetic yield gains are increasing about .008 t ha^{-1} $year^{-1}$ in public and private breeding programs and indicated gains of other physiological traits relating to yield. The third section explored the black pericarp trait and compared and contrasted gene expression of plant secondary metabolites of red and black pericarp lines. The fourth section explored the protein digestibility trait in sorghum, in which an NIR calibration curve was developed, and a strong genotype x environment effect was observed.

Future work remains to be completed on all three of these major projects. From the genetic gain study, research should now be done to determine how the emerging technologies (doubled haploids, genomic selection, high-throughput phenotyping, temporal male sterility, etc.) in sorghum breeding will impact genetic yield gains in the future. On-farm yield gains should also be explored in greater detail using existing datasets from the Texas A&M Variety Testing Program. In black pericarp study, genetic mapping of chromosomal reasons associated with the black pericarp trait should be performed using the existing phenotypic dataset to understand this complex trait further. In the protein study, the NIR predicted phenotypes should be mapped to determine which chromosomal regions are involved in the high-digestibility phenotype. Cereal chemists should also work to improve existing assays or create novel approaches to evaluate this important trait.

REFERENCES

Aboubacar, A., J.D. Axtell, C. Huang, and B.R. Hamaker. 2001. A Rapid Protein Digestibility Assay for Identifying Highly Digestible Sorghum Lines. Cereal Chem. 78(2): 160–165.

Anderson, J.W. 2003. Whole grains protect against atherosclerotic cardiovascular disease. Proc. Nutr. Soc. 62(1): 135–42.

Assefa, Y., and S.A. Staggenborg. 2011. Phenotypic Changes in Grain Sorghum Over the Last Five Decades. J. Agron. Crop Sci. 197(4): 249–257.

Awika, J. 2000. Sorghum phenols as antioxidants. Texsas A&M University.

Awika, J.M., and L.W. Rooney. 2004. Sorghum phytochemicals and their potential impact on human health. Phytochemistry 65(9): 1199–1221.

Awika, J., L. Rooney, and R. Waniska. 2004. Properties of 3-Deoxyanthocyanins from Sorghum. J. Agric. Food Chem. 52(14): 4388–4394.

Awika, J.M., L.W. Rooney, and R.D. Waniska. 2005. Anthocyanins from black sorghum and their antioxidant properties. Food Chem. 90(1–2): 293–301.

Awika, J.M., L.W. Rooney, X. Wu, R.L. Prior, and L. Cisneros-Zevallos. 2003. Screening methods to measure antioxidant activity of sorghum (sorghum bicolor) and sorghum products. J. Agric. Food Chem. 51(23): 6657–62.

Ayyangar, G.N.R., C. Vijiaraghavan, M.A.S. Ayyar, and V.P. Rao. 1933. Inheritance of characters in sorghum. III. Grain colours red, yellow, white. Indian J Agric Sci 3: 594–604.

Ayyangar, G.N.R., C. Vijiaraghavan, M.A.S. Ayyar, and V.P. Rao. 1934. Inheritance of

characters in sorghum—the great millet. VI. Pearly and chalky grains. Indian J. Agric. Sci 4: 96–99.

Bach Knudsen, K.E., and L. Munck. 1985. Dietary fibre contents and compositions of sorghum and sorghum-based foods. J. Cereal Sci. 3(2): 153–164.

Bansal, S., A. Mishra, A. Tomar, S. Sharma, V.K. Khanna, and G.K. Garg. 2008. Isolation and temporal endospermal expression of γ-kafirin gene of grain sorghum (Sorghum bicolor L. moench) var. M 35-1 for introgression analysis of transgene. J. Cereal Sci. 48(3): 808–815.

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 Chem. 83(1): 108–113.

Bean, S.R., B.P. Ioerger, B.M. Smith, and D.L. Blackwell. 2011. Sorghum protein structure and chemistry: Implications for nutrition and functionality. ACS Symp. Ser. 1089: 131–147.

Belton, P.S., I. Delgadillo, N.G. Halford, and P.R. Shewry. 2006. Kafirin structure and functionality. J. Cereal Sci. 44(3): 272–286.

Bennett, W.F., B.B. Tucker, and A.B. Maunder. 1990. Modern grain sorghum production. Iowa State University Press, Ames, Iowa, USA.

Beta, T., and H. Corke. 2001. Genetic and Environmental Variation in Sorghum Starch Properties. J. Cereal Sci. 34: 261–268.

Boddu, J., C. Jiang, V. Sangar, T. Olson, T. Peterson, and S. Chopra. 2006. Comparative structural and functional characterization of sorghum and maize duplications

containing orthologous Myb transcription regulators of 3-deoxyflavonoid biosynthesis. Plant Mol. Biol. 60(2): 185–199.

Boddu, J., C. Svabek, F. Ibraheem, A.D. Jones, and S. Chopra. 2005. Characterization of a deletion allele of a sorghum Myb gene yellow seed1 showing loss of 3-deoxyflavonoids. Plant Sci. 169(3): 542–552.

Boyles, R.E., B.K. Pfeiffer, E.A. Cooper, B.L. Rauh, K.J. Zielinski, M.T. Myers, Z. Brenton, W.L. Rooney, and S. Kresovich. 2016. Genetic dissection of sorghum grain quality traits using diverse and segregating populations. Theor. Appl. Genet. 130(4): 697-716.

Bubeck, D.M., M.R. Carlone, R.L. Fox, M.D. Hoffbeck, R.L. Segebart, and D.S. Stucker. 2006. Breeding progress measured in eight elite inbred families. Maydica 51(1): 141–149.

Butler, L.G., D.J. Riedl, D.G. Lebryk, and H.J. Blytt. 1984. Interaction of proteins with sorghum tannin: Mechanism, specificity and significance. J. Am. Oil Chem. Soc. 61(5): 916–920.

Campbell, B.T., D. Boykin, Z. Abdo, W.R. Meredith, S. Smith, B. Diers, J. Specht, and B. Carver. 2014. Cotton. p. 13–32. *In* Smith, S., Diers, B., Specht, J., Carver, B. (eds.), Yield Gains in Major U.S. Field Crops. 33rd ed. ASA, CSSA, and SSSA, Madison, Wisconsin, USA.

Campbell, B.T., P.W. Chee, E. Lubbers, D.T. Bowman, J.R. Meredith, J. Johnson, and D.E. Fraser. 2011. Genetic improvement of the Pee Dee cotton germplasm collection following seventy years of plant breeding. Crop Sci. 51(3): 955–968.

Casady, A.J. 1965. Effect of a Single Height (Dw3) Gene of Sorghum on Grain Yield, Grain Yield Components, and Test Weight. Crop Sci. 5(5): 385–388.

Chamba, E.B., N.G. Halford, J. Forsyth, M. Wilkinson, and P.R. Shewry. 2005. Molecular cloning of β-kafirin, a methionine-rich protein of sorghum grain. J. Cereal Sci. 41: 381–383.

Chen, F., P. Cole, Z. Mi, and L. Xing. 1993. Corn and wheat-flour consumption and mortality from esophageal cancer in shanxi, China. Int. J. Cancer 53(6): 902–906A.

Clifford, M.N. 2000. Anthocyanins – nature, occurrence and dietary burden. J. Sci. Food Agric. 80: 1063–1072.

Dalton, L.G. 1969. A postiive regression of yield on maturity in sorghum. Crop Sci. 7(3): 271.

Doggett, H. 1988. Sorghum. 2nd ed. Longman Scientific & Technical, London.

Duncan, R.R., P.J. Bramel-Cox, and F.R. Miller. 1991. Contributions of introduced germplasm to hybrid development in the USA. p. 69–102. *In* Shand, M., Wiesner, L. (eds.), Use of plant introductions in cultivar development. Crop Science Society of America, Madison, Wisconsin, USA.

Duodu, K.G., A. Nunes, I. Delgadillo, M.L. Parker, E.N.C. Mills, P.S. Belton, and J.R.N. Taylor. 2002. Effect of Grain Structure and Cooking on Sorghum and Maize in vitro Protein Digestibility. J. Cereal Sci. 35(2): 161–174.

Duodu, K., J.R.. Taylor, P. Belton, and B. Hamaker. 2003. Factors affecting sorghum protein digestibility. J. Cereal Sci. 38(2): 117–131.

Duvick, D.N. 1984. Genetic Contributions to Yield Gains of U.S. Hybrid Maize, 1930 to

1980. Genet. Contrib. to Yield Gains Five Major Crop Plants: 15–47.

Duvick, D.N. 1997. What is yield? p. 332–335. *In* Edmeades, G.O., Banzinger, B., Mickelson, H.R., Pena-Valdivia, C.B. (eds.), Developing Drought and Low N-Tolerant Maize. CIMMYT, El Batan, Mexico.

Duvick, D.N. 2005a. The Contribution of Breeding to Yield Advances in maize (Zea mays L.). p. 83–145. *In* Donald, L.S. (ed.), Advances in Agronomy. Academic Press.

Duvick, D.N. 2005b. Genetic progress in yield of Untied States maize (Zea mays L.). 50: 193–202.

Duvick, D.N., J. Smith, and M. Cooper. 2004. Long-term Selection in a Commercial Hybrid Maize Breeding Program. p. 109–151. *In* Janick, J. (ed.), Plant Breeding Reviews Volume 24; Part 2: Long-term Selecton: Crops, Animals, and Bacteria. John Wiley & Sons, Inc., New Jersey.

Dykes, L., G.C. Peterson, W.L. Rooney, and L.W. Rooney. 2011. Flavonoid composition of lemon-yellow sorghum genotypes. Food Chem. 128(1): 173–179.

Dykes, L., and L.W. Rooney. 2006. Sorghum and millet phenols and antioxidants. J. Cereal Sci. 44(3): 236–251.

Dykes, L., W.L. Rooney, and L.W. Rooney. 2013. Evaluation of phenolics and antioxidant activity of black sorghum hybrids. J. Cereal Sci. 58: 278–283.

Dykes, L., L.M. Seitz, W.L. Rooney, and L.W. Rooney. 2009. Flavonoid composition of red sorghum genotypes. Food Chem. 116(1): 313–317.

Eberhart, S.A., and W.A. Russell. 1966. Stability Parameters for Comparing Varieties.

Crop Sci. 6(1): 36.

Ejeta, G., M.M. Hassen, and E.T. Mertz. 1987. In vitro digestibility and amino acid composition of pearl millet (Pennisetum typhoides) and other cereals. Proc. Natl. Acad. Sci. U. S. A. 84(17): 6016–6019.

Elkin, R.G., M.B. Freed, B.R. Hamaker, Y. Zhang, and C.M. Parsons. 1996. Condensed tannins are only partially responsible for variations in nutrient digestibilities of sorghum grain cultivars. J. Agric. Food Chem. 44(3): 848–853.

Fehr, W. 1987. Principles of cultivar development. Macmillan, New York.

Ferreyra, M.L.F., S. Rius, J. Emiliani, L. Pourcel, A. Feller, K. Morohashi, P. Casati, and E. Grotewold. 2010. Cloning and characterization of a UV-B-inducible maize flavonol synthase. Plant J. 62(1): 77–91.

Finlay, K.W., and G.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Aust. J. Agric. Res. 14(6): 742–754.

Gabriel, K. 2006. A study of heterotic relationships in Sorghum.

Gizzi, G., and B.L. Gambin. 2016. Eco-physiological changes in sorghum hybrids released in Argentina over the last 30 years. F. Crop. Res. 188: 41–49.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2).

Gonzalez, A., J. Mendenhall, Y. Huo, and A. Lloyd. 2009. TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Dev. Biol. 325(2): 412–421.

Gous, F. 1989. Tannins and phenols in black sorghum. Texas A&M University

Graham, R.J.D. 1916. Pollination and Cross-fertilization in the Juar Plant (Andropogon Sorghum, Brot). Imperial Department of Agriculture in India.

Graham, D., and K.J. Lessman. 1966. Effect of Height on Yield and Yield Components of Two Isogenic Lines of Sorghum vulgare Pers. Crop Sci. 6(4): 372–374.

Guiragossian, V., B. a. K. Chibber, S. Van Scoyoc, R. Jambunathan, E.T. Mertz, and J.D. Axtell. 1978. Characteristics of proteins from normal, high lysine, and high tannin sorghums. J. Agric. Food Chem. 26(1): 219–223.

Haegele, J.W., K.A. Cook, D.M. Nichols, and F.E. Below. 2013. Changes in nitrogen use traits associated with genetic improvement for grain yield of maize hybrids released in different decades. Crop Sci. 53(4): 1256–1268.

Hamaker, B.R., and B.A. Bugusu. 2003. Overview: sorghum proteins and food quality. *In* Workshop on the proteins of sorghum and millets: enhancing nutritional and functional properties for Africa. West Lafayette, Indiana, USA.

Henley, E.C., J.R.N. Taylor, and S.D. Obukosia. 2010. The importance of dietary protein in human health: Combating protein deficiency in sub-Saharan Africa through transgenic biofortified sorghum. 1st ed. Elsevier Inc.

Heppel, S.C., F.W. Jaffé, A.M. Takos, S. Schellmann, T. Rausch, A.R. Walker, and J. Bogs. 2013. Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. Plant Mol. Biol. 82(4–5): 457–471.

Ibraheem, F., I. Gaffoor, Q. Tan, C.R. Shyu, and S. Chopra. 2015. A sorghum MYB

transcription factor induces 3-deoxyanthocyanidins and enhances resistance against leaf blights in maize. Molecules 20(2): 2388–2404.

Ioerger, B., S.R. Bean, M.R. Tuinstra, J.F. Pedersen, J. Erpelding, K.M. Lee, and T.J. Herrman. 2007. Characterization of polymeric proteins from vitreous and floury sorghum endosperm. J. Agric. Food Chem. 55(25): 10232–10239.

Izquierdo, L., and I.D. Godwin. 2005. Molecular Characterization of a Novel Methionine-Rich δ -Kafirin Seed Storage Protein Gene in Sorghum (Sorghum bicolor L .). 30(4): 313–318.

Jambunathan, R., E.T. Mertz, and J.D. Axtell. 1975. Fractionation of soluble proteins of high-lysine and normal sorghum grain. Cereal Chem. 52(1): 119–121.

Jampala, B., W.L. Rooney, G.C. Peterson, S. Bean, and D.B. Hays. 2012. Estimating the relative effects of the endosperm traits of waxy and high protein digestibility on yield in grain sorghum. F. Crop. Res. 139: 57–62.

JMP® Pro, Version 13.1.0. SAS Institute Inc., Cary, NC, 12016.

Johnson, J.W., D.T. Rosenow, G.L. Teetes, and J.M. Phillips. 1982. Registration of 19 greenbug resistant sorghum germplasm lines. 22: 1272.

Jun, J.H., C. Liu, X. Xiao, and R.A. Dixon. 2015. The Transcriptional Repressor MYB2 Regulates Both Spatial and Temporal Patterns of Proanthocyandin and Anthocyanin Pigmentation in Medicago truncatula. Plant Cell 27(10): 2860–79.

Kawahigashi, H., S. Kasuga, Y. Sawada, J. -i. Yonemaru, T. Ando, H. Kanamori, J. Wu, H. Mizuno, M. Momma, Z. Fujimoto, M.Y. Hirai, and T. Matsumoto. 2016. The Sorghum Gene for Leaf Color Changes upon Wounding (P) Encodes a Flavanone

4-Reductase in the 3-Deoxyanthocyanidin Biosynthesis Pathway. G3 Genes, Genomes, Genet. 6(5): 1439–1447.

Knuckles, B.E., D.D. Kuzmicky, and A.A. Betschart. 1985. Effect of Phytate and Partially Hydrolyzed Phytate on in vitro Protein Digestibility. J. Food Sci. 50(4): 1080–1082.

Kushi, L.H., K.A. Meyer, and D.R. Jacobs. 1999. Cereals, legumes, and chronic disease risk reduction: evidence from epidemiologic studies. Am. J. Clin. Nutr. 70(3): 451s–458s.

Laidlaw, H.K.C., E.S. Mace, S.B. Williams, K. Sakrewski, A.M. Mudge, P.J. Prentis, D.R. Jordan, and I.D. Godwin. 2010. Allelic variation of the β-, γ- and δ-kafirin genes in diverse Sorghum genotypes. Theor. Appl. Genet. 121: 1227–1237.

Landry, J., and T. Moureaux. 1970. Heterogeneite des glutelines du grain de mais: extraction selective et composition en acides amines des trois fractions isolees. Soc Chim Biol Bull.

Laubscher, F.X. 1945. A genetic study of Sorghum relationships. Union S. Afr. Sci. Bull. 242: 1–22.

Lemoine, R. 2000. Sucrose transporters in plants: Update on function and structure. Biochim. Biophys. Acta - Biomembr. 1465(1–2): 246–262.

Lin, C.-S., M.R. Binns, and L.P. Lefkovitch. 1986. Stability analysis: where do we stand? Crop Sci. 26(5): 894–900.

Liu, S.M., G.A. Constable, P.E. Reid, W.N. Stiller, and B.R. Cullis. 2013a. The interaction between breeding and crop management in improved cotton yield. F.

Crop. Res. 148: 49–60.

Liu, H., Y. Du, H. Chu, C.H. Shih, Y.W. Wong, M. Wang, I.K. Chu, Y. Tao, and C. Lo. 2010. Molecular dissection of the pathogen-inducible 3-deoxyanthocyanidin biosynthesis pathway in sorghum. Plant Cell Physiol. 51(7): 1173–1185.

Liu, Z., Y. Liu, Z. Pu, J. Wang, Y. Zheng, Y. Li, and Y. Wei. 2013b. Regulation, evolution, and functionality of flavonoids in cereal crops. Biotechnol. Lett. 35(11): 1765–1780.

Lo, S.C., and R.L. Nicholson. 1998. Reduction of light-induced anthocyanin accumulation in inoculated sorghum mesocotyls. Implications for a compensatory role in the defense response. Plant Physiol. 116(3): 979–989.

Mace, E.S., H.K. Buhariwalla, and J.H. Crouch. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459a-459h

Mazhar, H., and A. Chandrashekar. 1993. Differences in kafirin composition during endosperm development and germination in sorghum cultivars of varying hardness. Cereal Chem. 70(6): 667–671.

Mazza, G., and R. Brouillard. 1987. Color stability and structural transformations of cyanidin 3, 5-diglucoside and four 3-deoxyanthocyanins in aqueous solutions. J. Agric. Food Chem. 35(3): 422–426.

Menz, M.A., R.R. Klein, N.C. Unruh, W.L. Rooney, P.E. Klein, and J.E. Mullet. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci. 44: 1236–1244.

Mertz, E.T., M.M. Hassen, C. Cairns-Whittern, a W. Kirleis, L. Tu, and J.D. Axtell. 1984. Pepsin digestibility of proteins in sorghum and other major cereals. Proc. Natl. Acad. Sci. U. S. A. 81(January): 1–2.

Miller, F.R. 1984. Regustration of RTx430 sorghum parental line. Crop Sci. 24(December): 1224.

Miller, F.R. 1986. Registration of RTx435 sorghum parental line. Crop Sci. 26: 215–216.

Miller, F.R. 1992. Registration of A/BTx635 Sorghum. Crop Sci. 32(December): 1517–1518.

Miller, F.R., and Y. Kebede. 1984. Genetic Contributions to Yield Gains in Sorghum, 1950 to 1980. p. 1–14. *In* Fehr, W.R. (ed.), Genetic Contributions to Yield Gains of Five Major Crop Plants. CSSA, Madison, Wisconsin, USA.

Mizuno, H., H. Kawahigashi, Y. Kawahara, H. Kanamori, J. Ogata, H. Minami, T. Itoh, and T. Matsumoto. 2012. Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-Bipolaris sorghicola interaction. BMC Plant Biol. 12(1): 121.

Mizuno, H., T. Yazawa, S. Kasuga, Y. Sawada, H. Kanamori, Y. Ogo, M.Y. Hirai, T. Matsumoto, and H. Kawahigashi. 2016. Expression of Flavone Synthase II and Flavonoid 3′-Hydroxylase Is Associated with Color Variation in Tan-Colored Injured Leaves of Sorghum. Front. Plant Sci. 7(November): 1–10.

Monk, R., C. Franks, and J. Dahlberg. 2014. Sorghum. p. 293–310. *In* Smith, S., Diers, B., Specht, J., Carver, B. (eds.), Yield Gains in Major U.S. Field Crops. 33rd ed.

Madison, Wisconsin, USA.

Myers, R.J.K., and M.A. Foale. 1981. Row spacing and population density in grain sorghum — A simple analysis. F. Crop. Res. 4: 147–154.

Oria, M.P., B.R. Hamaker, J.D. Axtell, and C.P. Huang. 2000. A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc. Natl. Acad. Sci. U. S. A. 97(10): 5065–5070.

Oria, M.P., B.R. Hamaker, and J.M. Schull. 1995a. In vitro protein digestibility of developing and mature sorghum grain in relation to α-, β-, and γ-kafirin disulfide crosslinking. J. Cereal Sci. 22(1): 85–93.

Oria, M.P., B.R. Hamaker, and J.M. Shull. 1995b. Resistance of Sorghum α-, β-, and γ-Kafirins to Pepsin Digestion. J. Agric. Food Chem. 43(317): 2148–2153.

Peterson, G.. C., J.W. Jognson, G.L. Teetes, D.T. Rosenow, and R.E. Schaffert. 1984. Registration of Tx2783 greenbug resistant sorghum germplasm line. Crop Sci. 24: 1984.

Peterson, G.C., K. Schaefer, and B.B. Pendleton. 2009. Registration of 16 Sorghum Germplasm Lines. 3(2): 203–205.

Petti, C., R. Kushwaha, M. Tateno, A.E. Harman-Ware, M. Crocker, J. Awika, and S. DeBolt. 2014. Mutagenesis breeding for increased 3-deoxyanthocyanidin accumulation in leaves of sorghum bicolor (L.) moench: A source of natural food pigment. J. Agric. Food Chem. 62(6): 1227–1232.

Pfeiffer, B.K., and W.L. Rooney. 2015. Sunlight Induces Black Color and Increases Flavonoid Levels in the Grain of Sorghum Line Tx3362. Crop Sci. 55(4): 1703.

Pfeiffer, B.K., and W.L. Rooney. 2016. Inheritance of pericarp color, nutritional quality, and grain composition traits in black sorghum. Crop Sci. 56(1): 164–172.

Poloni, A., and J. Schirawski. 2014. Red card for pathogens: Phytoalexins in sorghum and maize. Molecules 19(7): 9114–9133.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rooney, W.L. 2000. Genetics and Cytogenetics. p. 261–308. *In* Smith, C.W., Frederiksen, R.A. (eds.), Sorghum: origin, history, technology, production. Wiley, New York.

Rooney, W.L. 2004. Sorghum Improvement - Integrating Traditional and New Technology to Produce Improved Genotypes. Adv. Agron. 83: 37–109.

Rooney, W.L., F.R. Miller, and L.W. Rooney. 2003. Registration of RTx437 sorghum parental line. Crop Sci. 43(1): 445–446.

Rooney, W.L., O. Portillo, and C. Hayes. 2013a. Registration of ATx3363 and BTx3363 Black Sorghum Germplasms. J. Plant Regist. 7(3): 342–346.

Rooney, W.L., L.W. Rooney, J. Awika, and L. Dykes. 2013b. Registration of Tx3362 Sorghum Germplasm. J. Plant Reg. 7: 104–107.

Rosenow, D., L.C. J, A. Dahlberg, R. Fredericksen, G. Odvody, G. Peterson, F. Miller, C. Woodfin, K. Schaefer, S. Collins, J. Jones, and A. Hamburger. 2002. Release of four A/B sorghum parental lines ATx642 through ATx645. p. 24–30. *In* International Sroghum and Millets Newsletter.

Russel, W.A. 1991. Genetic improvement of maize yields. Adv. Agron. 46: 245–298.

Sarath, G., R.B. Mitchell, S.E. Sattler, D. Funnell, J.F. Pedersen, R.A. Graybosch, and K.P. Vogel. 2008. Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. J. Ind. Microbiol. Biotechnol. 35(5): 343–354.

Schwartz, B.M., and C.W. Smith. 2008. Genetic gain in fiber properties of upland cotton under varying plant densities. Crop Sci. 48(4): 1321–1327.

Serna-Saldivar, S., and L.W. Rooney. 1995. Structure and chemistry of sorghum and millets. p. 69–124. *In* Sorghum and millets: chemistry and technology. American Association of Cereal Chemists: St. Paul, MN.

Sharma, M., C. Chai, K. Morohashi, E. Grotewold, M.E. Snook, and S. Chopra. 2012. Expression of flavonoid 3'-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. BMC Plant Biol. 12(1): 196.

Shih, C.H., I.K. Chu, W.K. Yip, and C. Lo. 2006. Differential expression of two flavonoid 3′-hydroxylase cDNAs involved in biosynthesis of anthocyanin pigments and 3-deoxyanthocyanidin phytoalexins in sorghum. Plant Cell Physiol. 47(10): 1412–1419.

Shih, C.-H.H., S.-O.O. Siu, R. Ng, E. Wong, L.C.M. Chiu, I.K. Chu, and C. Lo. 2007. Quantitative analysis of anticancer 3-deoxyanthocyanidins in infected sorghum seedlings. J. Agric. Food Chem. 55(2): 254–259.

Shull, J., A. Chandrashekar, A. Kirleis, and G. Ejeta. 1990. Development of Sorghum (Sorghum bicolor) endosperm in varieties of varying hardness. Food Struct. 9(3): 253–267.

Shull, J.M., J.J. Watterson, and A.W.I. Kirleis. 1992. Purification and immunocytochemical localization of kafirins in Sorghum bicolor (L . Moench) endosperm. Protoplasma 171: 64–74.

Singh, R., and J.D. Axtell. 1973. High Lysine Mutant Gene (hl that Improves Protein Quality and Biological Value of Grain Sorghum. Crop Sci. 13(5): 535–539.

Smith, S., M. Cooper, J. Gogerty, C. Löffler, D. Borcherding, K. Wright, S. Smith, B. Diers, J. Specht, and B. Carver. 2014. Maize. p. 125–172. *In* Smith, S., Diers, B., Specht, J., Carver, B. (eds.), Yield Gains in Major U.S. Field Crops. 33rd ed. ASA, CSSA, and SSSA, Madison, Wisconsin, USA.

Smith, J.S.C., D.N. Duvick, O.S. Smith, M. Cooper, and L. Feng. 2004. Changes in pedigree backgrounds of pioneer brand maize hybrids widely grown from 1930 to 1999. Crop Sci. 44(6): 1935–1946.

Specht, J.E., B.W. Diers, R.L. Nelson, J. Francisco, F. de Toledo, J.A. Torrion, P. Grassini, S. Smith, B. Diers, J. Specht, and B. Carver. 2014. Soybean. p. 311–356. *In* Smith, S., Diers, B., Specht, J., Carver, B. (eds.), ield Gains in Major U.S. Field Crops. 33rd ed. ASA, CSSA, and SSSA, Madison, Wisconsin, USA.

Stephens, J.C. 1946. A Second Factor for Subcoat in Sorghum Seed. Agron. J. 38(4): 340–342.

Stephens, J.C., and R.E. Karper. 1965. Release of breeding stocks of male-sterilized grain sorghum lines. College Station, TX.

Stich, K., and G. Forkmann. 1988. Biosynthesis of 3-deoxyanthocyanins with flower extracts from Sinningia cardinalis. Phytochemistry 27(3): 785–789.

Swarts, K., H. Li, J.A. Romero Navarro, D. An, M.C. Romay, S. Hearne, C. Acharya, J.C. Glaubitz, S. Mitchell, R.J. Elshire, E.S. Buckler, and P.J. Bradbury. 2014. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome 7(3): 0.

Sweeny, J.G., and G.A. Iacobucci. 1981. Synthesis of anthocyanidins-III: Total synthesis of apigeninidin and luteolinidin chlorides. Tetrahedron 37(8): 1481–1483.

Tesso, T., G. Ejeta, A. Chandrashekar, C.P. Huang, A. Tandjung, M. Lewamy, J.D. Axtell, and B.R. Hamaker. 2006. A novel modified endosperm texture in a mutant high-protein digestibility/high-lysine grain sorghum (Sorghum bicolor (L.) moench). Cereal Chem. 83(2): 194–201.

Tesso, T., B. Hamaker, and G. Ejeta. 2008. Sorghum protein digestibility is affected by dosage of mutant alleles in endosperm cells. Plant Breed. 127(6): 579–586.

Tohge, T., L.P. De Souza, and A.R. Fernie. 2017. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. J. Exp. Bot. 68(15): 4013–4028.

Tsuda, T., F. Horio, K. Uchida, H. Aoki, and T. Osawa. 2003. Dietary cyanidin 3-O-beta-D-glucoside-rich purple corn color prevents obesity and ameliorates hyperglycemia in mice. J. Nutr. 133(7): 2125–2130.

Unger, P.W., and R.L. Baumhardt. 1999. Factors related to dryland grain sorghum yield increases: 1939 through 1997. Agron. J. 91: 870–875.

USDA-NASS. 2017. Statistics by subject. USDA-National Agricultural Statistics Service. https://quickstats.nass.usda.gov/.

Vinall, H.N., and A.B. Cron. 1921. Improvement of sorghums by hybridization. J.

Hered. 12(10): 435–443.

Watterson, J.J., J.M. Shull, and A.W. Kirleis. 1993. Quantitation of α-, β-, and γ-kafirins in vitreous and opaque endosperm of Sorghum bicolor. Cereal Chem. 70(4): 452–457.

Weaver, B. a a, and D.W. Cleveland. 2006. Does aneuploidy cause cancer? Curr. Opin. Cell Biol. 18(6): 658–67.

Weaver, C. a., B.R. Hamaker, and J.D. Axtell. 1998. Discovery of grain sorghum germ plasm with high uncooked and cooked in vitro protein digestibilities. Cereal Chem. 75(5): 665–670.

Winn, J.A., R.E. Mason, A.L. Robbins, W.L. Rooney, and D.B. Hays. 2009. QTL mapping of a high protein digestibility trait in Sorghum bicolor. Int. J. Plant Genomics 2009: 471853.

Wolfrum, E., C. Payne, T. Stefaniak, and W. Rooney. 2013. Multivariate Calibration Models for Sorghum Composition using Near-Infrared Spectroscopy Multivariate Calibration Models for Sorghum Composition using Near-Infrared Spectroscopy.

Wong, J.H., T. Lau, N. Cai, J. Singh, J.F. Pedersen, W.H. Vensel, W.J. Hurkman, J.D. Wilson, P.G. Lemaux, and B.B. Buchanan. 2009. Digestibility of protein and starch from sorghum (Sorghum bicolor) is linked to biochemical and structural features of grain endosperm. J. Cereal Sci. 49(1): 73–82.

Wu, X., B. Jampala, A. Robbins, D. Hays, S. Yan, F. Xu, W. Rooney, G. Peterson, Y.-C. Shi, and D. Wang. 2010. Ethanol fermentation performance of grain sorghums (Sorghum bicolor) with modified endosperm matrices. J. Agric. Food Chem.

58(17): 9556–62.

Wu, X., and R.L. Prior. 2005. Identification and characterization of anthocyanins by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry in common foods in the United States: vegetables, nuts, and grains. J Agric Food Chem 53(8): 3101–3113.

Young, V.R., N.S. Scrimshaw, and P.L. Pellett. 1998. Significance of dietary protein source in human nutrition: Animal or plant proteins? p. 205–222. *In* JC Waterlow, DG Armstrong, L.F. and R.R. (ed.), Feeding a World Population of More Than Eight Billion People.