ENVIRONMENTAL AND GENETIC CONTROL OF BLACK PERICARP TRAITS IN SORGHUM (SORGHUM BICOLOR)

A Dissertation

by

LAUREN NANCY FEDENIA

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,	Patricia E. Klein
Committee Members,	Robert R. Klein
	William L. Rooney
	Luis Cisneros-Zevallos
	Linda Dykes
Head of Department,	Patricia E. Klein

May 2021

Major Subject: Horticulture

Copyright 2021 Lauren Nancy Fedenia

ABSTRACT

Sorghum [*Sorghum bicolor* (L.) Moench] is a major cereal crop grown for feed, silage, and human consumption. Specialty grain-types are typically red, yellow, or white and this coloration is associated with specific polyphenol profiles that serve as an important dietary source of diverse plant polyphenols. Unique amongst the common grain sorghums are the black grain genotypes that are all descendants of the accession 'Black Shawaya' from the Sudan. The black pigmentation of 'Black Shawaya' grain is associated with the accumulation of rare polyphenols, 3-deoxyanthocyanidins (3-DOAs), however, the expression of this trait is not observed in all environments conducive to grain sorghum production. The present research aimed to elucidate the environmental factors required for full penetrance of this trait and the distinct genetic pathways that condition the atypical accumulation of 3-DOA in the black sorghum pericarp.

Light spectrum and/or photoperiod were previously implicated as critical factors for the expression of this trait. In this research, the black sorghum genotype was grown under regimes of visible light, visible light supplemented with UVA or supplemented with UVA plus UVB (or dark control). Pericarp 3-DOAs and pericarp pigmentation were maximized in the black genotype exposed to a light regime supplemented with UVB revealing that a fluence of UVB is the determining environmental component for trait penetrance.

Multiple trait mapping methodologies were used to characterize the genomic loci associated with the black sorghum phenotype including genetic linkage map

construction with quantitative trait loci (QTL) analysis using an F_5 recombinant inbred line population segregating for pericarp 3-DOA accumulation and bulked segregant analysis (BSA) of opposing phenotypic extremes for 3-DOA accumulation. Four major overlapping QTL were detected in multiple mapping efforts along with several minor-effect QTL discovered in separate mapping methods. These mapping efforts revealed genomic regions of the sorghum genome controlling this trait and provided the foundation for additional elucidation of genes regulating the cascade of cellular events leading to the black pericarp phenotype.

Finally, a detailed transcriptome and co-expression analysis performed on pericarp, leaf, and root tissue from black- and red-grain sorghum genotypes under different light regimes revealed a potential metabolic and signaling pathway modeling 3-DOA induction in black sorghum pericarp tissue in which UVB triggers the generation of endogenous elicitors that initiate a signal cascade resulting in the high expression of flavonoid biosynthetic genes ultimately responsible for 3-DOA synthesis.

DEDICATION

This dissertation is dedicated to Nancy Brace and Daniel Aultman; two individuals who helped me discover my passion and purpose in life.

ACKNOWLEDGEMENTS

I thank my committee chair and supervisor, Dr. Patricia Klein, for the opportunity to conduct research in her program and for her enduring support and commitment to my education and development. Her mentorship was indispensable for the completion of this work and she has unknowingly served as my utmost role-model throughout my graduate career. For me, she represents all that is possible to achieve as a woman in science through strength and persistence. I would also like to thank Dr. Robert Klein for his continual guidance and dedication to my training. Without his years of incalculable support, this doctoral study would not be possible. His commitment to the development of his students cannot be overstated. I am forever grateful for his expertise and participation in my education.

I would like to express my appreciation to Dr. William Rooney, Dr. Linda Dykes, and Dr. Luis Cisneros-Zevallos for agreeing to serve on my committee. Their collaboration was essential for the completion of these studies. I would also like to thank my other team members including Julie McCollum, Natalie Patterson, and my fellow lab mates for all their contributions to this research. It was a privilege to have their teamwork, support, and friendship.

Finally, I would like to thank the College of Agriculture and Life Sciences, the Department of Horticultural Sciences, the Belsterling Foundation, and all of the organizations who contributed funding and scholarships to support my education.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Patricia Klein of the Department of Horticultural Sciences at Texas A&M University, Dr. Robert Klein of the USDA-ARS, Dr. William Rooney of the Department of Soil and Crop Sciences at Texas A&M University, Dr. Linda Dykes of the USDA-ARS, and Dr. Luis Cisneros-Zevallos of the Department of Horticultural Sciences at Texas A&M University.

This project utilized germplasm developed by Dr. William Rooney and Dr. Brian Pfeiffer. Chamber experiments were designed by Dr. Robert Klein. Julie McCollum and Carson Brandenburger assisted in phenotyping and collection and/or isolation of DNA for some of the genotypes characterized. Illumina sequencing was provided by the Texas A&M AgriLife Research Genomics and Bioinformatics Services. Dr. Patricia Klein analyzed and produced SNP datasets from sequencing data. All foundational expertise, equipment, reagents, and laboratory space for flavonoid quantification in black sorghum pericarp was provided by Dr. Linda Dykes. Dr. Linda Dykes conducted all procedures using Ultra High Performance Liquid Chromatography and established the correlation to the colorimetric assay. Additionally, Dr. Linda Dykes hosted Lauren Fedenia for multiple weeks to train her in her USDA laboratory on procedures which served as the basis to all flavonoid phenotyping conducted in this research. Finally, the co-expression analyses depicted in Chapter 4 were conducted with the guidance of Dr. Sabyasachi Mandal of the Department of Biology.

All other work conducted for this dissertation was completed independently by Lauren Fedenia.

Funding Sources

The research described in this dissertation was supported by my graduate research assistantship from the Department of Horticultural Sciences at Texas A&M and by the USDA-ARS, Texas A&M University Department of Horticultural Sciences and Department of Soil and Crop Sciences, Texas A&M AgriLife Research, USDA-ARS Non Assistance Cooperative Agreement 58−3091−9-020, and USDA Hatch funds.

The contents of this work are solely the responsibility of the author and do not necessarily represent the official views of the funding sources.

NOMENCLATURE

3-DOA	3-deoxyanthocyanidins
4CL	4-coumarate:CoA ligase
ANS	Anthocyanidin synthase
AOC	Allene oxide cyclase
AOX	Alternative oxidase
APX-C	L-ascorbate peroxidase (cytosolic)
APX-P	L-ascorbate peroxidase (peroxisomal)
BLUP	Best linear unbiased prediction
BSA	Bulk-segregant analysis
C4H	Trans-cinnamate 4-monooxygenase
CaM binding	Calmodulin-binding protein
CHI	Chalcone isomerase
CHS	Chalcone synthase
CS18	College Station 2018
CS19	College Station 2019
CYP P450	Cytochrome P450
DFR	Dihydroflavonol 4-reductase
ELIP	Early-light inducible protein
F3H	Flavanone 3-hydroxylase
F3'H	Flavonoid 3'-hydroxylase

FNR	Flavanone 4-reductase
FNSII	Flavone synthase II
GBS	Genotyping-by-sequencing
GPx	Glutathione peroxidase
GST	Glutathione S-transferase
LOX	Lipoxygenase
PAL	Phenylalanine ammonia lyase
PVE	Percent variation explained
PO	Polyphenol oxidase
PR-10a	Pathogenesis-related protein 10a
PR-5	Pathogenesis-related protein 5
PR-Bet v I family	Pathogenesis related protein bet v I family
PR-protein	Pathogenesis-related protein
Px Precursor	Class III peroxidase 14 precursor
QTL	Quantitative trait loci
RIL	Recombinant inbred line
ROS	Reactive oxygen species
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase
T10	Ten days post anthesis
T17	Seventeen days post anthesis
TLPR4 Precursor	Thaumatin-like pathogenesis-related protein 4 precursor

TO-m	Tocopherol *O*-methyltransferase
TPPT	Tocopherol polyprenyltransferase
TrxRs	Thioredoxin reductase
UPLC-PDA	Ultra-performance liquid chromatography with photodiode array detection
VIS	Visible light
WE18	Weslaco 2018
WGCNA	Weight gene co-expression network analysis
WGS	Whole genome resequencing
Y1	Yellow Seed 1

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

CONTRIBUTORS AND FUNDING SOURCES vi

NOMENCLATURE viii

TABLE OF CONTENTS xi

LIST OF FIGURES xiv

LIST OF TABLES xvi

CHAPTER I INTRODUCTION 1

Synopsis 1
Review of Literature 3
Genetic Control of Sorghum Grain Color 3
Genetic and Environmental Regulation of Black Pericarp Trait in Sorghum 6
Flavonoids and Plant Stress 8
Biosynthesis of Polyphenols 10
Transcription Factor Regulation of Polyphenol Biosynthesis 15
Environmental- and Tissue-Specific Biosynthesis of 3-DOAs in Sorghum 20
Objectives 22

CHAPTER II PHENOTYPIC, PHYTOCHEMICAL, AND TRANSCRIPTOMIC ANALYSIS OF BLACK SORGHUM PERICARP IN RESPONSE TO LIGHT QUALITY* 24

Synopsis 24
Introduction 25
Materials and Methods 29
Plant Material and Growth Room Conditions 29
Sample Preparation, Extraction, and Ultra-Performance Liquid Chromatography with Photodiode Array Detection (UPLC-PDA) of 3-DOAs 32

Pericarp Extraction and Sample Preparation for RNA-seq Analysis 34
RNA Isolation, Illumina Library Construction, and Transcriptome Sequencing..... 35
Results 37
Phenotype Development Under Select Light Regimes in a Controlled-Environment Growth Room 37
3-DOA Accumulation in Black Sorghum under Field Conditions versus Growth Room Conditions 44
Comparative Transcriptomics of Pericarp Samples under Different Light Regimes 45
Genes Involved in Secondary Metabolism 48
Genes Involved in Stress Response 51
Genes Involved in Defense Response 52
Discussion 53

CHAPTER III IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH 3-DOA PRODUCTION IN BLACK GRAIN SORGHUM 61

Synopsis 61
Introduction 63
Materials and Methods 69
Parental Material 69
Expression of Black Pericarp Phenotype and 3-DOA Accumulation after Multiple Generations of Grain Development in Darkness 69
Recombinant Inbred Line Population 70
Phenotyping 71
Genotyping, Linkage Mapping, and QTL Analysis of the F_5 RIL population 72
Bulked Segregant Analysis of Phenotypic Extremes in the F_5 RIL Population 75
Bulked Segregant Analysis of Phenotypic Extremes in the F_2 Population 78
Screening for Candidate Gene Deletion in the Phenotypic Extremes in the F_5 RIL Population 82
Results 83
Correlation between Colorimetric and UPLC-PDA Quantification of 3-DOAs 83
Black Pericarp Phenotype and 3-DOA Accumulation after Multiple Generations of Unilluminated Grain Development 84
Phenotyping F_5 RIL population for Pericarp 3-DOA Concentration in Field Environments 86
Genotyping, Linkage Mapping, and 'F_5 QTL_{GBS} analysis' of 224 F_5 RILs 87
F_5 RIL Population Whole Genome Resequencing-based Bulked Segregant Analysis, 'F_5 BSA_{WGS}' 91
Bulked Segregant Analysis of Phenotypic Extremes in an F_2 Population using GBS, 'F_2 BSA_{GBS}' 93
Collocating Genomic Regions Associated with Pericarp 3-DOA Concentration from the Different Mapping Methodologies 96
Discussion 105

CHAPTER IV CANDIDATE GENE NETWORK FOR ORGAN-SPECIFIC REGULATION OF 3-DOA PRODUCTION IN BLACK SORGHUM PERICARP TISSUE 113

Synopsis 113
Introduction 114
Materials and Methods 117
Comparative Transcriptomics of Sorghum Pericarp, Flag leaf, and Root Samples 117
Weighted Gene Co-Expression Network Analysis 119
Conserved Motif Distribution Analysis of Hub Genes 121
Results 121
Comparative Transcriptomics of Pericarp, Flag leaf, and Root Tissues 121
Weighted Gene Correlation Network Analysis 126
Conserved DNA Sequence Motif Distribution Analysis of Hub Genes 133
Discussion 135

CHAPTER V CONCLUSIONS 144

REFERENCES 147

APPENDIX SUPPLEMNTARY MATERIAL 168

Supplementary Figures 168
Supplementary Tables 170

LIST OF FIGURES

Page

Figure 1. Biosynthetic pathway of major sorghum grain polyphenols 14

Figure 2. Color development of sorghum rachis branches subjected to varying light regimes at 10 and 17 days after exposure during grain filling and at final grain maturity..... 39

Figure 3. Number of differentially expressed genes upregulated in RTx3362 exposed to VIS + UVA + UVB compared to all other treatments categorized by Gene Ontology (GO) functional groups 47

Figure 4. Heatmap showing transcript abundance of genes related to phenylpropanoid and flavonoid biosynthesis as well as stress and defense response in RTx3362 exposed to No light (A); VIS (B); VIS + UVA (C); VIS + UVA + UVB (D); and BTx378 exposed to VIS + UVA + UVB (E) 49

Figure 5. Schematic illustration of the F_2 and F_5 mapping populations generated from a biparental cross used for linkage mapping of genomic regions associated with 3-DOA accumulation in sorghum grain 81

Figure 6. Correlation between the colorimetric method and UPLC-PDA quantification of luteolinidin, the predominant 3-DOA in black sorghum pericarp..... 84

Figure 7. 3-DOA concentration and pericarp pigment development in three sorghum treatments grown in College Station, TX 2019 86

Figure 8. Genetic linkage map constructed from 224 F_5 RILs for each linkage group representing the 10 sorghum chromosomes as visualized using R/qtl. 88

Figure 9. Illustration of QTL detected for 3-DOA accumulation across 10 sorghum chromosomes in an F_5 RIL population utilizing the F_5 QTL_{GBS} analysis 91

Figure 10. WGS reads from BTx378 and RTx3362 illustrating the deletion of Sobic.004G185100 gene in RTx3362 genome on chromosome 4 103

Figure 11. PCR screening for presence or absence of coding region sequences of Sobic.004G185100 (230 bp) and Sobic.002G0074400 (116 bp) genes in the 20 highest- and 20 lowest- F_5's for pericarp 3-DOA concentration..... 104

Figure 12. Heatmap showing DEGs of network genes between pairwise comparisons. 130

Figure 13. Weighted Gene Co-expression Network Analysis of genes (nodes, gray circles) that are most correlated with the accumulation of 3-DOAs in the black pericarp and have the highest co-expression values 132

Figure 14. Model of UVB induced ROS, redox cycling, endogenous elicitor generation and signal transduction leading to 3-DOA accumulation in black pericarp tissue. 141

LIST OF TABLES

Page

Table 1. Light transmissivity of enclosures to block sorghum panicles from specific wavelengths of light. 32

Table 2. Analysis of variance for Ultra-Performance Liquid Chromatography with Photodiode Array Detection quantification of 3-DOAs in developing and mature *Sorghum bicolor* grain for five separate models (A-E). 42

Table 3. Connecting letter reports according to the all pairs Tukey-Kramer method for Ultra-Performance Liquid Chromatography with Photodiode Array Detection quantification of 3-DOAs in developing and mature *Sorghum bicolor* grain for five separate models (A-E). 43

Table 4. Summary of mapping and analytical methods used to map loci associated with 3-DOA accumulation in mature sorghum grain. 80

Table 5. Statistical summary of grain 3-DOA concentration of field-grown RILs from three environments. 87

Table 6. Mean squares of ANOVA for the colorimetric quantification of 3-DOA concentration of field-grown RIL grain from three environments 87

Table 7. Genetic linkage map marker coverage across the 10 sorghum chromosomes generated from 224 F_5 RILs. 89

Table 8. QTL detected for pericarp 3-DOA accumulation in the 224 F_5 RILs using BLUPs. 90

Table 9. Bulked segregant analysis using the QTLseqr method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_5 RILs 92

Table 10. Bulked segregant analysis using the QTLseqr method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_2 BSA_{GBS-42} 95

Table 11. Bulked segregant analysis using the QTLseq method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_2 BSA_{GBS-20} 96

Table 12. Summary of genomic regions significantly associated with pericarp 3-DOA concentration from the different mapping methodologies 98

Table 13. Intervals, differential expression, and Gene Ontology terms of genes within major QTL detected among mapping methods. 101

Table 14. Summary table of samples used for RNA-seq 119

Table 15. RNA-seq samples compared for DEG analysis 123

Table 16. Gene ontology (GO) terms enriched in 42 co-expressed genes discovered in network analysis 128

Table 17. The most highly conserved DNA sequence motifs identified from the promoter region of the 42 co-expressed hub genes. 134

CHAPTER I

INTRODUCTION

Synopsis

Sorghum [*Sorghum bicolor* (L.) Moench] is a multipurpose C4 grass bred for feed, silage, and human consumption (FAOSTAT 2019). Originating in the semi-arid tropics of Africa and Asia, sorghum is a major staple crop that can be successfully produced in stressful environments with reduced inputs (Morris et al. 2013). The environmental plasticity of sorghum and introgression of early flowering and dwarfing alleles has allowed for the expansion of tropical sorghum production to temperate environments (Olsen et al. 2013; Morris et al. 2013).

Sorghum is largely a self-pollinated diploid with two copies of 10 chromosomes (2n=2x=20), a moderate genome size of ~730 Mb, and high homology to the rice genome (Paterson et al. 2009). Many genomic resources have been developed for sorghum including several reference genome sequences, gene expression atlases, and a genotyping-by-sequencing pipeline for genotypic analysis in a high throughput and informative manner (Paterson et al. 2009; Olson et al 2014; McCormick 2017; Morishige et al. 2013). These resources permit detailed investigations into the environmental and genetic factors that control traits of importance to sorghum production in the US and worldwide.

A rare black pericarp trait found in grain sorghum has notable value in the specialty food and nutraceutical industries. The black appearance of these sorghum grains are associated with the biosynthesis of the polyphenols 3-deoxyanthocyanidins (3-

DOAs), a group of compounds not commonly present in higher plants. These polyphenols are unique from more common anthocyanins in that 3-DOAs are more stable in aqueous solutions due to the absence of a hydroxyl group at position C-3 of the flavylium ion. This stability is highly valued for use in natural food colorants, organic food preservatives, and antioxidant food additives (Khalil et al. 2010). Furthermore, phytochemical research has demonstrated the cytotoxicity of 3-DOAs on human cancer cells and proven that sorghum is a convenient and scalable source for the bioproduction of these compounds (Shih et al. 2007).

To better understand the mechanisms and pathways involved in the biosynthesis of 3-DOAs in sorghum tissues, previous research has concluded several important aspects regarding this trait. The black pericarp was determined to be controlled by two to twelve recessive genes and the trait is not fully penetrant in all environmental conditions (Pfeiffer and Rooney 2016). Studies (Pfeiffer and Rooney 2015) in which panicles were shaded during seed maturation showed that light is critical for 3-DOA accumulation and for the black pigmentation of pericarp. Observations in tropical nurseries, where the trait is not fully penetrant, suggest that environmental factors related to day length affect the full expression of this trait (WL Rooney, personal communication). Utilizing greenhouses constructed with UV-blocking or UV-transmitting plastics, Wu et al. (2017) suggested that exposure of maturing panicles to UV light increased the biosynthesis of 3-DOAs. The implication of light quality effecting the black pericarp parallels numerous studies with horticultural crops where UV- and/or blue light were critical for anthocyanin accumulation in grapes (Azuma et al.

2012; Koyama et al. 2012), petunia (Ryan et al. 2002) and apples (hu Dong et al. 1995; Ubi et al. 2006).

The biosynthesis of 3-DOAs in sorghum is not restricted to the pericarp. Upon leaf wounding or pathogen attack, sorghum plants with the *P* gene turn purple due to the induction of 3-DOA biosynthesis (Kawahigashi et al. 2016b), and the 3-DOAs have been implicated to act as phytoalexins leading to a hypersensitive response in resistant genotypes (Snyder and Nicholson 1990). In sorghum leaves, the induction of 3-DOA biosynthesis does not require exposure to light (Weiergang et al. 1996), which is in stark contrast to the induction of 3-DOA biosynthesis in sorghum pericarp (Pfeiffer and Rooney 2015). Thus, the biosynthesis of 3-DOAs in leaves and pericarp of sorghum represent unique developmental systems, which thereby provide an opportunity to elucidate those environmental, genetic, and cellular factors conditioning the black pericarp phenotype. The extensive research over the past decades into characterizing and developing black sorghum genotypes at Texas A&M University has established the necessary resources and infrastructure for elucidating the genetic, physiological, and environmental factors controlling the development of the black pericarp trait.

Review of Literature

Genetic Control of Sorghum Grain Color

Grain color of sorghum [*Sorghum bicolor* (L.) Moench] is influenced by a suite of genes including those that control pericarp color and thickness, the presence of a testa layer, and the color and structure of the endosperm (Smith and Frederiksen 2000). The sorghum kernel has several layers of tissue, each contributing some importance to

determining grain color. The pericarp (or epicarp) is the outermost layer of the kernel and is traditionally described as red, yellow, or white, and this trait is controlled by the *R* and *Y* genes. *R* and *Y* epistatically interact to produce red *(R_Y_)*, white/colorless *(R_yy/rryy)*, or yellow *(rrY_)* sorghum (Smith and Frederiksen 2000). Red or yellow types that also possess the dominant intensifier gene, *I*, will show an increase in pericarp color (Ayyangar et al. 1933a; Smith and Frederiksen 2000). From previous test crossing experiments, black sorghum was determined to be genetically red (*R_Y_*) (Rooney et al. 2013). Because of the unique phenotype of black sorghums, it was concluded that additional genetic factors beyond the *R* and *Y* genes are functioning to control seed color in black sorghums (Pfeiffer 2017; Rooney et al. 2013). The mesocarp, the layer below the pericarp, can vary in thickness depending on a single dominant *Z* gene. A thin mesocarp (*Z_*) can lead to a translucent appearance allowing the endosperm color to be more noticeable. In contrast, a thick (*zz*) mesocarp will make the grain appear chalky and can hide the appearance of a testa or other endosperm characteristics (Smith and Frederiksen 2000).

The testa is a thin layer of cells that may or may not be present directly below the mesocarp. The presence or absence of the testa layer is determined by complementary dominant genes, B_1 and B_2. For the testa layer to be present, the plant must possess both loci (B_1B_2) in their dominant form. If the testa is present, it can be pigmented depending on the *Tp* gene. A dominant *Tp_* genotype will have brown pigmentation while a recessive *tptp* testa will be purple (Smith and Frederiksen 2000). A spreader gene, *S*, will only be expressed in the presence of dominant B_1 and B_2. This gene causes

pigments from the testa to migrate into the pericarp (Vinall and Cron 1921). More specifically, in genotypes $B_1B_1B_2B_2$ with a dominant *S*, the pericarp will appear brown regardless of other genetic influence. Genotypes that have a partial testa layer for unknown reasons (Blakely et al. 1979) that possess a dominant *S* may accumulate brown splotches in the pericarp resembling those caused by the *Pb* gene (Smith and Frederiksen 2000). Some black sorghum varieties have a testa layer, such as in Black PI Tall or the commercial hybrid Onyx, while others lack a testa layer including Tx3362 and Black Tx430 (Dykes et al. 2005; Hayes 2012).

Sorghum grain can be categorized into three groups according to chemical and genetic makeup (Hahn and Rooney 1986; Rooney et al. 1981). Group I does not have a pigmented testa, are low in phenols, and contains no tannins. Group II have a pigmented testa with tannins and a recessive spreader gene, and group III sorghums contain a pigmented testa with tannins plus a dominant spreader allele (Hahn and Rooney 1986). Sorghum tannin levels are further controlled by functional alleles at the *Tan1* (Sobic.004G280800) and *Tan2* loci (Hahn and Rooney 1986). If one or both of these alleles are in their homozygous recessive state, the seed will lack tannin (Hahn and Rooney 1986; Wu et al. 2012).

Separate from grain color, sorghum plants are also characterized by leaf-sheath and glume color; tan, red, or purple (Ayyangar et al. 1933b). Secondary plant color is controlled by epistatic *P* and *Q* genes. Plants with *P_Q_* genes produce purple pigmented plants while those with *P_qq* are red. Plants with either *ppqq* or *ppQQ* genotypes will be tan-pigmented (Valencia and Rooney 2009). Pericarp and secondary

plant color genes are generally considered to act independently (Dykes et al. 2009); however, there are some instances where both sets of genes contribute to grain color. Secondary plant and glume color are said to influence the maternal tissue of the grain and modify effects of other direct color factors (Valencia and Rooney 2009). For example, a dominant *Pb* gene induces purple splotching in the pericarp of *P_Q_* or *P_qq* genotypes, but most grain sorghums lack these purple spots as this trait was selected against during sorghum domestication (Ayyangar et al. 1938). Sorghum plants with purple secondary plant color produce 3-DOAs (apigeninidin and luteolinidin) in the leaves as a defense response to wounding and pathogen attack whereas tan genotypes produce the flavones apigenin and luteolin. The loss of function of the dominant *P* gene will result in a tan secondary leaf color and the absence of biosynthesis of 3-DOAs in the leaves (Kawahigashi et al. 2016b). Finally, Dykes (2010) determined that the highest possible levels of 3-DOAs in black sorghum grain cultivars were observed in genotypes with purple secondary plant color and a thick pericarp.

Genetic and Environmental Regulation of Black Pericarp Trait in Sorghum

The black sorghum trait was originally discovered in the cultivar 'Black Shawaya' from Western Sudan (Hahn et al. 1984). Black pericarp sorghum was determined to be genetically red (*R_Y_*) from test crossing with parents of various pericarp colors (Rooney et al. 2013); however, it was concluded that there must be additional genetic factors to explain black sorghum's distinguishable phenotype from a typical red pericarp sorghum (Dykes et al. 2009; Rooney et al. 2013). Through efforts to introgress the black pericarp phenotype into adapted backgrounds, the complexity of the

trait's inheritance was revealed as the frequency of this phenotype in an F_2 population of ~1 black pericarp out of 1000 F_2 segregants (Pfeiffer 2017). This suggests that, in addition to the *R* and *Y* genes, the black pericarp trait in sorghum is conditioned by additional recessive genes (Pfeiffer 2017; Rooney et al. 2013). Using generation means analysis, Pfeiffer and Rooney (2016) showed multiple genes with additive, dominance, and epistatic interactions were controlling the black pericarp color.

Efforts to characterize the black pericarp revealed that the colorization of the pericarp correlates with enhanced biosynthesis of 3-DOAs suggesting that the same genetic factors control both traits (Dykes et al. 2005; Dykes et al. 2009; Dykes et al. 2013). As the two major 3-DOAs that accumulate in sorghum pericarp are orange (luteolinidin) and yellow (apigeninidin), the identity of the black metabolite remains unknown. As seen in pathogen-induced color variation in sorghum leaves, it was concluded that the darkest leaf lesions were due to high levels of luteolinidin and that this compound was responsible for the coloration (Mizuno et al. 2014); however, this study did not address why this naturally orange compound may appear as black in the leaf tissue. It has been speculated that the pericarp color is developed by the oxidation of flavonoids during maturation (Waniska and Rooney 2000).

Based on the reported correlation of the black appearance of pericarp and 3-DOA levels, Pfeiffer and Rooney (2016) examined the inheritance of the trait using three phenotyping methods: seed visual color ratings, seed quantitative colorimeter measurements, and pericarp 3-DOA levels. Using the different phenotypic methodologies (and the estimation equation employed), the authors concluded that

between 2 to 12 recessive genes control the black phenotype. It should be noted that the observed frequency of ~1:1000 black F_2 phenotypes based on visual color ratings would implicate ~5 segregating genes that condition the trait while heritability estimates established using 3-DOA measurements suggest 2 to 4 genes controlling this trait (Pfeiffer 2017; Pfeiffer and Rooney 2016).

In addition to the complex genetic inheritance, light quality influences the development of the black pericarp and the biosynthesis of grain 3-DOAs. When panicles of black sorghum varieties were shaded, the black colored pericarp and the biosynthesis of 3-DOAs was reduced, and both high intensity and duration of light exposure were found to be critical to the development of the black phenotype and accumulation of 3-DOAs (Pfeiffer and Rooney 2015). Interestingly, the intermediate flavanones naringenin and eriodictyol accumulated to similar levels regardless of shading treatment (Dykes et al. 2009) suggesting that light-induced accumulation of 3-DOAs may be due to activation of late biosynthetic steps in the phenylpropanoid pathway.

Flavonoids and Plant Stress

Flowering plants produce a variety of secondary metabolites for UV protection, insect attraction, pathogen defense, variation of flower color, male fertility, pollination, allelopathy, and auxin transport (Winkel-Shirley 2001). During abiotic stress, plants generate reactive oxygen species (ROS), which can damage DNA, lipids, proteins, and sugars. Plants evolved to produce antioxidants to combat oxidative stress. Flavonoids often suppress the generation of ROS or reduce ROS levels once formed by inhibiting

ROS-generating enzymes, recycling other antioxidants, or by chelating transition metal ions such as iron and copper (Baskar et al. 2018). Antioxidants' specific activities are dependent on their structural class, degree of hydroxylation, substitutions, conjugations, and polymerization (Heim et al. 2002). Flavonoids scavenge molecules such as superoxides and hydroxyl radicals once each have formed in the cell (Baskar et al. 2018). The use of elicitors such as methyl jasmonate (Portu et al. 2018; Wang et al. 2015), ethylene (Heredia and Cisneros-Zevallos 2009), harpin, chitosan (Ruiz-García and Gómez-Plaza 2013), and UV-B radiation (Pandey and Pandey-Rai 2014) have proven to enhance flavonoid production via induction of free radicles.

Although plants require light for photosynthesis, UV radiation is a component of sunlight that can be harmful to plant processes at high intensities (Baskar et al. 2018). Tolerance to UV radiation depends on the plant's ability to repair and acclimate to UV exposure, which is often associated with the biosynthesis of flavonoids (Baskar et al. 2018). Epidermal flavonoids absorb UV-B but do not absorb visible light as this would cause a reduction in photosynthetically active radiation (Lower et al. 1997). Additionally, flavonoids do not absorb but instead transmit UV-A as UV-A is critical to plant development via activation of the blue light/UV-A photoreceptor (Lower et al. 1997). In many plant species, flavonoid biosynthesis can be initiated through the stimulation of UV-B photoreceptors that triggers enzymes and transcription factors to activate genes in the phenylpropanoid pathway (Greenberg et al. 1997). In the epidermal layer of rye seedlings, UV-B induced *trans* to *cis* isomerization of the cinnamic acid enzyme which leads to the activation of phenylalanine ammonia lyase and the

subsequent synthesis of flavonoids (Braun and Tevini 1993). In parsley leaves, UV light induced the synthesis of phenylalanine ammonia lyase and chalcone synthase (Logemann et al. 2000). Based on existing literature, the black sorghum grain phenotype is known to require light (Pfeiffer and Rooney 2015), and UV-B (Wu et al. 2017) has been implicated as a critical factor for the development of black pericarp phenotype and the hyper-accumulation of 3-DOAs. Further research is necessary to better characterize the spectra of light that are critical for inducing 3-DOA biosynthesis in sorghum pericarp.

Biosynthesis of Polyphenols

Polyphenols encompass a group of phytochemicals characterized by their phenolic structural features, but are subdivided into several major classes based on source of synthesis, biological function, and chemical structure (Tsao 2010). Polyphenol subclasses include phenolic acids, flavonoids, polyphenolic amides, and other non-flavonoid polyphenols. Phenolic acids include benzoic and cinnamic acid (Tsao 2010) while flavonoids consist of six major subgroups: anthocyanidins, flavanols, flavonols, flavanones, flavones, and isoflavonoids (Sak 2014). Three groups of flavonoids predominantly exist in grains of various sorghum genotypes: 3-DOAs, flavones, and flavanones (Taleon et al. 2012). Non-flavonoid polyphenols contain compounds most familiar to the public such as resveratrol in wine, ellagic acid in berry fruits, lignans in flax and sesame, and curcumin found in turmeric (Tsao 2010). Polyphenolic amides include capsaicinoids such as those found in chili peppers responsible for hotness (Davis et al. 2007) and avenanthramides in oats (Bratt et al. 2003). In wheat, rice, and maize,

selection pressure during domestication reduced polyphenol levels in the grain thereby reducing the bitter taste and dark grain pigmentation associated with polyphenols. By contrast, sorghum is rather unique amongst the major domesticated cereals as much of the phenotypic and genetic diversity for polyphenol content still resides in sorghum germplasm (Morris et al. 2013; Olsen and Wendel 2013).

For clarification, anthocyanins differ from anthocyanidins in that anthocyanins are a glycosidic form and responsible for several red, blue, and purple pigments in flower petals, fruits, vegetables, and some grains (Tsao 2010). Anthocyanidins (besides 3-DOAs) include cyanidin, delphinidin, malvidin, pelargonidin, petunidin, and peonidin (Dykes 2008). The two major 3-DOAs, orange luteolinidin (3',4'-hydroxylated) and yellow apigeninidin (4'-hydroxylated) (Mizuno et al. 2014; Shih et al. 2006), are unique from other common antho-cyanins/cyanidins in that 3-DOAs lack a hydroxyl group at the 3' carbon position. This distinct hydroxylation pattern results in more stable molecules making 3-DOAs useful in the food and nutraceutical industries (Mazza and Brouillard 1987; Sweeny and Iacobucci 1981).

The biosynthesis of flavonoids (anthocyanins, proanthocyanidins, flavones, flavonols, flavanones, flavan-4-ols, isoflavonoids) arise from different metabolic branch points within the phenylpropanoid pathway, but all flavonoids share a series of common early biosynthetic steps (Petti et al. 2014; Pfeiffer 2017). Most relevant to the proposed research is the biosynthesis of flavones, flavanones, and 3-DOAs, which are derived from two partially overlapping and competing pathways in sorghum (Lo and Nicholson 1998). The phenylpropanoid and flavonoid pathway shown in Figure 1 begins with the

early biosynthetic steps resulting in the production of naringenin from phenylalanine. The early biosynthetic steps start with the first committed enzymatic step catalyzed by phenylalanine ammonia lyase (PAL) followed by trans-cinnamate 4-monooxygenase (C4H), 4-coumarate:CoA ligase (4CL), and chalcone synthase (CHS) (Mizuno et al. 2012). Chalcones are then isomerized into naringenin via chalcone isomerase (CHI). Naringenin is the common intermediate for all flavonoids and the point in the branch point in the pathway at which competition for this substrate occurs (Baskar et al. 2018; Miranda et al. 2012). The branch pathway leading to 3-DOA synthesis includes specific dihydroflavonol 4-reductases (DFR). These enzymes selectively reduce dihydroflavonols (Nakatsuka and Nishihara 2010) and are also proposed to encode a flavanone 4-reductase (FNR) enzyme (Halbwirth et al. 2003; Stich and Forkmann 1988). The FNR enzyme then converts flavanones (naringenin and eriodictyol) to flavon-4-ols (apiferol and luteoforol) (Nakatsuka and Nishihara 2010). The FNR reduces the carbonyl group at the number 4 carbon of naringenin and/or eriodictyol to yield these compounds (Lo and Nicholson 1998). The biosynthesis of luteoforol also requires the enzyme flavonoid 3'-hydroxylase (F3'H) that introduces a hydroxyl group at the 3' position of the B ring of naringenin. Flavon-4-ols, apiforol and luteoforol, are converted to 3-DOAs via the action of another F3'H (Mizuno et al. 2016). To date, the final enzymatic step(s) in the pathway to 3-DOAs have yet to be elucidated (Mizuno et al. 2016; Shih et al. 2006). It has been proposed that the final enzymatic step involves an unidentified putative anthocyanidin synthase (ANS) (Mizuno et al. 2012) or a more

recently proposed fast-acting FNR (Mizuno et al. 2016). Finally, 3-DOAs can also exist in a glucosylated form (Nakatsuka and Nishihara 2010).

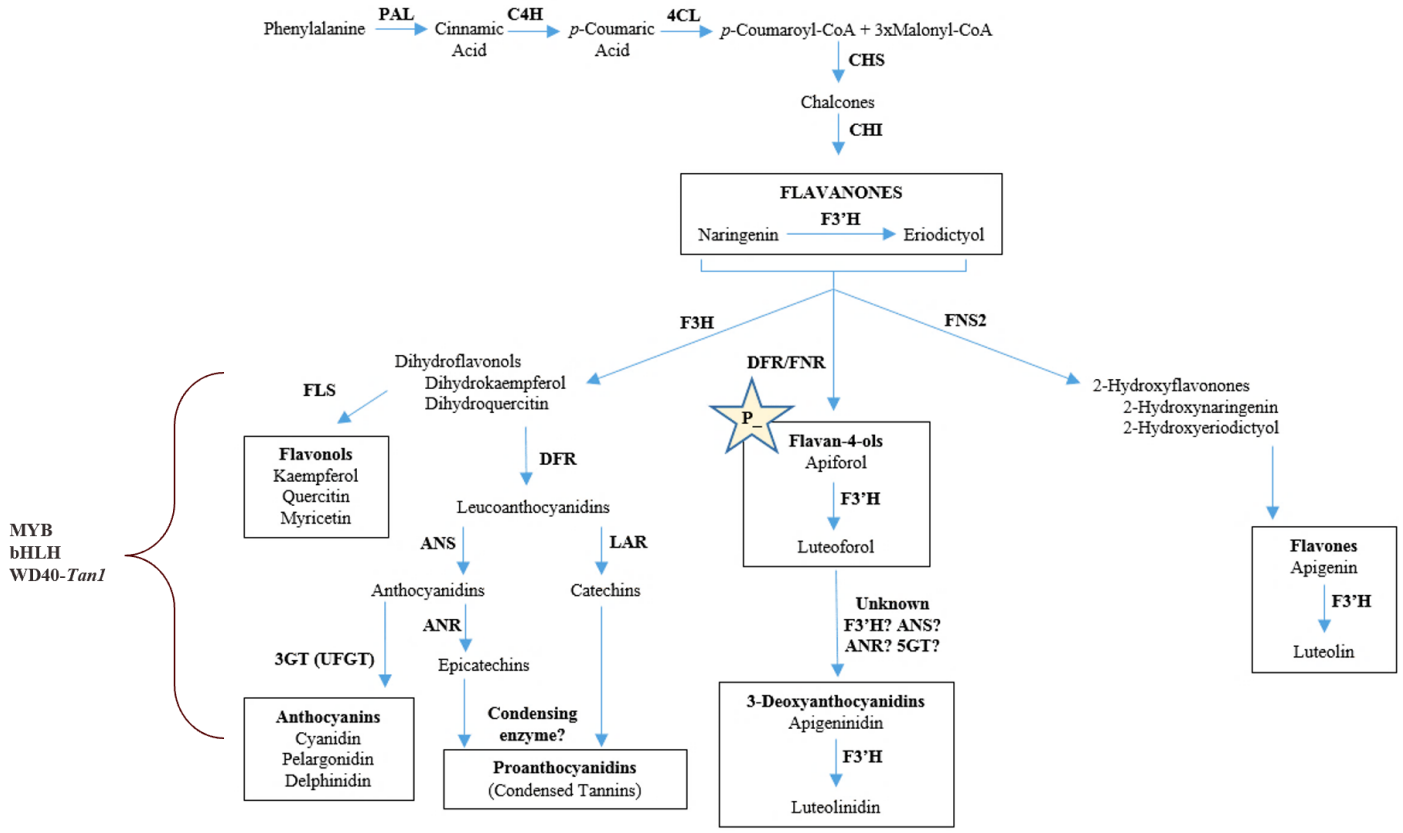


Figure 1. Biosynthetic pathway of major sorghum grain polyphenols. PAL phenylalanine ammonia lyase; C4H cinnamate 4-hydroxylase; 4CL coumarate 4-ligase; CHS chalcone synthase; CHI chalcone isomerase; F3H flavanone 3-hydroxylase; F3'H flavonoid 3'-hydroxylase; FNR flavanone 4-reductase; FNS2 flavone synthase II; FLS flavonol synthase; DFR dihydroflavonol reductase; ANS anthocyanidin synthase; ANR anthocyanidin reductase; 5GT anthocyanin 5-O-glucosyltransferase; LAR leucoanthocyanidin reductase; 3GT (UFGT) UDP glucose:flavonoid 3-O-glucosyltransferase. Adapted from Sharma et al. 2012; Kawahigashi et al. 2016a; Winefield et al. 2005.

Black pericarp sorghum produces elevated levels of 3-DOAs compared to flavones (high in red grain sorghum) and flavanones (high in yellow grain sorghum). Selection of the metabolic pathway from naringenin to produce the flavones apigenin and luteolin depends upon the enzymatic activity of flavone synthase II (FNSII) to produce apigenin and F3'H to produce luteolin (Mizuno et al. 2016). However, if FNR and F3'H are active, 3-DOAs are synthesized (Mizuno et al. 2016). From this research it was hypothesized that if FNR is expressed, it is dominant to the FNSII-dependent pathway when both pathways are activated. In summary, FNR and F3'H activity will direct the pathway towards 3-DOA synthesis but if the FNR is nonfunctional, the pathway will produce flavones and flavanones (Mizuno et al. 2016).

Transcription Factor Regulation of Polyphenol Biosynthesis

The flavonoid biosynthesis pathway is controlled by key transcription factors that regulate structural genes. Six families of transcription factors, MYB, basic helix-loop-helix (bHLH), WD40, WRKY, Zinc finger, and MADS box proteins, are known to be involved in the flavonoid biosynthesis pathway (Terrier et al. 2009). The transcription factors relevant to this research are MYB, bHLH, and WD40, which can work independently or as a complex to control several enzymatic steps in the flavonoid pathway (Mano et al. 2007). In *Arabidopsis*, it was shown that early biosynthetic genes are activated by MYB transcription factors resulting in the biosynthesis of flavonols whereas late biosynthetic genes require a ternary complex of transcription factors, MYB-bHLH-WD40, leading to the biosynthesis of proanthocyanidins (Li 2014). Regulation of different flavonoid pigments by distinct combinations of MYB, bHLH, and WD40

transcription factors was also seen in maize (Cone 2007). In the aleurone tissue of maize seeds, structural anthocyanin genes were activated by a MYB-WD40-bHLH triad (Cone 2007); however, phlobaphenes were synthesized under the control of just MYB domain P protein, independent of a bHLH partner (Sainz et al. 1997). A similar transcription activation pattern was observed in anthocyanin biosynthesis in highly pigmented apples (Espley et al. 2007).

MYB factors are a functionally diverse family of proteins containing five groups: R4-MYB, R3-MYB, R2R3 MYB, single-repeat MYB, and MYB-like (Dubos et al. 2010; Jin and Martin 1999; Rosinski and Atchley 1998). They are involved in numerous functions including the control of developmental regulations, metabolism, and biotic and abiotic stress responses (Baskar et al. 2018). The *Yellow seed1* (*Y1*) gene encodes an R2R3 MYB regulatory protein which plays a key role in the biosynthesis of 3-DOAs in sorghum vegetative tissue (Mizuno et al. 2012). Sorghum genotype Tx623, a tan secondary color plant with white pericarp, carries a deletion in the *Y1* gene resulting in leaves that lack 3-DOA pigments (Boddu et al. 2005). The *Y1* R2R3 MYB transcription factor regulates a suite of structural genes in the phenylpropanoid and flavonoid pathway including CHS, CHI, and DFR (Boddu et al. 2006). *Y1* also controls F3'H, an enzyme required for biosynthesis of 3-DOAs in leaves in response to pathogen attack or wounding (Ibraheem et al. 2010). Maize lines were transformed to express the sorghum *Y1* gene (containing 9164 bp of the y1 gene [AY860968 (Ibraheem et al. 2010)] with 2375 bp of the 5' regulatory region, 6946 bp sequence with three exons and two introns, and 820 bp of the 3′UTR) and exposed to *Colletotrichum graminicola*, a fungal pathogen

known to induce 3-DOA biosynthesis in sorghum leaves. *Y1*-maize transgenic lines accumulated high levels of luteolinidin and undetectable levels of apigeninidin (Ibraheem et al. 2015) supporting the importance of this R2R3 MYB transcription factor in 3-DOA accumulation. Although the accumulation of 3-DOAs via *Y1* gene action in both sorghum and maize leaves in response to fungal challenge has been demonstrated, the specific involvement of *Y1* in the regulation of 3-DOAs in black sorghum pericarp is unclear. In an initial transcriptome study of pericarp from red and black sorghum seed, Pfeiffer (2017) provided evidence that the black pericarp phenotype is not simply the result of the differential expression of the *Y1* R2R3 MYB factor between black and red pericarp tissues. While the R2R3 MYB transcription factor encoded by *Y1* is clearly important for determining seed color, this initial transcriptome analysis of Pfeiffer (2017) is consistent with inheritance studies of the black phenotype that concluded that the genetic control of this trait is complex and involves more than a single regulatory gene.

MYB transcription factors have been implicated in controlling flavonoid biosynthesis in a wide range of species and tissues including *Arabidopsis* (Nesi et al. 2001), apple (Espley et al. 2007), grape (Czemmel et al. 2009), and maize (Goff et al. 1992). Depending on the MYB, high levels of visible and UV-B radiation are known to up- (Mellway et al. 2009) or downregulate (Jin et al. 2000) MYB transcriptional activity. In *Arabidopsis* leaves, it was demonstrated that the expression of a MYB homolog was stimulated by light and downregulated by darkness (Nguyen et al. 2015). In an R2R3 MYB transcription factor in grapevine, *VvMYBF1,* gene expression was activated by UV

light (Czemmel et al. 2009). A similar induction pattern can be seen in *Populus spp.* where UV light as well as fungal infection and mechanical wounding activated gene expression of MYB transcription factors (Mellway et al. 2009). Evidence of transcriptional activation of the flavonoid biosynthetic pathway via UV-regulated MYB factors may be crucial to explaining the light requisite for the induction of black sorghum's phenotype. Furthermore, understanding that MYBs can be induced by different environmental factors (light or pathogen) is useful in clarifying the similarities and differences for the biosynthesis of 3-DOAs in leaves vs. the pericarp of sorghum.

In addition to being transcriptional activators, MYBs can function as transcriptional repressors (Nguyen et al. 2015). MYBs particularly belonging to the R3-MYB class act as repressors of the flavonoid biosynthesis pathway (Ravaglia et al. 2013). Examples of these MYB repressors have been reported in *Arabidopsis* (Rowan et al. 2009), apple (Lin-Wang et al. 2011), and strawberry (Aharoni et al. 2001). It was proposed that these MYB repressors interfere with the activity of the MYB complex (Dubos et al. 2008; Matsui et al. 2008). To add to the complexity, different MYBs can act as activators or repressors within the same pathway during anthocyanin synthesis (Albert et al. 2014). In stressed leaves in Eudicots, an R2R3 MYB activator, WD40, and bHLH form a complex that activates transcription of the genes that encode flavonoid structural genes as well as R2R3 MYB- and R3 MYB repressors. The activation of the MYB repressors allows for feedback repression and mediation of the response to environmental stress (Albert et al. 2014).

The bHLH transcription factors regulate several processes such as floral organogenesis (Heim et al. 2003), epidermal cell-fate determination (Serna 2007), light signaling, and flavonoid biosynthesis (Toledo-Ortiz et al. 2003). The bHLH family is defined by a unique domain consisting of a basic region involved in binding of DNA *cis*-regulatory elements and a hydrophobic HLH region (Heim et al. 2003; Toledo-Ortiz et al. 2003). Ludwig et al. (1989) described the first plant bHLH protein, *Lc,* in maize that is responsible for tissue-specific anthocyanin biosynthesis via cooperation with MYBs *C1* and *Pl1*. In *Arabidopsis*, *Transparent Testa 8* (*TT8*) was found to encode a protein with a basic helix-loop-helix at its C terminus, which is required for transcriptional activation of two flavonoid biosynthetic genes late in the pathway including DFR and *Banyuls* (*BAN*) (Nesi et al. 2000). While this class of transcription factors has been extensively characterized in *Arabidopsis*, the role of bHLH transcription factors in flavonoid biosynthesis in sorghum pericarp and leaf tissues remains to be elucidated.

The family of WD40 repeat proteins provide a network for protein interaction with other cellular components and aid in the control of signaling cascades, cellular transport, and apoptosis (Baskar et al. 2018). In *Mattiola incana*, the WD40 protein *Transparent Testa Glabra 1* (*TTG1*) regulates flavonoid biosynthesis specifically because expression of the DFR gene is dependent on *TTG1*. Mutants in this study had a substitution in the WD motif resulting in the lack of expression of DFR (Dressel and Hemleben 2009). A WD40 gene in sorghum, *Tan1*, is involved in the regulation of CHS, CHI, DFR, leucoanthocyanidin reductase (LAR) and anthocyanidin synthase (ANS) (Kawahigashi et al. 2016b). Demonstrating a trend similar to that in *M. incana*,

Tx623 has a mutant tannin allele from a 10-bp insertion in the exon region (referred to as *tan1-b*) causing a frame shift at position 923nt (309 aa) and a truncated protein of only 318 amino acids resulting in the disruption of the WD40 protein structure (Wu et al. 2012). This genotype does not produce significant levels of 3-DOAs (Hill 2014) potentially due to this nonfunctional WD40. Other sorghum genotypes were found to have a 1-bp deletion in the coding region causing a frame shift and a premature stop codon resulting in a nonfunctional allele (referred to as *tan1-a* in the literature). The frequency of the *Tan1*, *tan1-a*, and *tan1-b* alleles in major cultivated races and important US breeding lines are 78%, 20%, and 2%, respectively (Wu et al. 2012).

Environmental- and Tissue-Specific Biosynthesis of 3-DOAs in Sorghum

One distinction to make is that the induction of 3-DOAs differs depending on location of the tissue within the plant. Evidence shows that vegetative tissue produces 3-DOAs, classified as phytoalexins, in response to and defense against fungal infection and wounding. This can occur independent of light (Liu et al. 2010; Lo and Nicholson 1998). The 3-DOAs accumulate within inclusions in the epidermal cells of the leaves, become pigmented, then move within the cell to the site of penetration (Snyder and Nicholson 1990). The inclusions burst and release their contents in an attempt to protect against the stressor (Lo and Nicholson 1998). It has been reported that in infected leaf tissue, the expression of the genes encoding PAL C4H, 4CL, CHS, CHI, F3H, DFR, FNSII, and ANS are upregulated (Liu et al. 2010; Mizuno et al. 2012; Mizuno et al. 2016). In contrast, light is required for 3-DOA synthesis in sorghum grain (Pfeiffer and Rooney 2015). Pfeiffer (2017) described some of the similarities and differences in gene

expression of black pericarp compared to previously reported gene expression data from 3-DOA synthesis in pathogen-challenged leaf tissue (Mizuno et al. 2012). Pfeiffer (2017) found five CHS genes upregulated in both sorghum pathogen-challenged leaf tissue and black pericarp and two upregulated CHS genes unique to only the leaf tissue. Genes of two CHI family members were induced in sorghum leaves by pathogens and their expression was upregulated in black pericarp tissue (Pfeiffer 2017). Three of the four genes encoding F3'H were upregulated in black pericarp tissue whereas only one of the F3'H gene family members (Sobic.004G200800) was upregulated in response to pathogen attack in leaves (Pfeiffer 2017). Further, one gene coding a DFR (Sobic.009G043800) was not expressed in sorghum leaves under pathogen attack (Liu et al. 2010) but upregulated over 2,300-fold in black sorghum pericarp (Pfeiffer 2017). A gene encoding another DFR (Sobic.003G230900) contributing to anthocyanin biosynthesis was specifically found to be light-induced in sorghum leaves (Liu et al. 2010). This gene was not upregulated in black pericarp tissue (Pfeiffer 2017).

The sorghum gene, Sobic.006G226800 is known as the dominant *P* gene for secondary plant color. The *P* gene–sometimes described to regulate DFR and FNR–is required for the conversion of naringenin and eriodictyol into flavon-4-ols (Kawahigashi et al. 2016b). This *P* gene has also been suggested to act as an anthocyanindin reductase late in the 3-DOA pathway (Kawahigashi et al. 2016b). Genotypes with a dominant *P* gene will have purple leaf color and produce 3-DOAs in the leaves while genotypes with a recessive *p* gene will have tan leaf color and not produce 3-DOAs (Kawahigashi et al. 2016b). Previous studies have shown that tan plants lack a functional DFR and FNR

while also showing greater activity of FNSII. This leads to a greater accumulation of flavones in leaf tissue (Kawahigashi et al. 2016a; Mizuno et al. 2016). Variation of color can be seen within purple plants and can be explained by the ratio between the two 3-DOAs, luteolinidin and apigeninidin (Mizuno et al. 2016). This balance is said to be controlled by F3'H, an enzyme which hydroxylates the 3' position of the B-ring of naringenin and produces the precursor to luteolinidin (Shih et al. 2006). Tan plants do not express the F3'H enzyme during injury thus the anthocyanidin pathway is not activated (Mizuno et al. 2016). Despite the importance of the *P* and *Y1* genes for the biosynthesis of 3-DOAs in the leaves and the evidence of partially overlapping metabolic pathways for 3-DOA synthesis, it is recognized that transcriptional regulation for color differs in the leaf and grain (Kawahigashi et al. 2016b). For example, genotypes with purple secondary plant color and 3-DOA biosynthesis in the leaves do not always have black grain color. The present dissertation research aims to clarify the genetic similarities and differences of 3-DOA biosynthesis in different parts of the sorghum plant.

Objectives

The research detailed here addresses in part the long-term goal of the TAMU sorghum research team to utilize genomic and biotechnological resources to elucidate the genetic basis and genes controlling key agronomic traits critical to sorghum germplasm improvement.

This dissertation consists of three interrelated research projects that address the environmental and genetic regulation in sorghum of the rare black pericarp trait and the

associated accumulation of 3-DOAs. The first research objective addressed the critical light spectrum that leads to the accumulation of 3-DOAs in black pericarp, and to conduct an exploratory transcriptome analysis to provide insight into the molecular events signaling 3-DOA biosynthesis. The second research objective sought to identify regions of the sorghum genome and candidate genes within trait loci that control the black pericarp phenotype. The mapping of black sorghum trait loci involved multiple genetic mapping methodologies and two genetic mapping populations as I seek to discover genes and genic regions critical to the black pericarp phenotype. The final research objective focused on identifying the gene regulatory network governing the induction of pericarp-specific 3-DOA biosynthesis by implementing a Weighted Gene Co-expression Network Analysis (WGCNA). From these results, a model of cellular and molecular pathways associated with 3-DOA induction in black sorghum pericarp tissue will be proposed. Collectively, the overall goal of this research program is to elucidate the environmental signals that control this rare value-added black pericarp trait while also discovering the genes and gene networks that coordinate the cascade of cellular events leading to the accumulation of 3-DOAs in a tissue- and genotypic-specific manner.

CHAPTER II

PHENOTYPIC, PHYTOCHEMICAL, AND TRANSCRIPTOMIC ANALYSIS OF BLACK SORGHUM PERICARP IN RESPONSE TO LIGHT QUALITY*

Synopsis

Black sorghum [*Sorghum bicolor* (L.) Moench] is characterized by the black appearance of the pericarp and production of 3-deoxyanthocyanidins (3-DOAs), which are valued for their cytotoxicity to cancer cells and as natural food colorants and antioxidant additives. The black pericarp phenotype is not fully penetrant in all environments, which implicates the light spectrum and/or photoperiod as critical factors for trait expression. In this study, black- or red-pericarp genotypes were grown under regimes of visible light, visible light supplemented with UVA or supplemented with UVA plus UVB (or dark control). Pericarp 3-DOAs and pericarp pigmentation were maximized in the black genotype exposed to a light regime supplemented with UVB. Changes in gene expression during black pericarp development revealed that UV light activates genes related to plant defense, reactive oxygen species and secondary metabolism suggesting that 3-DOA accumulation is associated with activation of flavonoid biosynthesis and several overlapping defense and stress signaling pathways.

---

*Reprinted with permission from Fedenia L, Klein RR, Dykes L, Rooney WL, Klein PE (2020) Phenotypic, Phytochemical, and Transcriptomic Analysis of Black Sorghum (*Sorghum bicolor* L.) Pericarp in Response to Light Quality. Journal of Agricultural and Food Chemistry 68:9917-9929. https://doi.org/10.1021/acs.jafc.0c02657 Copyright 2020 American Chemical Society.

Introduction

The black grain trait in sorghum [*Sorghum bicolor* (L.) Moench] has notable value in the specialty food and nutraceutical industries. The black coloration of the grain is associated with the biosynthesis of the polyphenols 3-deoxyanthocyanidins (3-DOAs), a group of compounds not commonly present in higher plants. To date, sorghum is considered the only human dietary source that accumulates substantial concentrations of 3-DOAs in both seed from the black-grained genotype and more commonly in leaf tissue of varying genotypes in response to pathogen attack and wounding (Awika et al. 2004; Kawahigashi et al. 2016a). One source of this trait is traced to the cultivar 'Black Shawaya', a plant introduction from western Sudan (Hahn et al. 1984). These polyphenols are unique from more common anthocyanins in that 3-DOAs are more stable in aqueous solutions due to the absence of a hydroxyl group at position C-3 of the flavylium ion. This stability is valued for use as natural food colorants, organic food preservatives, and antioxidant food additives (Khalil et al. 2010). Furthermore, a recent review by Xiong et al. (2019) summarized sixteen different health properties of 3-DOAs, half of which were cancer related and reiterated that these compounds are more cytotoxic to cancer cells than common anthocyanin analogs. In the same report (Xiong et al. 2019), the authors compiled an exhaustive list of research studies which demonstrated that concentrated 3-DOAs from sorghum extract may induce apoptosis and inhibit the proliferation of different types of malignant cells ranging from digestive to blood cancers. While chemical synthesis of 3-DOAs is possible, the process is complex and low yielding (Xiong et al. 2019), thus, sorghum grain and leaves remain the only

economical source of these valued compounds provided proper stimuli are present to trigger their production.

Sorghum grains contain a wide variety of flavonoids including flavones, flavanones, and 3-DOAs (Gous 1989), however, the elevated accumulation of 3-DOAs (mainly luteolinidin and apigeninidin) is unique to the black-grained genotype. Following the activity of the enzymes phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), and 4-coumarate:CoA ligase (4CL) in the phenylpropanoid pathway, the biosynthesis of flavones, flavanones, and 3-DOAs are derived from different metabolic branch points within the flavonoid pathway (Kawahigashi et al. 2016a; Lo and Nicholson 1998). An updated comprehensive description of the known biosynthetic steps in the flavonoid pathway leading to 3-DOA or anthocyanin production has been recently published (Xiong et al. 2019). Briefly, activity by the first two enzymes in the flavonoid biosynthetic pathway, chalcone synthase (CHS) and chalcone isomerase (CHI), leads to the production of naringenin. Naringenin can be converted to eriodictyol by the action of flavonoid 3'-hydroxylase (F3'H). In sorghum grain, three main enzymes can compete for the substrate naringenin resulting in the generation of different flavonoids. These enzymes include dihydroflavonol 4-reductase (DFR) (or a similar enzyme with flavanone 4-reductase (FNR) activity), flavone synthase II (FNSII), and flavanone 3-hydroxylase (F3H) (Kawahigashi et al. 2016a). The action of the enzymes DFR or FNR on naringenin veers the flavonoid pathway to produce flavan-4-ols and eventually 3-DOAs (Kawahigashi et al. 2016a). The activity of FNSII catalyzes naringenin to flavones that predominantly accumulate in red and yellow grain sorghum

(Kawahigashi et al. 2016a; Yonekura-Sakakibara et al. 2019). The action of a third enzyme, F3H, leads to the production of anthocyanin which accumulates in sorghum grain of various colors (Kawahigashi et al. 2016a). It should be noted that while several reports (Kawahigashi et al. 2016a; Mizuno et al. 2012; Nakatsuka and Nishihara 2010; Stafford 1994) have attempted to identify the final enzymatic steps leading to 3-DOA biosynthesis, considerable uncertainty still exists concerning the enzymes and the genes encoding the penultimate and final step(s) in 3-DOA biosynthesis.

Flavonoid profiles of sorghum grains are affected by the genes responsible for pericarp color, *R* (gene ID unknown) and *Y* (Sobic.001G398100 (Nida et al. 2019)) (Smith and Frederiksen 2000). Based on the classical genetic models of grain color in sorghum, black pericarp sorghum is considered genetically red (*R_Y_*) (Pfeiffer and Rooney 2016). Despite possessing the same classical genetic background for pericarp color, the genetic differences underlying the varying flavonoid profiles between red and black pericarp sorghums (Dykes 2008), requires additional explanation. A recent genetic study of black sorghum demonstrated that additional genes (including recessive alleles) beyond the dominant alleles at the *R* and *Y* loci function to control the black grain phenotype (Pfeiffer and Rooney 2015). Unlike 3-DOA accumulation in pericarp tissue, 3-DOA production in sorghum leaf tissue is a dominant and common trait irrespective of grain color and stimulated by pathogen attack or mechanical wounding (Kawahigashi et al. 2016a). Sorghum leaves are characterized as either tan (nonpigmented), red, or purple; a trait referred to as “secondary plant color” (Smith and Frederiksen 2000). Sorghum plants with purple secondary plant color produce high

levels of 3-DOAs in the leaves as a defense response whereas tan genotypes produce the flavones apigenin and luteolin (Kawahigashi et al. 2016a). In leaves, 3-DOAs serve as phytoalexins that are synthesized as biotic/abiotic defenses (Nielsen et al. 2004). As 3-DOA biosynthesis in pericarp has not been associated with pathogen attack, it appears that the environmental signals conditioning 3-DOA production in pericarp are unique from sorghum leaf lesions.

A limited number of studies have examined the environmental cues controlling the development of the black pericarp phenotype in sorghum. Pfeiffer and Rooney (2015) examined the effect of panicle shading on pericarp pigmentation and determined that either a reduction in light intensity or an altered light spectrum inhibited pericarp pigmentation. Utilizing greenhouses constructed with UV blocking plastics, Wu et al. (2017) suggested that exposure of maturing sorghum panicles to UV light may increase biosynthesis of 3-DOAs. Numerous studies in horticultural crops have demonstrated that shading during fruit maturation has a substantial impact on the phytochemical composition. In particular, fruit shading in grape (Azuma et al. 2012) and apple (Ubi et al. 2006) reduced the level of anthocyanin accumulation. Additional studies have established that the light spectral composition to which fruit and vegetables are exposed is critical to their phytochemical composition. Previous studies have associated blue light (Kadomura-Ishikawa et al. 2013) and UV light (hu Dong et al. 1995) with the control of anthocyanin biosynthesis. Thus, the implication of light spectrum or light quantity controlling the development of black pericarp in sorghum parallels numerous

studies with horticultural crops where UV and/or blue light were critical for pigmentation of fruits, vegetables or flowers (Ryan et al. 2002; Koyama et al. 2012).

Given the growing interest in the perceived health benefits of black and red cereals, there exists an opportunity for black grain sorghum to help meet nutritional and health needs of the future. To improve black sorghum grain with high levels of 3-DOAs, it is critical to understand the environmental and cellular signals controlling the accumulation of these compounds in pericarp tissue. Herein, we identified the critical light spectrum that leads to the accumulation of 3-DOAs in the pericarp of black sorghum. Additionally, an exploratory transcriptome analysis of developing pericarp tissue provides insight into the molecular events signaling 3-DOA biosynthesis in black pericarp. Based on the unique upregulation of a suite of genes in black pericarp, we postulate that black sorghum grain tissue is more reactive to UVB-induced cellular stress, which triggers the production of 3-DOAs. We propose that 3-DOAs likely provide epidermal protection to the developing grain against UV light and may participate in the detoxification of damaging free radicals.

Materials and Methods

Plant Material and Growth Room Conditions

Sorghum genotypes, BTx378 (Stephens and Karper 1965) and RTx3362 (Rooney et al. 2013) were grown in a Conviron® BDW Growth Room (Conviron Products of America, Pembina, ND) under controlled conditions with a temperature range of 22 ± 2 °C (night) to 31 ± 2 °C (day), daylength of 14 hours, and 50% (day) and 20% (night) relative humidity. All plants were exposed to 400 $\mu E/m^2s$ of photosynthetically active

light (VIS) using metal halide and incandescent lights. Growth room temperatures, daylength, relative humidity and light intensity settings were chosen to simulate standard field conditions in College Station, TX, an environment in which the black phenotype has been observed to be fully penetrant. Plants were grown in 3-gallon pots containing Sunshine REPS soil mix (Sun Grow Horticulture, Agawam, MA). Each pot contained soil premixed with 18g of Osmocote (16% N, 3.5% P and 10% K), 15 g gypsum, 15 g dolomite and 5 g micromix (6% Ca, 3% Mg, 12% S, 0.1% B, 1% Cu, 17% Fe, 2.5% Mn, 0.05% Mo and 1% Zn). Plants were watered as needed using water-soluble Peters 20N–8.8P–16.6K fertilizer (0.5 g/L; Scotts, Inc., Marysville, Ohio) dissolved in reverse osmosis water at each irrigation. To provide supplemental UV radiation, the growth room was equipped with four florescent fixtures (1.22 m) each containing two UVA and UVB emitting bulbs (Universal UV – FR32T8 32-50 W; Solacure, Browns Summit, NC). The fixtures were mounted and manually adjusted to provide uniform exposure of each panicle to UV radiation. As panicles reached full anthesis, UV lights were programed to expose plants to UV light for 12 h daily starting 2 h after the start of the VIS light cycle. These light conditions were monitored using a photometer (Solar Light Co. Inc., Philadelphia, PA) and maintained throughout grain development until physiological maturity (approximately 35 days after full anthesis).

Studies were conducted to ascertain those regions of the light spectrum (VIS, UVA, UVB) that are involved in the development of the black pericarp phenotype and the associated biosynthesis of 3-DOAs in pericarp. A series of shields (cladding materials) were positioned to block panicles from exposure to specific wavelengths of

light (Table 1). Those shields included; UV-filtering Plexiglas® acrylic sheet OP3/UF-5 to block UVA and UVB (ePlastics, San Diego, CA), Mylar® polyester film to block UVB (U.S. Plastics, Lima, OH), and acrylic-latex foam coated blackout lining fabric to block all light reaching the panicle (Tri-Tex Enterprises, Inc., Dallas, TX). For plant treatments exposed to filtered light spectra, rectangular shields were suspended over individual panicles using two bamboo stakes (1 cm diameter) inserted into the pot and oriented parallel to each plant stem. Shields covered the entirety of the panicles in all directions and light conditions surrounding and underneath the shields were monitored as described above. To reduce humidity around panicles shielded with Plexiglas and Mylar, fans were positioned to circulate the air surrounding the panicles. For each light regime, three randomly positioned plants of genotype RTx3362 (black pericarp) were assigned per each light treatment (Table 1). For comparative purposes, three randomly positioned plants of genotype BTx378 (red pericarp) were also exposed to VIS light supplemented with UVA and UVB (Table 1). A visual record of seed/pericarp colorization was documented photographically at day 10 (T10) and day 17 (T17) after the initiation of light spectral treatments, and grain were again photographed at physiological maturity. Sampling time points were chosen based upon visual ratings of pericarp coloration; T10 corresponded with the initial phase of pericarp pigmentation (~25% of pigmentation of mature seed) and T17 corresponded with ~90% pericarp pigmentation of the mature grain. To compare pericarp pigmentation and 3-DOA accumulation under field conditions, genotype RTx3362 was planted in 5.48 m rows in College Station, TX on March 15, 2019 and grown using standard agronomic practices

for grain sorghum. Field-grown grain was harvested at physiological maturity (approximately 35 days after full anthesis) and analyzed for 3-DOA accumulation as detailed below.

Table 1. Light transmissivity of enclosures to block sorghum panicles from specific wavelengths of light. Reprinted from Fedenia et al. 2020.

Genotype	**Light Treatment**	**Cladding Material**	**UVA** mW/cm^2	**UVB** µW/cm^2	**PAR** µE/m^{2}s
RTx3362 (Black)	No Light	Blackout Cloth	0.0	0.0	0.0
	VIS	Plexiglass	0.016	0.0	400
	VIS + UVA	Mylar	3.25	1.5	400
	VIS + UVA + UVB	None	4.9	10	450
BTx378 (Red)	VIS + UVA + UVB	None	4.9	10	450

Sample Preparation, Extraction, and Ultra-Performance Liquid Chromatography with Photodiode Array Detection (UPLC-PDA) of 3-DOAs

Immature grain (~100 seed per plant at each time point) was harvested from panicles at days 10 (T10) and 17 (T17) after the initiation of light regimes. Pericarp tissue was mechanically excised using a razor blade or scalpel (while minimizing contamination with endosperm tissue) and immediately submerged in liquid nitrogen (LN_2) prior to being pulverized with a mortar and pestle under LN_2. For 3-DOA quantification, ground pericarp tissue (~0.1 g) and whole mature grain (~8 g) was freeze-dried (Labconco, Kansas City, MO) for 48 h and stored at room temperature in an air-tight container containing DrieRite® (Carolina Biological Supply Co., Burlington, NC) until further analysis. Prior to 3-DOA extraction, the mature grain was ground using an IKA® Tube mill control (IKA-Werke, Staufen, Germany) for 3 min at 20,000 rpm.

Extraction of 3-DOAs was conducted according to previously established methods (Dykes et al. 2009; Dykes et al. 2013) with some modifications. In brief, freeze-dried ground pericarp (0.02 g) and grain samples (1 g) were extracted in 600 µL and 10 mL of 1% HCl in methanol (v/v), respectively, for 2 h in a reciprocal shaker (Eberbach, Belleville, MI) set at low speed. The extracts were centrifuged in a Sorvall Legend centrifuge (Thermo Fisher Scientific, Hanover Park, IL) at 4,000 *g* for 8 min and then decanted. Each extract was filtered using a 0.2-µm nylon membrane filter (Whatman Inc., Maidstone, UK) prior to UPLC-PDA analysis. All reagents for extraction and UPLC-PDA analysis were high performance liquid chromatography (HPLC) or analytical grade. Extracts were analyzed on a Nexera-i LC-2040C 3D liquid chromatograph equipped with a photodiode array detector (Shimadzu Scientific Instruments, Kyoto, Japan). The 3-DOAs were separated using a Kinetex C18 column (150 x 2.1 mm, 2.6) (Phenomenex, Torrance, CA). Column temperature was maintained at 40° C. Injection volume was 1 µL. The mobile phase consisted of 2% formic acid in water (v/v) (Solvent A) and acetonitrile (Solvent B). Solvent flow rate was 0.4 mL/min. The 3-DOAs were separated using the following gradient: 0-11 min, 12-25% B; 11-15 min, 25-75% B; 15-18.5 min, 75% B. All 3-DOAs were detected at 485 nm. Data was collected and processed using the LabSolutions (ver. 5.73) software (Shimadzu Scientific Instruments, Kyoto, Japan). The 3-DOAs were identified based on the retention times of commercial standards and UV-Vis spectra and were quantified as described by Dykes et al. (2009 & 2013). The commercial standards including luteolinidin chloride, apigeninidin chloride, and 7-methoxy-apigeninidin chloride were

obtained from ChromaDex (Irvine, CA). The 3-DOA concentrations (μg/g) were expressed as means ± standard deviation from three biological replicates with two technical replicates per biological replicate. One-way analysis of variance (ANOVA) was performed using JMP®, Version 14 (SAS Institute Inc., Cary, NC, 2019) for all comparisons. In all evaluations, all effects were considered fixed except for replication. The respective models for 3-DOA accumulation were A) light treatments conducted on developing pericarp of genotype RTx3362 at T10 and T17 (y = light treatment + timepoint + rep + (timepoint x light treatment) + error); B) developing pericarp in genotypes RTx3362 and BTx378 exposed to VIS + UVA + UVB at T10 and T17 (y = genotype + timepoint + rep + (timepoint x genotype) + error); C) light treatments conducted on genotype RTx3362 at grain maturity (y = light treatment + rep + error); D) mature grain from genotypes RTx3362 and BTx378 exposed to VIS + UVA + UVB (y = genotype + rep + error); E) chamber-grown RTx3362 exposed to treatment VIS + UVA + UVB and field grown RTx3362 (y = environment + error). Effects tables were used to present significant terms in each model. The Tukey-Kramer (Tukey HSD) test was conducted on the aforementioned comparisons to determine significant differences ($p < 0.05$) of the means.

Pericarp Extraction and Sample Preparation for RNA-seq Analysis

For transcriptomic analysis of pericarp, a select timepoint (T10) during grain development was chosen. Rachis branches were removed from panicles exposed to the different light regimes, placed on ice, and each pericarp was mechanically excised (while scraping away adhering endosperm) and immediately submerged in LN_2. Pericarp

from ~100 grain per each treatment were then pulverized to a fine powder with a mortar and pestle under LN_2. Replicated samples (~0.1 g) were placed in 1.5mL Eppendorf microcentrifuge tubes and stored at -80 ºC prior to RNA extraction and RNA-sequencing analysis (RNA-seq).

RNA Isolation, Illumina Library Construction, and Transcriptome Sequencing

Total RNA was isolated using the PureLink™ Plant RNA Reagent System (Thermo Fisher®, Invitrogen, Carlsbad, CA) according to the manufacturer's manual with minor modifications including an overnight precipitation in isopropyl alcohol followed by increased centrifugation time of 25 min. RNA-enriched nucleic acid samples were subsequently treated with the Turbo DNA-free Kit (Ambion®, Austin, TX) for removal of DNA contamination. RNA was analyzed for purity at 260 nm and 280 nm on a DeNovix DS-11 spectrophotometer (DeNovix Inc., Wilmington, DE) and for integrity on the Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA) prior to Illumina template library preparation and sequencing.

Library construction with Illumina's TruSeq Stranded Total RNA Library preparation kit (Illumina Inc. San Diego, CA) and sequencing was performed by AgriLife Genomics and Bioinformatics Services (Texas A&M University) and Novogene (Sacramento, CA). Paired-end (2 x 150bp) sequencing of the TruSeq libraries was performed on the NovaSeq 6000 (Illumina Inc.) for six biological replications per treatment of the black-pericarp genotype RTx3362 and two biological replications for the treatment of the red-pericarp genotype BTx378. Sequence cluster identification, quality pre-filtering, base calling and uncertainty assessment was conducted in real time

using Illumina's HCS 2.2.58 and RTA 1.18.64 software with default parameter settings. The resulting reads were analyzed for quality and quantity using FastQC Software v.0.11.5. (Andrews 2010). Trimmomatic (Bolger et al. 2014) was used for removal of primers, adapters, and other contaminate sequences from the fastq files with default parameter settings.

Trimmed fastq files for each sample were imported into the CLC Genomics Workbench version 12.0.1 (Qiagen, Valencia, CA) and mapped to the *Sorghum bicolor* BTx623 reference genome (Sbicolor_454 v3.0.1, www.phytozome.jgi.doe.gov). The RNA-seq Analysis tool was used with default parameters for stranded libraries to produce gene expression (GE) and transcript expression (TE) track files. Analysis of the GE tracks was performed using the RNA-seq Statistics Workflow Differential Expression for RNA-seq multifactorial tool (Robinson and Oshlack 2010) with default parameters for stranded libraries in the CLC Genomics Workbench between treatments and Pearson's Correlations were examined between biological replications. For examination of gene expression, transcript abundance of each gene was calculated as reads per kilobase of transcript per million mapped reads (RPKM) in addition to transcripts per million (TPM) based on standard parameters within the CLC Genomics Workbench. RPKM values were used to determine differential expression using the Differential Expression for RNA-seq tool within the CLC Workflow while TPM values were used for heatmap graphics. The confounding variables biological experimental replication, date of extraction and sequencing provider, were controlled for during the analysis. Differentially expressed genes (DEGs) were determined by considering a false

discovery rate (FDR) of $p \leq 0.05$ and a fold change cut off of $\log_2 \geq 1$ or ≤ -1 as limiting conditions. Biological function predictions were performed on DEGs using the Gene Ontology (GO) tool from PlantTFDB 4.0 (Jin et al. 2016) according to the *Sorghum bicolor* annotation (Sbicolor_454 v3.0.1). Significant GO terms were found using an FDR of $p \leq 0.05$.

Results

Phenotype Development Under Select Light Regimes in a Controlled-Environment Growth Room

In the absence of light, grain from genotype RTx3362 (black pericarp) was a creamy-white color at T10 and T17, eventually developing a medium brown coloration at grain maturity (Figure 2). Under the VIS light regime (400 $\mu E/m^2s$) or the light regime of VIS light supplemented with UVA (3.25 mW/cm^2), RTx3362 pericarp developed a medium green color at T10, light brown coloration at T17, and a medium to dark brown color in mature grain (Figure 2). When UVB (10 $\mu W/cm^2$) was included in the action spectra, grain of RTx3362 began to develop small black patches at T10 and progressed to nearly 50% darkening of the grain at T17 (Figure 2). Including UVB in the action spectra resulted in RTx3362 grain being uniformly black at physiological maturity, except for a few small patches that were ruby-red and likely the result of shading of the pericarp by their position within the rachis branches (Figure 2). Under the same light regime (VIS supplemented with UVA and UVB), red grain genotype BTx378 developed small red-orange patches at T10 and achieved a near uniform red color at T17 before slightly darkening at grain maturity (Figure 2). The pigmentation of

the pericarp of RTx3362 under VIS light supplemented with UVA and UVB was similar to that reported under normal agronomic field conditions where the black pericarp phenotype has proven to be fully penetrant (Pfeiffer and Rooney 2016).

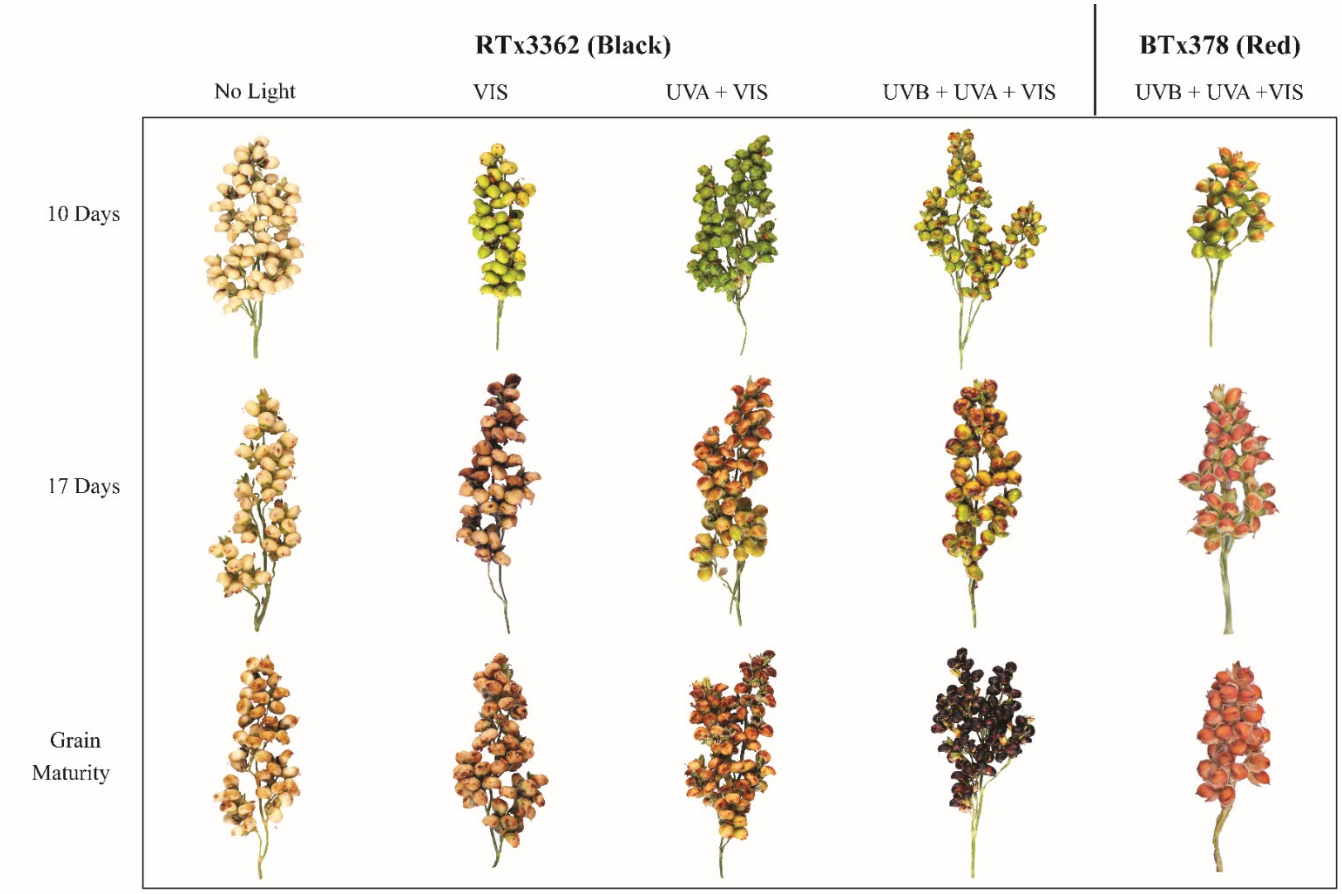


Figure 2. Color development of sorghum rachis branches subjected to varying light regimes at 10 and 17 days after exposure during grain filling and at final grain maturity. Reprinted from Fedenia et al. 2020.

Utilizing the conditions previously established for sorghum grain (Dykes et al. 2009; Dykes et al. 2013), four major 3-DOAs were detected in excised pericarp of developing grain: luteolinidin, apigeninidin, 5-methoxy-luteolinidin, and 7-methoxy-apigeninidin (Figure S1). All statistical models containing a main effect of timepoint, light treatment, and/or genotype and an interaction term of (timepoint x light treatment) or (timepoint x genotype) were significant (Table 2A-D; $p < 0.05$) for all 3-DOAs. The rep effect was not significant across all models for all 3-DOAs (Table 2A-D; $p < 0.05$). The concentration of pericarp 3-DOAs was highest in RTx3362 at T17 when exposed to VIS light supplemented with UVA and UVB with an average total 3-DOA accumulation of 2063 ± 173 µg flavonoid/g tissue (Table 3A; Figure S2A). While all 3-DOAs increased under VIS light supplemented with UVA and UVB light, luteolinidin accumulated to very high levels (1139 ± 75 µg/g) at T17 in pericarp from RTx3362 (Table 3A; Figure S2A). Under blackout conditions or exposure to VIS light only, 3-DOAs were detected in RTx3362 pericarp, although quantities did not exceed 61 ± 15 µg/g for any 3-DOA (Table 3A; Figure S2A). While greater coloration was visually observed during grain development under VIS light compared to blackout conditions in RTx3362 (Figure 2), the apparent pigmentation was not associated with greater accumulation of 3-DOAs (Table 3A; Figure S2A). Exposure of RTx3362 to VIS light plus UVA light resulted in greater accumulation of 3-DOAs at T17 compared to exposure to VIS light only or no light, however, the amount of 3-DOAs in each of these treatments was significantly less than when RTx3362 was exposed to VIS with UVA and UVB (Table 3A; Figure S2A). When developing pericarp of red grain genotype

BTx378 was exposed to light containing VIS with UVA and UVB, the total 3-DOA accumulation was less than 16 ± 0.9 µg/g at T17 (Table 3B; Figure S2A), which was less than the total level of 3-DOAs that accumulated in pericarp of RTx3362 under blackout conditions (61 ± 15 µg/g, Table 3A; Figure S2A).

Table 2. Analysis of variance for Ultra-Performance Liquid Chromatography with Photodiode Array Detection quantification of 3-DOAs in developing and mature *Sorghum bicolor* grain for five separate models (A-E). Reprinted from Fedenia et al. 2020.

			LUT[a]			AP[b]			5-MeO-LUT[c]			7-MeO-AP[d]			Total 3-DOA		
	variable	**DF[e]**	**SS[f]**	**FR[g]**	**P > F[h]**	**SS**	**FR**	**P > F**	**SS**	**FR**	**P > F**	**SS**	**FR**	**P > F**	**SS**	**FR**	**P > F**
A	time	1	518827	239	<.0001	40407	84	<.0001	117160	67	<.0001	8030	50	<.0001	1831174	172	<.0001
	rep	2	677	0	0.8575	497	1	0.6114	4043	1	0.3521	169	1	0.6053	3218	0	0.8618
	light treat.[i]	3	1085335	167	<.0001	71326	49	<.0001	180894	34	<.0001	15315	32	<.0001	3441271	107	<.0001
	time*light treat.	3	1043291	160	<.0001	71326	49	<.0001	176972	34	<.0001	15026	31	<.0001	3342225	104	<.0001
B	time	1	639103	453	0.0002	44770	32	0.0111	113330	1028	<.0001	9333	20	0.0203	2085943	281	0.0005
	rep	2	1997	1	0.56	1448	1	0.6439	360	2	0.3315	502	1	0.6267	9759	1	0.5793
	genotype	1	639426	453	0.0002	42974	30	0.0118	117181	1063	<.0001	9289	20	0.0204	2089880	282	0.0005
	time*genotype	1	637391	452	0.0002	44858	32	0.0111	114728	1041	<.0001	9906	22	0.0187	2097886	283	0.0005
C	rep	5	43437	1	0.6178	992	1	0.6955	1314	1	0.7195	144	1	0.3792	66583	1	0.6559
	light treat.	3	276742	8	0.0028	5789	6	0.008	12059	9	0.0016	432	6	0.0088	525952	9	0.0016
D	genotype	1	209822	8	0.0436	6075	21	0.0104	10411	15	0.0188	465	11	0.03	435061	12	0.0272
	rep	5	109032	1	0.5643	2212	2	0.356	3999	1	0.4692	260	1	0.4361	182381	1	0.5259
E	environment	1	648	0.2	0.6398	4116	7	0.0376	83	0.2	0.6476	596	6	0.0529	15154	1.4	0.2793

A) developing RTx3362 pericarp exposed to all light treatments (No Light, VIS, VIS + UVA, VIS + UVA + UVB) at two timepoints (10 and 17 days post anthesis); **B**) developing RTx3362 and BTx378 pericarp exposed to the same light treatment (VIS + UVA + UVB) at two timepoints; **C**) RTx3362 grain exposed to all light treatments at maturity; **D**) RTx3362 and BTx378 grain exposed to the same light treatment at maturity; **E**) chamber-grown RTx3362 exposed to treatment VIS + UVA + UVB and field-grown RTx3362; [a]luteolinidin; [b]apigeninidin; [c]5-methoxy-luteolinidin; [d]7-methoxy-apigeninidin; [e]degrees of freedom; [f]sum of squares; [g]F Ratio; [h]Prob > F; [i]light treatment

Table 3. Connecting letter reports according to the all pairs Tukey-Kramer method for Ultra-Performance Liquid Chromatography with Photodiode Array Detection quantification of 3-DOAs in developing and mature *Sorghum bicolor* grain for five separate models (A-E). Reprinted from Fedenia et al. 2020.

	Genotype	Timepoint	Treatment	LUT[a]		AP[b]		5-MeO-LUT[c]		7-MeO-AP[d]		Total 3-DOA	
					LSM[e]		LSM		LSM		LSM		LSM
A	RTx3362	T17	VIS + UVA + UVB	A	1139	A	299	A	485	A	140	A	2063
		T10	VIS + UVA + UVB	C	18	B	0.0	C	14	B	3.5	C	36
		T17	VIS + UVA	B	213	B	56	B	156	B	22	B	447
		T10	VIS + UVA	C	11	B	0.0	C	9.4	B	3.4	C	24
		T17	VIS	C	13	B	8.7	B C	13	B	4.7	C	40
		T10	VIS	C	7.5	B	0.0	C	9.2	B	3.0	C	20
		T17	No light	C	14	B	9.1	B C	26	B	10	C	61
		T10	No light	C	5.8	B	0.0	C	14	B	2.5	C	22
B	RTx3362	T17	VIS + UVA + UVB	A	1139	A	299	A	485	A	140	A	2063
		T10	VIS + UVA + UVB	B	18	B	0.0	B	14	B	3.5	B	36
	BTx378	T17	VIS + UVA + UVB	B	8.6	B	3.0	B	3.7	B	1.7	B	16
		T10	VIS + UVA + UVB	B	7.8	B	3.2	B	5.1	B	3.7	B	20
C	RTx3362	mature grain	VIS + UVA + UVB	A	276	A	46	A	60	A	12	A	394
		mature grain	VIS + UVA	B	45	A B	26	B	8.3	B	3.7	B	83
		mature grain	VIS	B	36	B	16	B	6.8	B	2.7	B	62
		mature grain	No light	B	8.2	B	2.9	B	11	B	3.0	B	25
D	RTx3362	mature grain	VIS + UVA + UVB	A	276	A	46	A	60	A	12	A	394
	BTx378	mature grain	VIS + UVA + UVB	B	2.7	B	2.2	B	1.8	B	0.9	B	7.6
E	RTx3362	mature grain	field-grown	A	155	A	97	A	49	A	36	A	337
		mature grain	chamber-grown	A	134	B	45	A	42	A	16	A	236

A) developing RTx3362 pericarp exposed to all light treatments (No Light, VIS, VIS + UVA, VIS + UVA + UVB) at two timepoints (10 and 17 days post anthesis); **B**) developing RTx3362 and BTx378 pericarp exposed to the same light treatment (VIS + UVA + UVB) at two timepoints; **C**) RTx3362 grain exposed to all light treatments at maturity; D) RTx3362 and BTx378 grain exposed to the same light treatment at maturity; E) chamber-grown RTx3362 exposed to treatment VIS + UVA + UVB and field-grown RTx3362; [a]luteolinidin**;** [b]apigeninidin; [c]5-methoxy-luteolinidin; [d]7-methoxy-apigeninidin; [e]least square mean (µg/g); means not connected by same letter are significantly different ($p < 0.05$).

To determine whether the effect of the light spectrum regimes persisted after physiological maturation, 3-DOA levels were quantified in mature grain. Consistent with the trend seen in developing pericarp tissue, the highest accumulation of 3-DOAs in mature grain occurred in RTx3362 exposed to VIS light supplemented with UVA and UVB (Table 3C-D; Figure S2B). The quantitative differences seen in 3-DOA levels in immature pericarp samples vs. mature grain reflect the methodology required to obtain a sample enriched in pericarp from mature grain, which results in a pericarp-enriched fraction contaminated with endosperm. Nevertheless, the results obtained during grain development and in mature grain support the assertion that UVB is critical for the accumulation of 3-DOAs and full penetrance of the black pericarp phenotype.

3-DOA Accumulation in Black Sorghum under Field Conditions versus Growth Room Conditions

In this study, efforts were made to establish growth room conditions (daylength, temperature, humidity, lighting conditions) that would allow for extrapolating our findings to field-grown sorghum with respect to light quality regimes and their effects on 3-DOA accumulation. To do so, we quantified 3-DOA levels in mature grain of RTx3362 grown under field conditions where the black pericarp phenotype was fully penetrant. At solar noon during grain filling, average levels of VIS, UVA, and UVB intensity were 1350 $\mu E/m^2s$, 3.7 mW/cm^2, and 8.7 $\mu W/cm^2$, respectively. The fluences of UVA and UVB established in the growth room were similar to values at solar noon while the levels of VIS were 3- to 3.5-fold lower in the growth room (Table 1). UPLC-PDA quantification of 3-DOAs in mature field-grown RTx3362 grain revealed that

apigeninidin concentration was statistically greater under field conditions; however, luteolinidin, 5-methoxy-luteolinidin, 7-methoxy-apigeninidin and total 3-DOA concentration were not significantly different in field vs. growth room-grown RTx3362 (Table 2E & Table 3E). While we could not determine why higher levels of apigeninidin accumulated under field conditions, the similar levels of luteolinidin, 5-methoxy-luteolinidin, 7-methoxy-apigeninidin and total 3-DOA concentration in pericarp under growth room and field conditions indicate that the growth room light regimes are a reasonable approximation of field conditions where high levels of 3-DOAs accumulate and the black pericarp phenotype is fully penetrant in RTx3362.

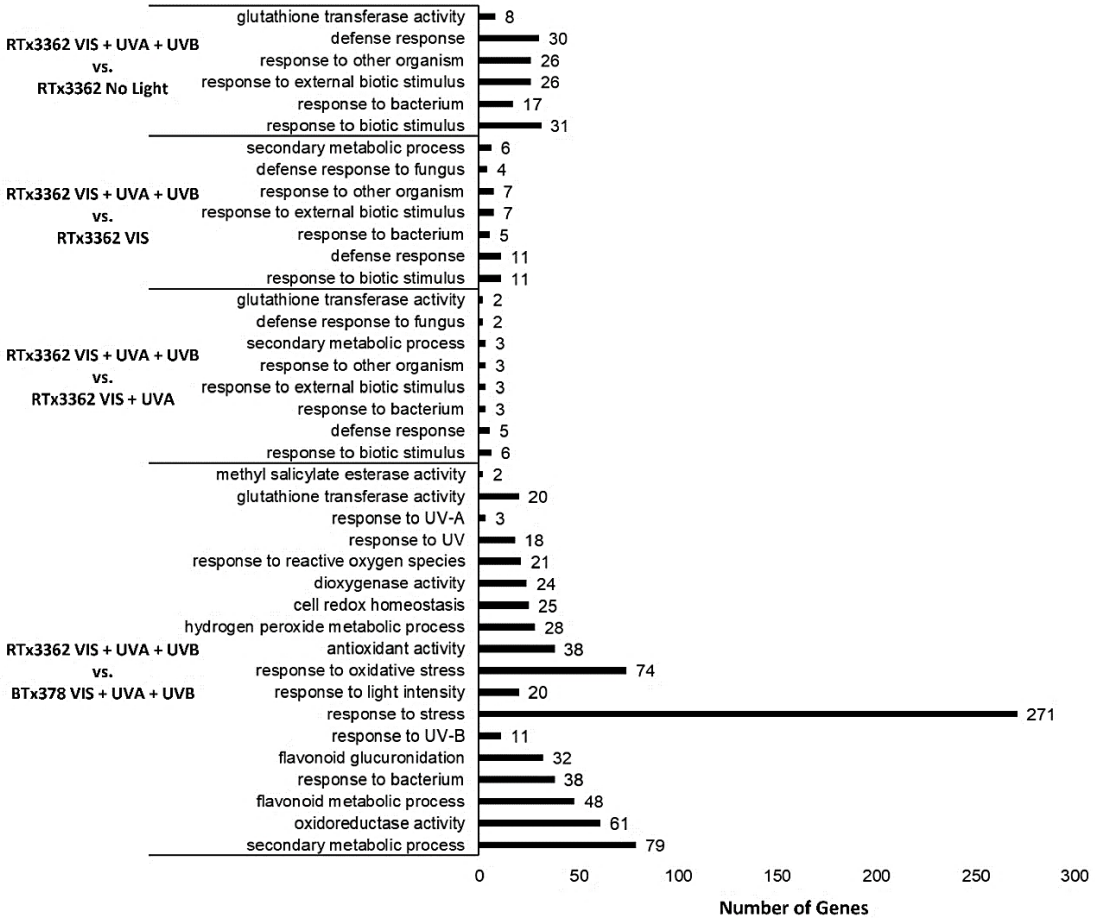
Comparative Transcriptomics of Pericarp Samples under Different Light Regimes

To gain insight into the cascade of cellular events in pericarp leading to the biosynthesis of 3-DOAs, changes in pericarp gene expression were quantified at T10 which was the developmental stage at which pericarp pigmentation was first visually apparent. The series of light spectral regimes to which panicles of BTx378 (red grain) and RTx3362 (black grain) genotypes were exposed permitted seven pairwise comparisons of pericarp gene expression. Utilizing this RNA-seq data, six pairwise comparisons were conducted to quantify changes in pericarp gene expression in genotype RTx3362 under the different light spectral regimes. A seventh comparison of pericarp gene expression between genotypes RTx3362 and BTx378 exposed to VIS with UVA and UVB was also conducted. From this RNA-seq dataset, a total of 1,327 x 10^6 reads were obtained with an average of 51 million reads per sample after adaptor and low-quality paired-end sequences were trimmed and/or removed (Table S1). An average

of 43.7 million clean paired-end reads per sample were aligned to the reference sorghum genome (Sbicolor_454 v3.0.1) using the RNA-seq Analysis tool within the CLC Genomics Workbench (Table S1). After filtering, a total of 9,299 DEGs across all 7 pairwise comparisons were detected; when examining the DEGs within each pairwise comparison, the number of differentially expressed genes ranged from 2 to 6867 (Table S2).

For preliminary characterization of the functional genetic differences underlying the accumulation of 3-DOAs in RTx3362 pericarp, Gene Ontology (GO) enrichment analysis was performed using the DEGs identified in RTx3362 in response to the different light quality regimes (Figure 3). When contrasting gene expression in pericarp from RTx3362 exposed to VIS with UVA and UVB to each of the other light regimes (i.e, no light, VIS, and VIS with UVA), multiple GO terms were found across all pairwise comparisons. Prominent amongst the GO terms highly enriched in DEGs upregulated by UVB in RTx3362 included, but are not limited to, secondary metabolic process (GO:0019748) and defense response (GO:0006952) (Figure 3). In general, the functional characterization of DEGs in RTx3362 pericarp exposed to UV light (in particular, UVB) related to the perception/response of plant tissue to biotic or abiotic stress signals (Figure 3).

Figure 3. Number of differentially expressed genes upregulated in RTx3362 exposed to VIS + UVA + UVB compared to all other treatments categorized by Gene Ontology (GO) functional groups. A subset of significant GO terms are shown (FDR of $p < 0.05$). Reprinted from Fedenia et al. 2020.



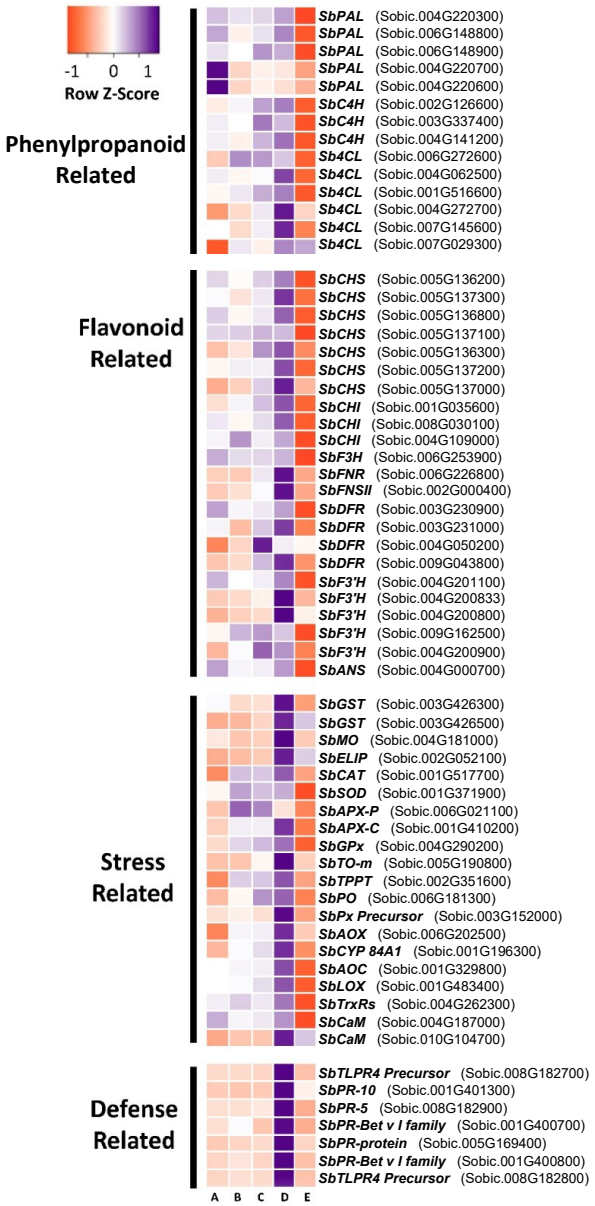
By exposing BTx378 to the same UVB-containing light regime that induced 3-DOA accumulation in RTx3362, we sought to gain an initial understanding of cellular events leading to the genotype-specific accumulation of 3-DOAs in black pericarp sorghum. Examination of GO terms enriched in DEGs up-regulated in RTx3362 compared to BTx378 revealed a number of categories related to cellular stress signaling or response including but not limited to: response to stress (GO:0006950), response to oxidative stress (GO:0006979), oxidoreductase activity (GO:0016491), antioxidant

activity (GO:0016209), hydrogen peroxide metabolic processes (GO:0042743), response to reactive oxygen species (GO:0000302), and secondary metabolic processes (GO:0090487) (Figure 3). While pericarp of BTx378 accumulated minor amounts of 3-DOAs when exposed to VIS with UVA and UVB (Table 2 & Table 3), the present gene expression analysis revealed that BTx378 pericarp does not either perceive and/or respond to UV light as a stress similar to what was observed for RTx3362 pericarp.

Genes Involved in Secondary Metabolism

To examine in more detail the changes in gene expression during pericarp development in response to the various light regimes, we used transcripts per million (TPM) to display the transcript abundance of specific DEGs (fold change cut off of $\log_2 \geq 1$ or ≤ -1; FDR p-value of ≤ 0.05) within the phenylpropanoid and flavonoid biosynthetic pathways and a panel of genes related to cellular response to stress as well as defense related genes (Figure 4).

Figure 4. Heatmap showing transcript abundance of genes related to phenylpropanoid and flavonoid biosynthesis as well as stress and defense response in RTx3362 exposed to No light (A); VIS (B); VIS + UVA (C); VIS + UVA + UVB (D); and BTx378 exposed to VIS + UVA + UVB (E). Sb indicates *Sorghum bicolor.* Reprinted from Fedenia et al. 2020.



Transcript abundance of genes encoding structural proteins upstream and downstream in the flavonoid biosynthetic pathway varied depending on the genotype and the light regime to which the pericarp was exposed. In general, the highest transcript abundance for flavonoid biosynthesis genes was observed in pericarp of RTx3362 exposed to VIS with UVA and UVB light (Figure 4). Pericarp transcripts of multiple genes encoding structural proteins upstream in the phenylpropanoid pathway including PAL, C4H, and 4CL genes coding for proteins upstream in the flavonoid biosynthetic pathway including CHS and CHI and genes encoding proteins downstream in the flavonoid biosynthetic pathway including F3H, FNSII, F3'H, FNR, DFR, and ANS were highly abundant in RTx3362 exposed to VIS with UVA and UVB light (Figure 4). The lowest transcript abundance of nearly all known upstream and downstream flavonoid biosynthetic genes were observed for BTx378 (red pericarp) under the same UVB-containing light regime (Figure 4). There are several notable exceptions that included PAL (Sobic.004G220700; Sobic.004G220600) (upregulated in RTx3362 under blackout conditions) and a DFR (Sobic.004G050200) (upregulated in RTx3362 exposed to VIS with UVA light). With these few exceptions, it appears that the entire flavonoid biosynthetic pathway was upregulated in black pericarp under a light regime containing UVB (Figure 4).

While UV light was required for the elevated transcript abundance of flavonoid biosynthesis genes, RTx3362 pericarp grown in blackout conditions accumulated higher steady state transcript levels for most upstream and downstream flavonoid biosynthetic genes than BTx378 pericarp under the VIS with UVA and UVB light regime (Figure 4).

These results indicate that genotype-specific differences exist between pericarp of BTx378 and RTx3362 in the basal expression of genes in the flavonoid biosynthesis pathway. Moreover, genotype-specific differences exist relating to the UVB-dependent transcriptional activation in RTx3362 pericarp of the flavonoid genes encoding structural proteins and the subsequent accumulation of 3-DOAs. The transcription factor ultraviolet-B receptor 8 (UVR8) is well characterized in the literature for UVB mediated signaling transduction (Tilbrook et al. 2013), however, the gene encoding the homolog of UVR8 in sorghum (Sobic.006G85300 (Ambawat et al. 2013)) was not differentially expressed between any pairwise light regime comparisons. In addition, differential expression of the *Y1* gene (Sobic.001G398100 (Nida et al. 2019)) encoding the MYB transcription factor proposed to regulate both upstream and downstream genes in the flavonoid biosynthetic pathway in sorghum (Boddu et al. 2005), was not observed among any comparison between light regimes in genotype RTx3362, however, this gene was upregulated 4-fold in BTx378 compared to RTx3362 exposed to the same light conditions.

Genes Involved in Stress Response

Transcript abundance for genes frequently associated with stress-related cellular processes were generally most abundant in RTx3362 pericarp exposed to the light regime containing UVB (Figure 4). As was the case for a limited number of flavonoid biosynthetic genes, select stress-related genes showed higher transcript abundance (i.e. APX-P (Sobic.006G021100)) under other light regimes (Figure 4). Nevertheless, TPM for the majority of genes related to stress, secondary metabolism, and defense were

greatest for RTx3362 pericarp developing under the light regime containing UVB (Figure 4).

Among the genes upregulated in RTx3362 under UVB light were those encoding enzymes related to cell detoxification and metabolism of toxic compounds (glutathione S-transferase, catalase, superoxide dismutase, l-ascorbate peroxidase, glutathione peroxidase, tocopherol O-methyltransferase, tocopherol polyprenyltransferase, class III peroxidase precursor, cytochrome P450, monooxygenase), prevention of photooxidative stress (early-light inducible protein), oxidation (lipoxygenase, polyphenol oxidase, alternative oxidase), protection from oxidation (thioredoxin reductase), and biosynthesis of signaling hormones (allene oxide cyclase) (Figure 4). In addition, genes encoding two calmodulin-binding proteins were upregulated in RTx3362 black pericarp exposed to VIS with UVA and UVB vs. all other light regimes (Figure 4). Calcium (Ca^{2+}) acts as a secondary messenger in concert with ROS and increases in cytosolic Ca^{2+} indicate the occurrence of a stress defense response (Görlach et al. 2015).

Genes Involved in Defense Response

Although little is known regarding UVB induction of defense genes, it has been observed in the literature that expression of pathogenesis-related (PR) defense genes are often induced in response to various elicitors including UVB light (Brederode et al. 1991; Kuhn et al. 1984). Transcript abundance for a panel of genes encoding PR proteins showed a strict genotype-specific and UVB-dependent response (Figure 4). In the presence of UVB light, pericarp of RTx3362 accumulated PR transcripts, a response not observed under the light regimes lacking UVB nor in pericarp of genotype BTx378

despite exposure to the same fluence of UVB light. Pairwise comparisons of the different light regimes revealed that genes encoding PR proteins made up over a quarter of the limited number of DEGs found upregulated in RTx3362 black pericarp under VIS with UVA and UVB vs. VIS with UVA providing a clear and novel distinction between gene expression in response to UVB vs. UVA in black sorghum grain.

Discussion

Flavonoids are a diverse collection of secondary metabolites that are widely recognized as the primary determinants of plant color and for providing a visual signal for pollinators in flowers and seed dispersers in ripe fruit. Flavonoids also play critical roles in plants for pathogen defense, male fertility, and protection from a wide array of abiotic stresses (Brederode et al. 1991; Kuhn et al. 1984). There is interest in pigmented grains as useful sources of dietary flavonoids as potential health beneficial metabolites and for their visual appeal to niche consumer groups (Abdel-Aal et al. 2006). Most notable within the pigmented cereals are the varieties of rice that exhibit brown, red or black (purple) pericarp (Abdel-Aal et al. 2006), purple pericarp wheat (Ficco et al. 2016), and red/purple aleurone maize genotypes (Paulsmeyer et al. 2017). Unique amongst the colored grains is black grain sorghum that accumulates high levels of 3-DOAs within the pericarp rather than the accumulation of anthocyanin and/or proanthocyanidin as in corn, rice and wheat (Zhu 2018). Naturally occurring gene duplications in the flavonoid biosynthetic pathway have resulted in diverse grain pigmentation and polyphenol accumulation (Rausher 2006). The resulting phytochemical diversity in grains has fostered detailed investigations into the genetic

basis and environmental signals controlling grain pigmentation (Liu et al. 2013; Rhodes et al. 2014; Xu et al. 2015). In the present study, we utilized a series of cladding materials with different optical properties (e.g., UV transmission) to manipulate the light spectra and thereby gain a better understanding of the environmental signals controlling the genotype-specific biosynthesis of 3-DOAs in the pericarp of black grain sorghum genotype RTx3362 and the cascade of transcriptional activity associated with this unique trait.

The present study revealed that exposure of pericarp from genotype RTx3362 to UVB irradiation during grain development is critical for full penetrance of the black pericarp phenotype and the associated biosynthesis of the four major 3-DOAs. The effect of UVB light on 3-DOA accumulation in black sorghum pericarp parallels the biosynthesis of flavonoids in many fruits and vegetables that are sensitive to the quality of light, with shorter wavelengths of light in the blue or UV region reportedly having the greatest effect on flavonoid biosynthesis (Kadomura-Ishikawa et al. 2013; Koyama et al. 2012; Zoratti et al. 2014). Depending on the variety, anthocyanin and melanin production in the kernel of some wheat genotypes have been correlated with sunlight exposure (Yang and Jie 2007). Reddy et. al (1994) summarized investigations that demonstrated UVB increased plant pigments in certain rice cultivars, however, elevated pigment production stimulated from UVB exposure can significantly reduce plant biomass, height, leaf area, dry weight, net assimilation rate, and overall productivity. UVB radiation is a key environmental signal that initiates a diverse array of metabolic and developmental responses, and the response to UVB is dependent on factors

including the fluence/duration of the UVB treatment and the genotype of the plant (Jenkins 2009). In pericarp of RTx3362, full penetrance of the black grain phenotype was associated with UVB fluences and duration similar to that observed in natural environments where the trait fully develops. Environmental conditions in which the black pericarp phenotype does not fully develop are often associated with lower fluences of UVB light. In particular, the present use of Mylar cladding material, which prevented the black pigmentation of pericarp, transmitted a low fluence (1-2 $\mu W/cm^2$) of longer wavelength UVB light (300-315 nm). In addition, the lack of full penetrance of the black pericarp phenotype in field environments where the total fluence of UV light is reduced (due to daylength or UV index) supports the assertion that the black grain phenotype in sorghum requires a fluence of UVB greater than those associated with photomorphogenic responses (Tilbrook et al. 2013; Jenkins 2009).

In sorghum genotypes of purple plant color, 3-DOAs accumulate in localized lesions of vegetative tissues in response to wounding or pathogen attack (Lo et al. 1999). This stress-induced response in vegetative tissue results in the activation of specific genes and gene-family members in the flavonoid biosynthetic pathway and the rapid accumulation of high levels of 3-DOA phytoalexins near the site of attack (Nielsen et al. 2004). In the present study, a comparative transcriptome analysis of pericarp from black- and red-grain sorghum genotypes facilitated an initial characterization of cellular events leading to the accumulation of 3-DOAs in pericarp in response to UVB light. In UVB irradiated pericarp of black grain genotype RTx3362, transcripts of upstream and

downstream flavonoid biosynthetic genes were abundant, and multiple members of specific gene families were upregulated in response to UVB (Figure 4).

The activation of the flavonoid biosynthetic pathway by UVB light was genotype-specific as transcripts for these genes were relatively unabundant for UVB irradiated pericarp of red-grain genotype BTx378. It should be noted that, while the majority of genes in the biosynthetic pathway for secondary metabolites are well-characterized, the final enzymatic step(s) specific to the biosynthesis of 3-DOAs have yet to be conclusively identified (Liu et al. 2010). Various studies have proposed that the final structural enzymes in 3-DOA biosynthesis may include the following; anthocyanidin synthase (Turnbull et al. 2000), a recently proposed fast-acting flavanone-4-reductase (Kawahigashi et al. 2016a), and 5-O-glucosyltransferase (Nakatsuka and Nishihara 2010). While beyond the scope of the present study, the identification of the final enzymatic steps in 3-DOA biosynthesis and the gene family members encoding these enzymes will require additional molecular and biochemical analyses that are essential to our understanding of this critical pathway.

Plant responses to UV light are mediated by both non-specific signaling pathways that involve reactive oxygen species (ROS) and wound/defense signaling pathways, and UVB specific pathways that regulate photomorphogenic responses to low fluence UVB irradiation (Jenkins 2009). Low fluence UVB-mediated photomorphogenesis responses, which are controlled by the photoreceptor UVR8, include the promotion of flavonoid biosynthesis, morphological changes (stem extension, cotyledon opening), and the regulation of expression of a range of structural

and regulatory gene networks (Jenkins 2009). Higher fluence UVB responses are mediated by nonspecific signaling pathways and are accompanied by activation of pathways to protect plants from oxidative stress (Brosche and Strid 2003). In pericarp of black sorghum, the activation of the flavonoid biosynthetic pathway by UVB was associated with the expression of genes encoding stress-related compounds and enzyme systems that protect against oxidative stress (Figure 4). Some of the antioxidant agents and free-radical scavenging systems activated within UVB-irradiated RTx3362 included the genes encoding glutathione S-transferase, catalase, superoxide dismutase, l-ascorbate peroxidase, and tocopherol O-methyltransferase (Figure 4). Flavonoids (including 3-DOAs) in black pericarp of RTx3362 also have the potential to act as effective secondary protection mechanisms against UVB light due to their reported highly effective ROS scavenging activity as well as direct absorption of UVB light (i.e., sunscreen) (Saewan and Jimtaisong 2013). While red-grain sorghum genotype BTx378 was irradiated with the same fluence of UVA and UVB light, transcripts for genes encoding antioxidant and free-radical scavenging systems were relatively unabundant in red pericarp indicating that black- and red-pericarp genotypes differ markedly in their tolerance or responsiveness to UVB (Figure 4).

The nonspecific signaling pathway activated by UVB light in black pericarp also included biotic stress response genes. The present observations are consistent with reports that plant defense genes are responsive to environmental stressors beyond pathogen attack (Casañal et al. 2013). In tobacco, UVB was proven to be a fluence-dependent inducer of a PR-gene and this response was mediated by ROS (Green and

Fluhr' 1995). In strawberry, *PR-10* was shown to control preferential flavonoid accumulation by binding to metabolic intermediates (Casañal et al. 2013. Moreover, one study found a strong correlation between the rapid accumulation of *PR-10*, *CHS*, and 3-DOA accumulation in pathogen-challenged sorghum leaves (Lo et al. 1999). Consistent with these studies, the activation of biotic-defense related genes that occurred in concert with genotype-specific activation of flavonoid biosynthetic and stress related genes substantiates the assertion that UVB controls the accumulation of 3-DOAs and the black pericarp phenotype in RTx3362 by a nonspecific stress signaling pathway (Figure 4).

While the nature of the response to UVB light is dependent on fluence, duration, and the level of adaptation or acclimation of plant tissue, genotypic differences within a plant species exist with respect to their tolerance and response to UV irradiation (Jenkins 2009; Lo et al. 1999; Casati and Walbot 2005; Frohnmeyer and Staiger 2003; Ulm and Nagy 2005). In addition, the tolerance or response to UVB light within a given genotype is often tissue-specific especially as it pertains to vegetative vs. reproductive/floral tissues (Burchard et al. 2000; Samanta et al. 2011; Ormrod et al. 1995). In sorghum, genotypes of red or purple secondary plant color accumulate high levels of 3-DOAs (apigeninidin and luteolinidin) in vegetative tissues in response to wounding or pathogen attack (Kawahigashi et al. 2016a; Liu et al. 2010). Vegetative pigmentation is controlled by two dominant epistatic genes, *P* and *Q,* and possession of the dominant allele for both genes leads to the development of 3-DOAs in sorghum leaf tissue (Valencia and Rooney 2009). By contrast, pericarp pigmentation has been proposed to be controlled in sorghum by *Y1*, a MYB transcription factor that is a homolog of maize *Pericarp color1*

(*P1*) (Boddu et al. 2005). In high altitude genotypes of maize, acclimation to higher UVB fluences results in the biosynthesis of the flavone maysin in both silks and leaves (Casati and Walbot 2005). The accumulation of these flavones was associated with *P1,* an R2R3 MYB transcription factor, being highly expressed in leaves and silks of landraces adapted to higher fluences of UVB light (Casati and Walbot 2005). Based on the evolutionary relationship of maize and sorghum, *Y1* represents an obvious candidate regulatory gene controlling 3-DOA accumulation in black pericarp in response to UVB; however, examination of transcript levels of *Y1* (Sobic.001G398100) did not indicate that *Y1* was differentially expressed in RTx3362 exposed to different light regimes. *Y1* was upregulated 4-fold in the pericarp of BTx378 compared to RTx3362 under VIS with UVA and UVB. BTx378 and RTx3362 both harbor a dominant *Y1* allele (Pfeiffer and Rooney 2016), indicating that additional regulatory factors likely control the accumulation of 3-DOAs in RTx3362 in response to UVB. Regulation of 3-DOA biosynthesis in black pericarp appears more complex than control by the R2R3 MYB transcription factor encoded by *Y1* and activation by UVB since basal expression of the flavonoid pathway (and accumulation of 3-DOAs) in dark-grown RTx3362 pericarp was greater than that in red-pericarp genotype BTx378 under the full light spectrum (VIS with UVA and UVB) (Table 3; Figure 4; Figure S2). The complexity of the black pericarp trait is supported by heritability studies that estimate that 2 to 12 genes (some alleles are recessive) are controlling the trait (Pfeiffer and Rooney 2016). Investigation of genomic differences underlying the tissue-specific accumulation of 3-DOAs in

RTx3362 are presently underway to obtain a more complete understanding of the molecular mechanisms leading to this trait.

Black pericarp with the associated accumulation of 3-DOAs is a rare and valued phenotype whose origin can be traced to a single plant introduction (‘Black Shawaya,’ PI673843) from Sudan. The underlying genes controlling the inheritance of the black pericarp trait are complex, although the present results implicate heritable, genotype- and tissue-specific differences in the sensitivity of black and red sorghum to higher fluences of UVB. These higher UVB fluences activate flavonoid biosynthesis and other stress-responsive pathways specifically in pericarp of black sorghum, and these cellular responses collectively represent protective mechanisms to UVB light and the associated production of reactive oxygen species. When combined with a detailed understanding of the genetic basis of the black pericarp phenotype, the present understanding of the environmental signals and nature of the cellular response should allow breeders to more effectively develop new black pericarp genotypes and should permit the selection of growth environments to maximize the expression of this trait.

CHAPTER III

IDENTIFICATION OF QUANTITATIVE TRAIT LOCI ASSOCIATED WITH 3-DOA PRODUCTION IN BLACK GRAIN SORGHUM

Synopsis

Sorghum [*Sorghum bicolor* (L.) Moench] is a major cereal crop bred for feed, silage, and human consumption across nearly 41 million ha worldwide (FAOSTAT 2019). Amongst the grain-type sorghums, pericarp color is typically red, yellow, or white (Smith and Frederiksen 2000) however, a rare black grain trait was discovered in the Sudanese cultivar 'Black Shawaya' (Hahn et al. 1984). Black pigmentation of sorghum grain is associated with the high accumulation of rare polyphenols, 3-deoxyanthocyanidins (3-DOAs) (Awika et al. 2004). Valued for their health promoting properties, there is an increasing interest from the food and health industries to utilize 3-DOAs as nutraceutical additives (Xiong et al. 2019). The selection and advancement of this important sorghum grain phenotype could be aided by a better understanding of the genetic loci and genes that control black pigmentation and the associated accumulation of 3-DOAs.

Multiple trait mapping methodologies were used to identify the genomic loci associated with the black sorghum phenotype that included genetic linkage map construction with quantitative trait loci (QTL) analysis and bulked segregant analysis (BSA) of opposing phenotypic extremes for 3-DOA accumulation. The first mapping method utilized replicated field trails of a recombinant inbred line (RIL) population of

224 F_5 individuals generated from a controlled cross between BTx378 (red pericarp) and RTx3362 (black pericarp). Genotyping-by-sequencing (GBS) was implemented on the population to construct genetic linkage maps and conduct QTL analysis (F_5 QTL_{GBS}). The second mapping method subjected the F_5 RIL individuals representing the phenotypic extremes of the population (20 highest- and 20 lowest- accumulators of 3-DOA) to whole genome sequencing (WGS) to 35X depth and BSA (F_5 BSA_{WGS}). To further clarify the genomic loci associated with the black sorghum phenotype, a large F_2 population segregating for pericarp color was screened to identify phenotypic extremes that were then subjected to GBS and QTL analysis identified using BSA. Each F_2 plant was barcoded with a unique DNA sequence to preserve genotypic identity during the BSA analysis, which included 42 individuals per pool (F_2 $BSA_{GBS\text{-}42}$) and a subsequent analysis which included the top and bottom 20 out of the 42 individuals presenting the most extreme phenotypes of the F_2 population (F_2 $BSA_{GBS\text{-}20}$). Four major QTL were detected on chromosomes 1, 2, 4 and 10 across a combination of mapping methods. These mapping efforts also revealed multiple minor effect QTL in 2 or fewer mapping techniques. Sequence alignment was examined in QTL regions by comparing WGS data from RTx3362 and BTx378. Within the QTL interval detected on chromosome 4, WGS data revealed a deletion of a major detoxification gene in RTx3362 that may be responsible for higher cellular stress in the black pericarp tissue as proposed in a previous study (Fedenia et al. 2020). Additionally, structural variation was observed in the genomic region spanning the *Y1* gene previously characterized to regulate 3-DOA production in sorghum vegetative tissue. While these results imply an involvement of

the *Y1* gene in conditioning the black pericarp phenotype, further molecular-based investigation into the genomic region flanking the *Y1* gene is warranted. The genomic regions and candidate genes identified using multiple mapping approaches may ultimately enhance the ability to characterize germplasm via marker-assisted selection and facilitate efficient development of new cultivars possessing the ability to accumulate high levels of 3-DOAs in food-grade specialty sorghum grain.

In addition to the mapping efforts conducted in this study, a preliminary search for signatures of epigenetic regulation of this trait was conducted, and it was determined that multiple generations of grain development in the absence of light could not reverse the UVB-induced black pericarp phenotype. Together, these results provide the foundation for additional cellular and molecular studies that will reveal the intricate cascade of subcellular events leading to the tissue-specific accumulation of these phytochemicals in black seed sorghum.

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a multipurpose C4 grass commonly utilized for feed, silage, and human consumption (FAOSTAT 2019). Sorghum is known for its superior adaptability to environmentally diverse conditions such as high heat and drought in comparison to other cereals such as rice, wheat, and maize (Mundia et al. 2019). The environmental plasticity of sorghum is associated with the persistence of diverse polyphenol production in the grain whereas these bitter or dark compounds were selected against during domestication of most other cereal crops (Olsen et al. 2013; Morris et al. 2013). Given the growing interest in the perceived health benefits of

polyphenol-rich foods, there exists an opportunity to promote and increase production of certain cultivars of grain sorghum to help meet nutritional and agronomic needs of the future.

Sorghum grain contains a wide variety of polyphenols including phenolic acids (benzoic or cinnamic acid derivatives), flavonoids, and condensed tannins; and polyphenol composition highly influences grain pigmentation (Dykes 2008). Grain pigments are formed in the pericarp, testa, and endosperm and are mainly influenced by eight genes: *R* and *Y* (pericarp color), *I* (intensifier of pericarp color), *Z* (pericarp thickness), *S* (spreader of pigment), B_1 and B_2 (presence or absence of testa), and *Tp* (testa color) (Smith and Frederiksen 2000). The two pericarp color genes, *R* and *Y*, epistatically interact to produce the predominant grain colors of white (*R_yy, rryy*), lemon-yellow (*rrY_*), and red (*R_Y_*) (Smith and Frederiksen 2000). Two additional genes, *P* and *Q*, interact to control secondary plant color to thereby determine the color and polyphenol content of leaf, stalk, sheath, and glume tissue. Secondary plant color classifies vegetative tissue into either tan (*ppqq*, *ppQ_*), red (*P_qq*), or purple (*P_Q_*) pigmented plants (Smith and Frederiksen 2000). Of the pericarp- and secondary plant color genes, only the *Y* gene has been officially identified within the current sorghum genome (Sbicolor_454 v3.0.1) (Nida et al. 2019). Certain studies suggest that secondary plant color genes may modify pericarp polyphenol content (Dykes et al. 2005, 2006, 2011). For example, red pericarp sorghum with a red/purple secondary plant color does not predominantly accumulate one flavonoid over another in the grain whereas red pericarp sorghums with tan secondary plant color accumulate high levels of flavones

(Dykes 2008). Various combinations of the genes described above will predominantly influence the final color and polyphenol content of the pericarp tissue.

Amongst the common sorghum grain germplasm, a single accession emerged from western Sudan that displayed a novel black grain color (Hahn et al. 1984). Anomalous to other sorghum grain colors, black grain coloration is associated with the hyperaccumulation of 3-deoxyanthocyanidin (3-DOAs) located in the pericarp tissue (Awika et al. 2004). These compounds are valued for their stability and antioxidant capacity for use in the natural food and nutraceutical industries (Xiong et al. 2019). The genetic inheritance of high 3-DOA accumulation in sorghum pericarp tissue is complex and the phenotype cannot be sufficiently explained by the classic major effect genes impacting sorghum grain color described above. Moreover, the phenotype is not constitutively expressed in all environments and genetic improvement and introgression of this trait has been challenging (Pfeiffer and Rooney 2016). Heritability studies that included segregation distributions revealed that black pericarp coloration and high 3-DOA accumulation was in part, a recessive trait (Pfeiffer and Rooney 2016). Generation means analysis demonstrated that significant additive, dominance, and epistatic effects were also associated with the black grain phenotype, and that the estimated number of genes controlling this phenotype range from 2 to 12 (Pfeiffer and Rooney 2016). Furthermore, black grain sorghum was determined to contain the same pericarp color alleles responsible for red grain color (*R_Y_*), suggesting there are additional genetic factors contributing to this unique phenotype (Pfeiffer and Rooney 2016). The *Y* gene encodes an R2R3 MYB transcription factor that has been proposed to regulate the

flavonoid biosynthetic pathway in sorghum, including 3-DOA production in vegetative tissue (Boddu et al. 2005). Although upregulation of expression of the *Y1* gene was not observed in RTx3362 pericarp developing under UVB light (Fedenia et al. 2020), the involvement of the *Y* gene (acting alone or in concert with other MYB transcription factors) in the expression of the black pericarp trait cannot be excluded based on the limited studies of pericarp gene expression conducted to date. It is well known that transcription factors such as basic helix-loop-helix (bHLH) and WD40 interact with MYBs to regulate the flavonoid pathway (Mano et al. 2007) and thus, multiple transcription factors in addition to the *Y1* R2R3 MYB are potentially involved in the control of this trait. Despite the complex genetic mechanisms governing 3-DOA accumulation in sorghum pericarp tissue, sufficient genetic and genomic resources for sorghum exist to permit a deeper investigation into the genes responsible for this trait.

Sorghum is a self-pollinating diploid with two copies of 10 chromosomes (2n=2x=20) and a genome size of ~730 Mb (Paterson et al. 2009). Critical to the genotyping efforts in sorghum was the release of a high-quality genome sequence for sorghum elite inbred BTx623 (Sbicolor_454 v3.0.1, www.phytozome.jgi.doe.gov; Paterson et al. 2009). Since the first sorghum genome sequence was released in 2009, there have been additional complete genome sequences and annotations released, however, none have originated from a black pericarp genotype. Genomic resources have been developed and subsequently implemented for sorghum germplasm evaluation that include genotyping-by-sequencing (GBS) to identify single nucleotide polymorphisms (SNPs) within restriction site-associated DNA sequences at loci across the genome

(Morishige et al. 2013). High throughput genotyping of DNA samples with sequence-based barcodes ligated to the restriction enzyme-digested DNA of each sample permits efficient and cost-effective detection of SNPs especially when analyzing a large number of individuals (Morishige et al. 2013). The resulting SNP data can be used to construct linkage maps for QTL analysis that aim to identify regions of the genome associated with the trait of interest. This approach is valuable for application in marker-assisted breeding efforts and gene discovery.

An alternative mapping methodology to traditional linkage map construction and subsequent QTL analysis is to deploy BSA in which the DNA from several individuals showing extreme opposite trait values for a given phenotype is bulked together into two respective pools (Giovannoni et al. 1991; Michelmore et al. 1991). Given the accessibility of resequencing resources, the BSA pools are often subjected to WGS and then compared to parental sequence data using analytical techniques such as QTLseqr (Mansfeld and Grumet 2018; Takagi et al. 2013). The strong detection power in BSA resides in the utilization of only the phenotypic extremes of the population. Apart from the loci controlling the trait of interest, all other loci should (given a large enough bulk) assort independently in both bulk pools. The analysis reveals genomic regions with unequal representation of both parental genotypes which permits the trait loci to be identified. Although the major advantage of BSA is to reduce the time, labor, and constraints imposed by traditional linkage mapping, BSA is also a useful tool to increase marker depth across trait loci and to confirm the genomic intervals defined by traditional linkage mapping for the trait of interest (Takagi et al. 2013). Recently, the BSA

approach has successfully detected trait loci for economically valuable traits in a range of crops including squash (Vogel et al. 2020), rice (Takagi et al. 2013), tomato (Illa-Berenguer et al. 2015), watermelon (Branham et al. 2018), and broccoli (Branham and Farnham 2019).

The objective of the present research was to identify regions of the sorghum genome responsible for high 3-DOA production in pericarp tissue using a series of trait mapping methodologies and to subsequently identify a suite of candidate genes residing within these genomic regions. To this end, a bi-parental RIL population was derived from a single cross between two elite Texas inbreds with opposing pericarp color, BTx378 (red pericarp) and RTx3362 (black pericarp) and used for linkage map construction and QTL analysis. To complement this analysis, the phenotypic extremes of this RIL population (20 highest- and 20 lowest- accumulators of 3-DOAs) were also subjected to WGS-BSA. The two bulks were resequenced to ~30X depth and the variants were called using the Genome Analysis Toolkit (GATK) pipeline (Van der Auwera et al. 2013). The SNP data from the two bulks were imported into the R package QTLseqr to identify QTL. To further elucidate the genomic loci associated with the black sorghum phenotype, individuals from a large F_2 population (~10,000 plants) which exhibited phenotypic extremes for pericarp color were identified from the original parental cross (BTx378/RTx3362) and subjected to BSA and QTLseqr. Finally, it has been suggested that multiple generations of grain maturation under pollination bags may lessen the penetrance of the black pericarp phenotype under full sunlight (B. Pfeiffer, personal communication). To address this issue and search for signatures of

epigenetic regulation of this trait, a preliminary investigation was conducted to determine if the black pericarp phenotype and associated 3-DOA accumulation were reversible after multiple generations of grain development in the absence of light. Collectively, the studies reported herein examine the reversibility of the genetic regulation of this trait (i.e. epigenetic) and report the discovery of a series of loci and candidate genes associated with 3-DOA accumulation using a series of contrasting mapping methodologies.

Materials and Methods

Parental Material

To examine the genetic control of black pericarp pigmentation and 3-DOA accumulation, a series of experiments were conducted utilizing the sorghum genotypes RTx3362 (Rooney et al. 2013) and BTx378 (Stevens and Karper 1965). BTx378, a red pericarp sorghum inbred, does not possess a pigmented testa layer or condensed tannins (genetically $YYRRPPQQb_1b_1B_2B_2$) (Pfeiffer 2014; Stephens and Karper 1965). The inbred RTx3362 has a black pericarp that also lacks a pigmented testa layer and condensed tannins (genetically $YYRRPPQQB_1B_1b_2b_2$) (Pfeiffer 2014; Rooney et al. 2013). The black pericarp coloration of RTx3362 is associated with the high accumulation of 3-DOAs.

Expression of Black Pericarp Phenotype and 3-DOA Accumulation after Multiple Generations of Grain Development in Darkness

A preliminary investigation was conducted to determine if the black pericarp phenotype and associated 3-DOA accumulation were reversible after multiple

generations of grain development in the absence of light. Sorghum genotype RTx3362 was grown in a Conviron® BDW Growth Room (Conviron Products of America, Pembina, ND) under environmental conditions previously described (Fedenia et al. 2020). After full anthesis, an acrylic-latex foam coated blackout lining fabric was used to block all light from reaching the panicle throughout grain maturation. This process, which allowed RTx3362 grain to develop in complete darkness, was repeated for an additional 7 self-pollination generations. Following the 8th generation of grain development in the dark, seed of RTx3362$_{blackout}$ were planted in field conditions in College Station, TX (2019) under standard agronomic practices (see Chapter II). As a control, seed of RTx3362 that matured under normal illumination conditions was planted for comparative phenotyping. Panicles from RTx3362$_{blackout}$ and RTx3362 control plots were allowed to mature under full sunlight, and panicles were harvested at physiological maturity (approximately 35 days after full anthesis). Following removal of adhering glumes and grain threshing, 3-DOA concentration of mature grain was quantified using the colorimetric method (Fuleki and Francis 1968) as described in Chapter II and as reported previously (Dykes et al. 2005).

Recombinant Inbred Line Population

To identify regions of the sorghum genome controlling the black pericarp trait, a parental cross between BTx378 and RTx3362 was used to construct an F_5 RIL population (Pfeiffer and Rooney 2016). In 2012, the original parental cross between BTx378 and RTx3362 was made to produce F_1 seed that was self-pollinated to create segregating F_2 individuals. In subsequent generations of selfing and visual screening for

pericarp color, lines whose pericarp color spanned the range from red to intense black coloration were advanced to the F_5 generation to create a resource to elucidate the genetic mechanisms controlling the black pericarp trait (Pfeiffer 2017). In 2016, five representative F_4 panicles from each individual were bulked to comprise the F_5 generation, however, due to a lack of uniformity in panicles in the F_4 generation, single-head panicles selections were made in 2017 and used to reconstruct the F_5 population. Starting in 2018, seed from 224 F_5 RILs from the BTx378/RTx3362 mapping population were grown in three locations: College Station 2018 (CS18), Weslaco 2018 (WE18), and College Station 2019 (CS19). Standard agronomic practices (e.g., fertilization, tillage, pest control, irrigation) for grain sorghum production were applied at all field locations. Each plot was planted 5.2 m long with 0.8 m row width. Panicles were harvested at physiological maturity (approximately 35 days after full anthesis), threshed, and cleaned from adhering glumes to prevent grain weathering and trait degradation prior to quantifying pericarp 3-DOA accumulation.

Phenotyping

To quantify 3-DOA concentration in mature grain, the colorimetric method (Fuleki and Francis 1968) was utilized following a previously established protocol (Dykes et al. 2005). Mature grain samples (~5 g) were ground for 3 min at 12,000 *g* using an IKA® Tube mill control (IKA-Werke, Staufen, Germany). Ground samples (~0.3 g) were extracted in 25 mL 1% HCl in methanol (v/v) for 2 h using an orbital shaker (Bellco Biotechnology, Vineland, NJ) set at low speed (~120 revolutions/minute). Samples were then centrifuged at 2,500 *g* for 13 min. Each extract

was decanted and the supernatant was stored in the dark for 2 h at room temperature (22-23°C). Absorbance of supernatant was read at 485 nm using a spectrophotometer (Beckman DU® 640). The published extinction coefficient of luteolinidin chloride in extract solvent (ε = 27,400) was used to calculate the concentration of 3-DOAs using Lambert-Beer's Law equation (Dykes et al. 2013). All samples were run in duplicate and reported in mg luteolinidin equivalents/g (mg LutE/g).

The linear relationship between 3-DOA quantification using the colorimetric method and ultra-performance liquid chromatography with photodiode array detection (UPLC-PDA) was validated by selecting a subset of individuals from the RIL population that displayed varying pericarp pigmentation and re-quantifying 3-DOA concentration in mature grain samples using UPLC-PDA (Fedenia et al. 2020). The 3-DOA accumulation in samples measured using UPLC-PDA was regressed upon colorimetric measurements to determine how well the data fit the regression line using the coefficient of determination (R^2). All reagents for both quantification methods were HPLC or analytical grade.

Genotyping, Linkage Mapping, and QTL Analysis of the F_5 RIL population

To genotype the RILs for linkage map construction and traditional QTL analysis (F_5 QTL_{GBS}), genomic DNA was extracted from bulked leaf tissue from approximately 10 seedlings per genotype sown for 10-12 days in greenhouse conditions. Leaf tissue was harvested, pulverized, and total genomic DNA extracted using the Quick-DNA™ Plant/Seed Miniprep Kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol. GBS was conducted following the methodology published by Morishige et al.

(2013) with the template libraries created using the methylation-sensitive restriction enzyme *Fse*I (New England BioLabs, Inc.). Single-end sequencing was conducted on the Illumina HiSeq2500 using standard Illumina protocols (Illumina, San Diego, CA) performed by AgriLife Genomics and Bioinformatics Services (Texas A&M University). Reads (150 bp) were analyzed for quality and quantity using FastQC Software v.0.11.5 (Andrews 2010) and further processed using a series of custom Perl scripts to remove sequences lacking the barcode plus partial *Fse*I restriction site, trim the 12 bp barcode, sort sequences corresponding to each individual and compress/count duplicate reads. Filtered reads from each parental line were aligned to the BTx623 reference sorghum genome (Sbicolor_454 v3.0.1, www.phytozome.jgi.doe.gov) using BLASTN. Reads that mapped identically to more than one region of the reference genome were removed. Polymorphisms were identified between aligned reads from each of the parental lines and scored through the progeny using custom Perl and Python scripts following previously described methods (Morishige et al. 2013). An output file consisting of marker allele calls, read alignment to the BTx623 reference genome, and read depth for each RIL was generated. SNP markers were filtered following a previously established pipeline (Bentley et al. 2019) to produce a file of high-quality markers for linkage map construction.

Linkage groups were determined through independence test logarithm of odds (LOD) score and a genetic map was constructed using the R package R/qtl (Broman et al. 2003). To estimate map distances from recombination frequencies, Kosambi's mapping function was implemented. For QTL analysis, 3-DOA accumulation for all

individuals in the RIL population was measured for each field environment (CS18, WE18, CS19) using the colorimetric method described above. A one-dimension scan (*scanone*) and composite interval mapping (CIM) was performed using the R package R/qtl (Broman et al. 2003) with the EM algorithm. The LOD threshold was established using 1000 permutations at $p < 0.05$ and confidence intervals were approximated using Bayesian credible intervals (*bayesesint*) with a probability of 0.95. Phenotypic data from each location was initially analyzed separately. One-way analysis of variance (ANOVA) was performed on raw phenotypic data using JMP®, Version 15 (SAS Institute Inc., Cary, NC, 2020). All effects were considered fixed. Minimal differences were previously observed between plot replications within environments (WL Rooney, personal communication) thus plot replications within environments were not conducted. The model used for 3-DOA accumulation for individual environments was: y = genotype + environment + (genotype x environment) + error. The datasets from the three environments were then combined and transformed into best linear unbiased predictors (BLUPs) using JMP®, Version 15.0 and used for QTL analysis. QTLs observed for separate environments (CS18, WE18, CS19) as well as the BLUPs were compared and major QTL contributing to the black pericarp phenotype were recorded. Annotated genes within significant confidence intervals from each QTL were examined using the current sorghum reference genome and biological function predictions were performed using the gene ontology (GO) tool from PlantTFDB 4.0 using an FDR of $p < 0.05$.

Bulked Segregant Analysis of Phenotypic Extremes in the F_5 RIL Population

The phenotypic extremes of the RIL population (20 highest- and 20 lowest- F_5's for pericarp 3-DOA concentration) were selected using the average colorimetric data collected from all three field locations. A cutoff value of 1.15 and 0.31 mg LutE/g was used to define individuals for the black and red pericarp bulks, respectively. Two bulked pools of DNA were constructed that were then subjected to WGS. Variants between the two bulks were identified using the GATK pipeline (Van der Auwera et al. 2013). The SNP data was then imported into the R package QTLseqr (Mansfeld and Grumet 2018) for identification of QTL (F_5 BSA_{WGS}). To obtain genomic DNA for WGS, 10 seeds per genotype were harvested from single-head descent panicles and sown for 4-5 days in the dark (30 °C). Etiolated tissue was harvested from each F_5 RIL individual in the pools, immediately submerged in LN_2 and pulverized with a mortar and pestle. High molecular-weight DNA (greater than 30 Kbp) was isolated using a conventional cetyl trimethylammonium bromide (CTAB)-based method (Doyle 1991) with modifications. A lysis solution was prepared using 2% CTAB, 1.4 M NaCl, 20 mM EDTA (pH 8.0), 100 mM Tris HCl (pH 7.5), and 1% PVP-40 (with β-mercaptoethanol [0.2%] added to the lysis solution immediately before use). Pulverized tissue (0.15g) was added to a 1.5 mL tube, gently mixed by inversion with 800 µl lysis solution for 5 min and incubated in a water bath (65 °C, 1.5 h). Samples were chilled on ice (2 min), centrifuged (18,500 *g*, 23 °C, 15 min), and the supernatant transferred to 2 mL tubes with 800 µL of Chloroform:Isoamyl alcohol (24:1). Samples were gently mixed by inversion to form an emulsion (~20 min), centrifuged (16,000 *g,* 23 °C, 20 min), and the upper phase

transferred to a 1.5 mL tube. Nucleic acid was precipitated by gently mixing with 0.6 volumes of isopropanol, incubated (23 °C, 10 min), and centrifuged (21,000 *g,* 23 °C, 20 min). The pellet was washed with 70% EtOH (1 mL) then 80% EtOH (1 mL) and allowed to air dry (3 min). Following nucleic acid extraction and precipitation, the pellet was gently resuspended in 1x TE and treated with RNase (Ambion® RNase Cocktail Mix, 37 °C, 1 h). Following RNase digestion, DNA was purified by the addition of SDS (0.5% v/v) and proteinase K (400 µg/ml) and then incubated for 1 h (55 °C). RNA-treated DNA was then phenol:chloroform extracted and precipitated (-20 °C overnight) as detailed above. Pelleted genomic DNA was rinsed with 70% and 80% ethanol and gently resuspended in 1x TE (stored at -20 °C until sequenced). All pipetting was conducted using wide-bore P200 filter tips to minimize shearing of genomic DNA (gDNA). DNA was analyzed for purity at 260 nm and 280 nm on a DeNovix DS-11 spectrophotometer (DeNovix Inc., Wilmington, DE) and for integrity and concentration on the Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA). The concentration of individual DNA samples was adjusted to 40 ng/µl prior to pooling an equal ng-aliquot of each DNA sample for each pool. Paired-end (2 x 150 bp) sequencing of the gDNA libraries was performed on the NovaSeq 6000 (Illumina Inc., San Diego, CA). Library construction with the TrueSeq DNA PCR FREE preparation kit (Illumina Inc., San Diego, CA) and sequencing to ~30x depth was performed for the parents of the population and the two bulked DNA pools by AgriLife Genomics and Bioinformatics Services (Texas A&M University). Sequence cluster identification, quality pre-filtering, base calling and uncertainty assessment was conducted in real time

using Illumina's HCS 2.2.58 and RTA 1.18.64 software with default parameter settings. FastQC (Andrews 2010) was used to confirm the quality of the resulting fastq files before the reads were processed using the GATK pipeline (Van der Auwera et al. 2013). Raw reads in fastq format were converted to binary alignment map (bam) files and adapter-trimming tags were added using *MarkIlluminaAdapters* (Picard tools v.2.4.1). Reads were then aligned to the current sorghum reference genome (Sbicolor_454 v3.0.1) using Burrows-Wheeler alignment (BWA-MEM) v0.7.4.1 (Li and Durbin 2010). The GATK *HaplotypeCaller* (Van der Auwera et al. 2013) was used to perform SNP calling and generate variant call format (VCF) files. The VCF files were filtered using the GATK *VariantFiltration*. The SNP data from the GATK pipeline was imported into the QTLseqr package (Mansfeld and Grumet 2018) to calculate the Δ(SNP-index) between the two DNA bulks.

To assess the coverage and the quality of the alignment of the parental WGS data to the current sorghum reference genome (Sbicolor_454 v3.0.1), fastq files were imported into the CLC Genomics Workbench v21.0.3 (Qiagen, Valencia, CA) and mapped to the current sorghum reference genome (Sbicolor_454 v3.0.1) where read alignment parameters were used as follows: mismatch cost = 2; insertion and deletion cost = 3; 50% minimum read length required to match reference; 90% similarity between reads and the reference genome. The Local Realignment Tool was used after mapping to the reference genome to correct for mapping errors and perform coverage analysis using default parameters in the CLC Genomics Workbench v21.0.3 (Qiagen, Valencia, CA).

Bulked Segregant Analysis of Phenotypic Extremes in the F_2 Population

A population of approximately 10,000 F_2 progeny originating from the same parental cross used to construct the F_5 RILs was grown in CS19 using standard agronomic practices. At grain maturity, the F_2 panicles were visually rated for color development. A representative panicle from each parent grown in CS19 was used to visually rate and select F_2 individuals representing the two phenotypic extremes for grain color (red, black). Panicles were hand-harvested and the plants were flagged with tape for later identification. The panicles were threshed with seed cleaned to remove adhering glumes prior to colorimetric quantification of 3-DOA concentrations. To select which F_2 progeny to include in BSA, a 3-DOA cutoff of 1.00 and 0.31 mg LutE/g was used for the black and red individuals, respectively. For DNA isolation, young leaf tissue was harvested from newly emerged tillers from each F_2 individual visually selected as a phenotypic extreme for seed color. Following colorimetric determination of pericarp 3-DOA concentrations, leaf tissue from F_2's selected for bulk inclusion (42 F_2's per bulk) was ground, homogenized, and total gDNA was extracted with the use of the Quick-DNA™ Plant/Seed Miniprep Kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol. GBS template was generated using the methylation-sensitive restriction enzyme *Ngo*MIV and each individual was assigned a unique sequence-based barcode to maintain the identity of the sequences (Morishige et al. 2013). Single-end sequencing was conducted on the Illumina HiSeq2500 using standard Illumina protocols (Illumina, San Diego, CA) performed by AgriLife Genomics and Bioinformatics Services (Texas A&M University). F_2 DNA barcoding permitted

analyses of the dataset via construction of different size pools by inclusion/exclusion of individuals based on varying stringency cutoff values for 3-DOA concentration. GBS reads (150 bp) were analyzed for quality and quantity using FastQC Software v.0.11.5 (Andrews 2010) and further processed using a series of custom Perl scripts to remove sequences lacking the barcode plus partial *Ngo*MIV restriction site, trim the 12 bp barcode and sort sequences corresponding to each individual. The processed read files were imported into the GATK pipeline (Van der Auwera et al. 2013), the SNPs were imported into QTLseqr, and the Δ(SNP-index) was calculated between the two DNA bulks as described above. To investigate the effect of pool size on QTL detection, two separate analyses were conducted as follows: 1) the barcoded GBS data was filtered to conduct an analysis that included the top 42 individuals per bulk for each extreme phenotype (F_2 BSA_{GBS-42}); and 2) analysis that included only the top 20 individuals per bulk for each extreme phenotypes (F_2 BSA_{GBS-20}). A summary of the trait mapping methods employed in this chapter is listed in Table 4 and graphically depicted in Figure 5.

Table 4. Summary of mapping and analytical methods used to map loci associated with 3-DOA accumulation in mature sorghum grain.

Analytical Technique	**Abbreviation**	**Description**	**Generation[a]**	**Number**	**Genotyping Method**
Linkage mapping and QTL analysis	F_5 QTL_{GBS}	F_5 RIL population individuals	F_5	224	Genotyping-by-sequencing (*Fse*I)
Bulked segregant analysis, QTLseqr	F_5 BSA_{WGS}	Subset of F_5's 20 highest and 20 lowest pericarp 3-DOA level	F_5	20 per bulk	Whole genome resequencing
Bulked segregant analysis, QTLseqr	F_2 BSA_{GBS-42}	Subset of F_2's 42 highest and 42 lowest pericarp 3-DOA level	F_2	42 per bulk	Genotyping-by-sequencing (*Ngo*MIV)
Bulked segregant analysis, QTLseqr	F_2 BSA_{GBS-20}	Subset of F_2's 20 highest and 20 lowest pericarp 3-DOA level	F_2	20 per bulk	Genotyping-by-sequencing (*Ngo*MIV)

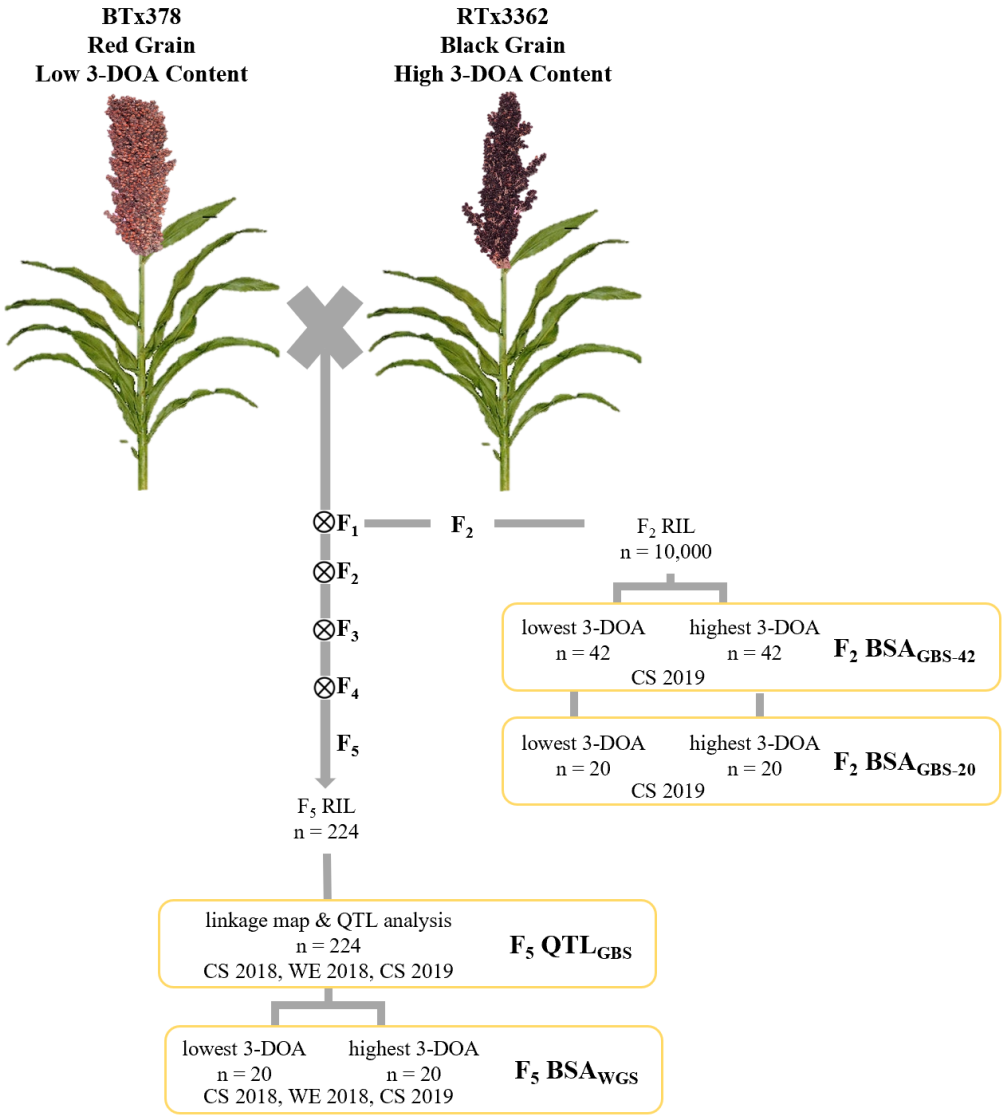


Figure 5. Schematic illustration of the F_2 and F_5 mapping populations generated from a biparental cross used for linkage mapping of genomic regions associated with 3-DOA accumulation in sorghum grain. The number of individuals (n) included in each analysis is provided and the location and year of phenotyping. Yellow boxes represent the four trait mapping methodologies and their abbreviations as detailed in Table 4.

Screening for Candidate Gene Deletion in the Phenotypic Extremes in the F_5 RIL Population

A PCR-based assay was used to screen for the presence or absence (gene deletion) of the coding region of Sobic.004G185100 in RILs representing the phenotypic extremes of the population (20 highest- and 20 lowest- F_5's for pericarp 3-DOA concentration). For comparative purposes, the phenotypic-extreme RILs was also screened for the presence of the coding region of Sobic.002G074400, a gene belonging to the same gene family as Sobic.004G185100 but detected in both the BTx378 and RTx3362 genomes. PCR amplification of genomic DNA was carried out using GoTaq® DNA Polymerase (Promega, USA) with the 10 µL reaction consisting of 1X GoTaq® Reaction Buffer, 0.2 mM of each dNTP, 0.5 µM of each primer, 1.25u GoTaq® DNA Polymerase, 1.5 mM $MgCl_2$, and 10 ng of template DNA. Primers used are as follows—Sobic.004G185100, forward: 5′-GTACACGAGTAGTCATTCATCATG-3′, reverse: 5′-CAAGACCAAGAGCGACAACAAG-3′; Sobic.002G074400, forward: 5′-CACTGCTCACGTATGCTCTCT-3′, reverse: 5′-AGCAGCCACGGTGATTTCTT-3′. Thermal cycling parameters for the PCR reaction were set to 95 °C, 2 min followed by 95 °C one min, 60 °C one min, 72 °C one min for 30 cycles followed by a seven min final extension at 72 °C. PCR reaction were subsequently electrophoresed on 3% SFR agarose (Super Fine Resolution, VWR Life Science) gels and visualized using ethidium bromide staining.

Results

Correlation between Colorimetric and UPLC-PDA Quantification of 3-DOAs

Visual rating of panicles for pericarp color has been routinely used in sorghum improvement as a field-based indirect selection tool for 3-DOA concentration of mature grain. While a relationship between pericarp color and 3-DOA concentration does exist, visual color ratings were determined to lack the accuracy required for quantifying 3-DOA concentration in this study. While UPLC-PDA is very accurate and precise, the method is not easily amenable to high throughput screening as was required in the present trait mapping study. To utilize a high throughput and accurate method of quantifying pericarp 3-DOA concentration, a linear relationship was established between 3-DOA levels quantified using the colorimetric method (Fuleki and Francis 1968; Dykes et al. 2005) and UPLC-PDA. Using a representative sampling of grain from field-grown RILs, a strong correlation ($R^2 = 0.9867$) was achieved between the colorimetric assay and the UPLC-PDA method thereby verifying the utility of the colorimetric method as a high throughput, accurate method for quantifying 3-DOA concentration for all grain examined in this study (Figure 6).

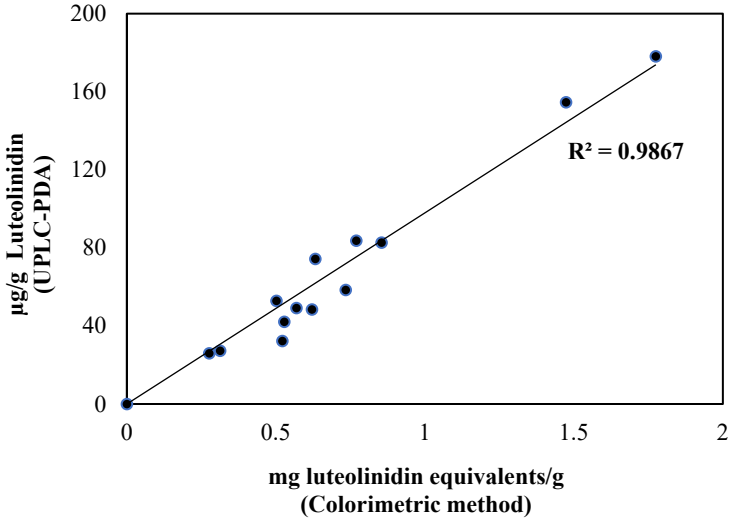


Figure 6. Correlation between the colorimetric method and UPLC-PDA quantification of luteolinidin, the predominant 3-DOA in black sorghum pericarp.

Black Pericarp Phenotype and 3-DOA Accumulation after Multiple Generations of Unilluminated Grain Development

During the development of the present RIL population, anecdotal evidence suggested that multiple generations of grain maturation under pollination bags lessened the penetrance of the black pericarp phenotype under full sunlight (B. Pfeiffer, personal communication). Reversibility in the genetic control of traits expressed in response to environmental stimuli is often a signature of epigenetic regulation and has been previously observed in phytochemical accumulation in response to UVB in high altitude maize landraces (Rius et al. 2016). To address this issue and gain information on the potential reversibility of the genetic control of 3-DOA accumulation in sorghum pericarp, RTx3362 grain was permitted to mature in complete darkness for 8 selfing generations, and then grain development in the subsequent generation was allowed to

occur under full sunlight. The results are presented in Figure 7. After 8 generations of grain development in the dark, RTx3362$_{blackout}$ grain under full sunlight was black and was visually indistinguishable from the grain of control RTx3362 panicles. Quantifying 3-DOA levels in mature grain from field-grown plants revealed an average 3-DOA concentration of RTx3362 $_{blackout}$ grain of 1.63 mg LutE/g, which was similar to the average 3-DOA concentration (1.46 mg LutE/g) of control RTx3362 panicles. If field-grown panicles of RTx3362$_{blackout}$ were covered for an additional generation (9th generation), the pericarp appeared red and the average 3-DOA accumulation was 0.53 mg LutE/g. Based on these observations, multiple generations of pericarp development in the absence of the light-signal responsible for the expression of the black pericarp trait did not reverse or inhibit the penetrance of the trait under full sunlight.

RTx3362$_{blackout}$ covered for 9th generation	RTx3362$_{blackout}$ grain under full sunlight	RTx3362 control under full sunlight
0.53 mg LutE/g	1.63 mg LutE/g	1.46 mg LutE/g

Figure 7. 3-DOA concentration and pericarp pigment development in three sorghum treatments grown in College Station, TX 2019.

Phenotyping F_5 RIL population for Pericarp 3-DOA Concentration in Field Environments

The average 3-DOA accumulation of the parental genotypes BTx378 and RTx3362 in different environments ranged from 0.14 ± 0.01 to 0. 23 ± 0.03 and 1.29 ± 0.09 to 1.74 ± 0.07 mg LutE/g, respectively (Table 5). Within a given environment, significant differences in 3-DOA levels were observed amongst the F_5 RILs (Table 6, $p < 0.001$). A significant environment and genotype x environment effect was observed

(Table 6, $p < 001$), however, this did not change the ranking of individuals that were considered the highest and lowest accumulators of 3-DOAs in the population. In all but the CS19 environment, transgressive segregants were observed as the maximum 3-DOA accumulation exceeded that of the black parent (data not shown).

Table 5. Statistical summary of grain 3-DOA concentration of field-grown RILs from three environments.

		3-DOA concentration (mg LutE/g)		
Year	**Environment**	**BTx378 Ave. ± SD[a]**	**RTx3362 Ave. ± SD**	**RIL Population Ave. ± SD**
2018	Weslaco, TX	0.18 ± 0.02	1.37 ± 0.02	0.45 ± 0.31
2018	College Station, TX	0.23 ± 0.03	1.29 ± 0.09	0.65 ± 0.37
2019	College Station, TX	0.14 ± 0.01	1.74 ± 0.07	0.41 ± 0.24

[a]Standard Deviation

Table 6. Mean squares of ANOVA for the colorimetric quantification of 3-DOA concentration of field-grown RIL grain from three environments. Degrees of freedom (df) for all sources of variation in the combined environments are shown.

	Environment				
Source	**CS18**	**WE18**	**CS19**	**Combined**	**df**
Genotype	0.28***	0.19***	0.11***	0.48***	223
Environment				7.42***	2
Genotype x Environment				0.06***	443

***Significant at 0.001 probability level

Genotyping, Linkage Mapping, and 'F5 QTLGBS analysis' of 224 F5 RILs

GBS analysis of the F_5 RIL population resulted in the identification of 2471 total SNP markers that were scored across the 224 RILs. After discarding 1297 SNPs that were redundant or had considerable missing data, the final linkage map contained 1174 informative SNPs with an estimated coverage of 1406.2 centimorgans (cM) mapped to 10 linkage groups representing the 10 sorghum chromosomes. By comparison to the sequenced sorghum reference genome, SNP coverage was determined and found to be well distributed across the 10 chromosomes with an average of one marker per 1.3 cM

(Figure 8 & Table 7); however, SNPs in pericentromeric regions were sparse due to the use of a methylation-sensitive enzyme in Illumina template preparation (Morishige et al. 2013).

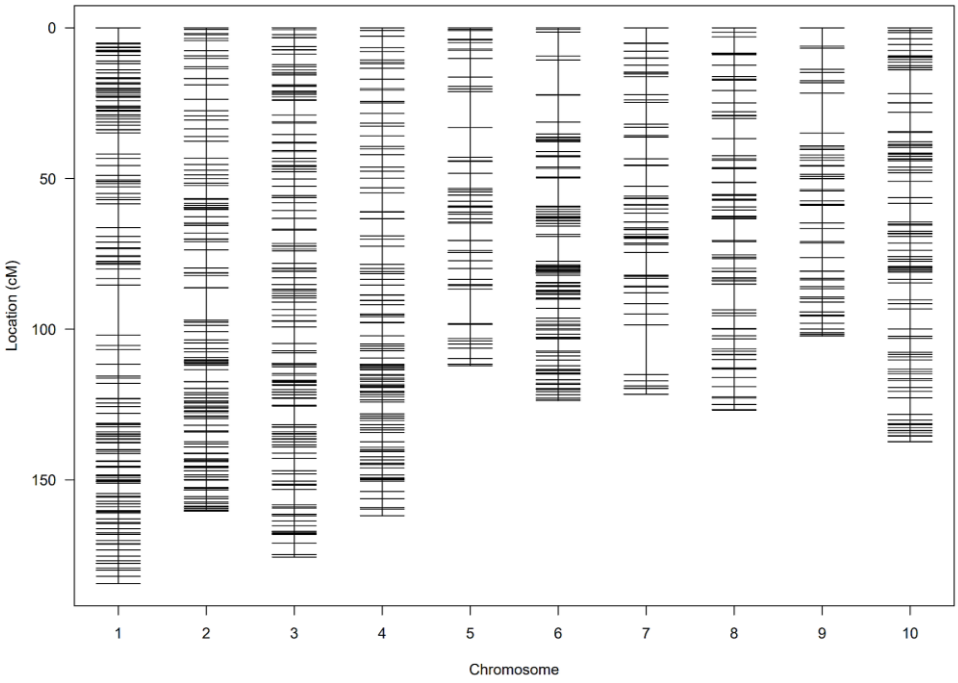


Figure 8. Genetic linkage map constructed from 224 F_5 RILs for each linkage group representing the 10 sorghum chromosomes as visualized using R/qtl.

Table 7. Genetic linkage map marker coverage across the 10 sorghum chromosomes generated from 224 F_5 RILs.

	Chromosome										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Chromosome Length (cM[a])	184.5	160.3	175.6	161.9	112.1	123.7	121.6	126.9	102.3	137.4	1406.2
Chromosome Length (Mb[b])	79.6	76.8	74.2	67.5	71.7	60.4	64.9	62.0	54.5	60.3	672.0
Number of Markers	197	160	165	123	67	130	72	85	62	113	1174
Marker Density (cM)	0.9	1.0	1.1	1.3	1.7	1.0	1.7	1.5	1.6	1.2	
Largest Interval Between Markers (cM)	16.6	10.7	6.1	6.3	11.9	11.5	16.5	8.5	13.3	7.9	
Largest Interval Between Markers (Mb)	11.9	31.2	29.7	37.0	19.9	21.7	14.6	30.5	26.3	13.7	

[a]Centimorgans
[b]Megabases

Marker-trait associations for 3-DOA concentration in sorghum grain were analyzed from phenotypic data gathered across all three environments. Minimal differences were observed between QTL detected when analyzing each environment separately (data not shown), therefore, the observations of 3-DOA accumulation were transformed into BLUP values and used as the final phenotypic input for QTL analysis. The results of the F_5 QTL_{GBS} analysis are shown in Table 8 and Figure 9.

Two QTL for 3-DOA accumulation in sorghum grain were detected in the F_5 QTL_{GBS} analysis using the LOD threshold score of 4.18 established using the *scanone* function in R/qtl with 1000 permutations (Table 8 & Figure 9). One QTL was observed on chromosome 1 (71.9-73.1 Mb, LOD 8.9) that accounted for 16.7% of the phenotypic variation observed within the population. The 95% Bayes credible confidence interval for this QTL contained 147 genes based on the current sorghum reference genome. The

QTL found on chromosome 2 (8.2-10.6 Mb, LOD 4.8) explained 9.4% of the phenotypic variation (Table 8 & Figure 9) and the 95% Bayes credible confidence interval encompassed 195 genes. Collectively, using the F_5 QTL_{GBS} methodology, these two trait loci explain 26.1% of the PVE in 3-DOA levels in this RIL population. If the threshold LOD score is relaxed to LOD 3.7, two additional QTL are detected: chromosome 4 (52.4-61.4 Mb, LOD 3.9, 7.7% PVE) and chromosome 10 (56.3-58.1 Mb, LOD 3.8, 7.5% PVE) (Table 8 & Figure 9). The 95% Bayes credible confidence interval encompassed 1048 and 207 genes within the QTL on chromosomes 4 and 10, respectively. When including the QTL detected with a slightly relaxed LOD threshold, the four trait loci explain 41.3% of the PVE in 3-DOA levels in this RIL population. While QTL observed by reducing the LOD score below the threshold predicted by permutation analysis may be suspect, it provides insight into genomic regions that may control this complex trait if more powerful mapping populations were developed and accurately phenotyped.

Table 8. QTL detected for pericarp 3-DOA accumulation in the 224 F_5 RILs using BLUPs.

Chr[a]	Peak Position	Confidence Interval[b]	Left Marker	Peak Marker	Right Marker	LOD	PVE[e]	Genes in Interval
	cM[c]		Mb[d]					
1	146.4	144.1-150.6	chr01_71.9	chr01_72.5	chr01_73.1	8.9	16.7%	147
2	47.3	43.4 – 48.9	chr02_8.2	chr02_9.8	chr02_10.6	4.8	9.4%	195
4[f]	98.6	91.4 – 120.3	chr04_52.4	chr04_54.1	chr04_61.4	3.9	7.7%	1048
10[f]	117.5	113.7 – 128.7	chr10_56.3	chr10_57.1	chr10_58.6	3.8	7.5%	207

[a] Chromosome
[b] $p < 0.05$ (± 1 LOD)
[c] Centimorgans
[d] Megabases
[e] Percent variance explained
[f] Detected using 3.7 LOD threshold

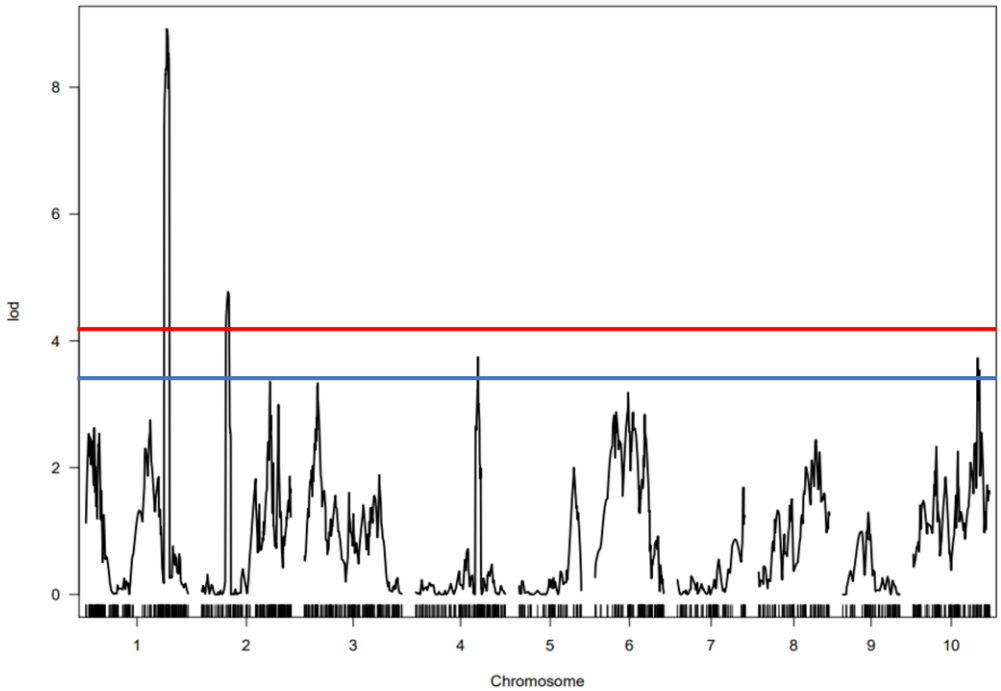


Figure 9. Illustration of QTL detected for 3-DOA accumulation across 10 sorghum chromosomes in an F_5 RIL population utilizing the F_5 QTL_{GBS} analysis. QTL were predicted using BLUPs of 3-DOA accumulation generated from observations in three environments. The Composite Interval Mapping (CIM) method from R/qtl calculated regions on chromosomes 1 and 2 that harbored QTL exceeding the threshold LOD score of 4.18 (red line; $p < 0.05$). The blue line depicts additional QTL (chromosomes 4 and 10) exceeding a less stringent LOD threshold of 3.7 ($p < 0.10$).

F_5 RIL Population Whole Genome Resequencing-based Bulked Segregant Analysis, 'F_5 BSA_{WGS}'

To provide supportive evidence of the QTL intervals detected by traditional linkage and QTL analysis, we applied a BSA method to the phenotypic extremes of individuals within the same RIL population using WGS genotyping (F_5 BSA_{WGS}). Twenty RILs per pool were defined to consistently accumulate either high or low levels of 3-DOAs. The 3-DOA levels in the bulked genotypes ranged from 0.21-0.31 mg LutE/g for the pool of low 3-DOA RILs (red grain) vs. 1.15-2.12 mg LutE/g for the pool of high 3-DOA concentration RILs (black grain). WGS of bulk DNAs yielded

167,772,350 and 178,385,606 total reads for the black and red pool, respectively, with 81% of the reads from the black pool mapped in pairs (135,829,762) and 81.45% of the reads from the red pool mapped in pairs (145,300,606). The average genome coverage was approximately 34x for the black pool and 36x for the red pool. After trimming and filtering the original 3,483,772 SNPs called using the Sbicolor_454 v3.0.1 sorghum genome for variant call format (VCF) within the GATK pipeline, 2,102,988 SNPs remained that were imported into QTLseqr. After calculating the Δ(SNP-index) between the red and black bulks for each SNP, a total of nine genomic regions were significantly ($p < 0.05$) associated with 3-DOA accumulation (Table 9). The intervals of the nine QTL contained 612 genes annotated in the current sorghum genome. The position of the highest Δ(SNP-index) was 0.40 on chromosome 1 within the interval of 62,066,498-62,596,122 bp.

Table 9. Bulked segregant analysis using the QTLseqr method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_5 RILs. Each bulk was comprised of the 20 highest- (black bulk) and 20 lowest- (red bulk) 3-DOA RILs subjected to WGS (F_5 BSA_{WGS}).

Chr[a]	Start (bp)	End (bp)	Interval (bp)	Number of SNPs	Peak Δ(SNP-index) Value[b]	Peak Δ(SNP-index) Position	Average Δ(SNP-index) Value[b]	Genes in Interval
1	62,066,498	62,596,122	529,624	1577	0.40	62,556,115	0.40	66
1	66,658,640	66,687,954	29,314	105	-0.35	66,687,954	-0.34	4
1	66,919,920	67,058,694	138,774	355	-0.38	66,979,691	-0.38	20
1	70,772,536	70,952,205	179,669	338	-0.38	70,952,205	-0.36	22
1	70,977,921	72,081,769	1,103,848	1581	-0.46	71,405,124	-0.42	137
4	51,680,645	52,700,848	1,020,203	5369	0.40	52,033,554	0.39	82
6	39,669,568	40,949,688	1,280,120	3985	0.39	40,691,413	0.38	46
6	47,389,063	48,220,901	831,838	4782	0.38	47,871,912	0.37	112
10	59,457,876	60,231,482	773,606	2112	0.40	59,798,222	0.39	123

[a] Chromosome

[b] If Δ(SNP-index) < 0, the allele contributing to the trait is from the red bulk. If Δ(SNP-index) > 0, the allele contributing to the trait is from the black bulk.

Bulked Segregant Analysis of Phenotypic Extremes in an F_2 Population using GBS, 'F_2 BSA_{GBS}'

To complement the analytical techniques conducted using the F_5 RIL population that was limited to 224 lines, F_2 individuals from the same parental cross (BTx378/RTx3362) were used to construct a large mapping resource consisting of nearly 10,000 plants. After preliminary field-based selection of mature panicles based on mature grain color, the colorimetric method identified those individuals that possessed either the highest 3-DOA (black bulk, 42 individuals) or the lowest 3-DOA concentration in pericarp (red bulk, 42 individuals). The 3-DOA concentrations from the 42 individuals in the black bulk ranged from 1.01-2.02 mg LutE/g, which indicated that transgressive segregant F_2 individuals for 3-DOA concentration existed in the population (see Table 5, RTx3362). The 3-DOA accumulation in the individuals comprising the red F_2 bulk ranged from 0.14-0.31 mg LutE/g which was comparable to the levels observed in the red parent, BTx378 (0.14 mg LutE/g, College Station 2019, Table 5).

To identify genomic regions in the F_2 phenotypic extremes that are associated with 3-DOA accumulation, GBS-BSA was utilized to identify polymorphic loci between the two bulks of individuals displaying phenotypic extremes. GBS analysis of the 42 F_2 individuals from both phenotypic bulks resulted in 90,868 SNPs and INDELS with quality filtering reducing the final number of polymorphisms to 40,942 high-quality variants for further analysis. F_2 BSA_{GBS-42} analysis resulted in the identification of 24 significant ($p < 0.05$) QTL detected across chromosomes 1, 2, 3, 4, 6, 9, and 10 (Table

10). Given the large number of marker-trait associations observed using bulks of 42 individuals, the stringency of the phenotypic cutoff delineating each pool size was increased to include only the top 20 extreme F_2's for 3-DOA concentration for either pool. The range of 3-DOA concentrations for the 20 individuals in the black and red pool was 1.19-2.02 mg LutE/g and 0.14-0.23 mg LutE/g, respectively. Analysis of the smaller pools resulted in a total of 76,744 SNPs and INDELS with quality filtering reducing the final number of high-quality markers to 37,767 for use in QTLseqr. F_2 $BSA_{GBS\text{-}20}$ analysis resulted in the identification of 26 significant ($p < 0.05$) QTL detected across chromosomes 1, 2, 4, 6, 7, 9, and 10 (Table 11).

Table 10. Bulked segregant analysis using the QTLseqr method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_2 BSA_{GBS-42}. Each bulk was comprised of the 42 highest- (black bulk) and 42 lowest- (red bulk) 3-DOA F_2's subjected to GBS.

Chr[a]	Start (bp)	End (bp)	Interval (bp)	Number of SNPs	Peak Δ(SNP-index) Value[b]	Peak Δ(SNP-index) Position	Average Δ(SNP-index) Value[b]	Genes in Interval
1	62,346,774	62,635,734	288,960	28	0.28	62,536,355	0.27	38
1	65,662,769	66,574,739	911,970	61	-0.33	66,324,562	-0.3	102
1	69,927,640	70,137,814	210,174	55	0.3	70,082,611	0.29	24
1	71,221,508	71,894,293	672,785	91	-0.32	71,391,095	-0.29	78
1	76,483,472	79,052,762	2,569,290	368	0.37	77,665,672	0.34	333
1	79,173,708	79,361,409	187,701	35	0.27	79,173,708	0.26	21
2	9,013,718	11,183,559	2,169,841	247	0.30	9,772,173	0.28	35
2	20,432,550	22,202,413	1,769,863	90	0.31	21,499,667	0.30	38
2	28,505,205	33,605,241	5,100,036	15	0.66	31,375,976	0.40	16
2	38,300,258	41,437,717	3,137,459	10	0.43	40,519,198	0.40	13
2	43,647,412	43,647,412	1	1	0.54	43,647,412	0.54	0
2	46,825,739	47,254,441	428,702	13	0.32	47,254,441	0.31	12
2	50,004,004	50,696,658	692,654	16	0.29	50,358,378	0.28	20
2	55,189,021	58,878,908	3,689,887	369	0.41	57,593,798	0.35	257
3	23,271,973	23,271,973	1	1	-0.33	23,271,973	-0.33	0
4	25,923,167	25,923,167	1	1	0.41	25,923,167	0.41	1
4	42,283,955	42,283,955	1	1	0.33	42,283,955	0.33	0
4	47,487,905	47,487,905	1	1	0.32	47,487,905	0.32	0
4	60,881,779	62,228,675	1,346,896	335	-0.31	61,151,317	-0.3	166
6	12,449,226	13,137,279	688,053	6	0.37	12,855,474	0.34	4
6	20,976,158	20,976,158	1	1	-0.42	20,976,158	-0.42	0
9	26,809,543	27,503,117	693,574	10	0.37	27,502,979	0.36	1
10	28,595,719	28,604,731	9,012	4	0.37	28,595,719	0.37	0
10	36,403,826	36,802,816	398,990	2	0.34	36,403,826	0.32	0

[a] Chromosome

[b] If Δ(SNP-index) < 0, the allele contributing to the trait is from the red bulk. If Δ(SNP-index) > 0, the allele contributing to the trait is from the black bulk.

Table 11. Bulked segregant analysis using the QTLseq method for quantitative trait loci ($p < 0.05$) detection associated with 3-DOA accumulation in F_2 BSA_{GBS-20}. Each bulk was comprised of the 20 highest- (black bulk) and 20 lowest- (red bulk) 3-DOA F_2's subjected to GBS.

Chr[a]	Start (bp)	End (bp)	Interval (bp)	Number of SNPs	Peak Δ(SNP-index) Value[b]	Peak Δ(SNP-index) Position	Average Δ(SNP-index) Value[b]	Genes in Interval
1	66,159,589	66,564,898	405,309	44	-0.40	66,324,562	-0.37	44
1	71,312,036	71,618,917	306,881	36	-0.35	71,409,681	-0.34	35
1	76,718,523	78,191,056	1,472,533	182	0.39	77,074,062	0.37	196
1	78,280,512	78,301,562	21,050	6	0.34	78,280,512	0.34	2
2	9,013,718	10,398,840	1,385,122	155	0.37	9,772,173	0.36	121
2	15,957,114	15,957,114	1	1	0.36	15,957,114	0.36	1
2	17,707,024	17,990,827	283,803	29	0.36	17,707,024	0.36	10
2	20,565,081	20,565,081	1	1	0.34	20,565,081	0.34	1
2	20,701,233	21,156,716	455,483	28	0.35	20,701,233	0.35	17
2	29,232,722	29,555,390	322,668	4	0.40	29,232,722	0.38	3
2	38,300,258	41,437,717	3,137,459	6	0.48	40,519,198	0.42	13
2	46,167,486	47,204,480	1,036,994	16	0.37	46,167,486	0.36	17
2	49,656,835	52,394,715	2,737,880	35	0.53	51,563,139	0.40	69
2	56,649,280	58,724,576	2,075,296	248	0.42	57,593,798	0.39	148
4	25,923,167	25,923,167	1	1	0.48	25,923,167	0.48	0
6	8,154,356	8,664,449	510,093	7	0.37	8,154,356	0.36	1
6	9,322,566	9,700,464	377,898	8	0.35	9,334,747	0.34	4
6	11,836,990	12,855,474	1,018,484	4	0.39	12,196,136	0.38	5
6	16,768,444	17,133,432	364,988	9	0.46	17,133,432	0.43	0
6	19,231,779	19,282,816	51,037	5	0.41	19,282,816	0.40	1
6	28,733,225	30,801,609	2,068,384	5	0.48	29,540,608	0.43	16
6	33,538,142	35,677,867	2,139,725	11	0.39	33,982,708	0.38	16
6	41,887,592	42,667,833	780,241	79	0.36	42,092,594	0.35	47
7	31,733,725	31,733,798	73	3	0.41	31,733,725	0.41	1
9	31,638,250	31,638,293	43	2	0.59	31,638,250	0.59	0
10	26,172,023	26,172,023	1	1	-0.44	26,172,023	-0.44	0

[a] Chromosome
[b] If Δ(SNP-index) < 0, the allele contributing to the trait is from the red pool. If Δ(SNP-index) > 0, the allele contributing to the trait is from the black pool.

Collocating Genomic Regions Associated with Pericarp 3-DOA Concentration from the Different Mapping Methodologies

Amongst the four mapping methods, the F_2 BSA_{GBS-20} analysis detected the largest number of QTL (26) followed by the larger F_2 bulk, F_2 BSA_{GBS-42}, (24) (Table 10 & Table 11). The F_5 BSA_{WGS} and F_5 QTL_{GBS} analysis detected 9 and 4 total QTL, respectively (Table 8 & Table 9). As the individuals included in the F_2 bulk were varied

from 42 to 20, the number of high-quality markers available for analysis were somewhat comparable (40,942 vs. 37,767) (Table 12).

The overlap in detected trait loci were first examined between the methods deployed on either the F_5 population or the F_2 population. Herein, confidence intervals for overlapping QTL were reported as the sum of the respective LOD ranges instead of only the portion of actual overlap. Intervals were reported this way to recognize that candidate genes may reside close but not always within a 1-LOD cutoff, thus, the entire sum of the interval regions was considered for further investigation. When including the additional loci found using the relaxed LOD threshold (3.7), two overlapping genomic regions significantly associated with 3-DOA accumulation in pericarp were detected between the F_5 QTL_{GBS} and F_5 BSA_{WGS} analysis. The first region of overlap resided on chromosome 1 spanning 71.9 – 72.1 Mb (Table 12). The second region was detected on chromosome 4 which ranged from 52.4 – 61.4 Mb (Table 12). Although there were no overlapping QTL on chromosomes 2, 6, and 10, the QTL detected on chromosome 10 for both methodologies resided less than 1 Mb apart (Table 12). Reducing the stringency of the analyses and/or slightly expanding the confidence intervals flanking these trait loci on chromosome 10 could potentially create overlap in these QTL. Overlap in QTL was also observed between analyses conducted on the F_2 population. The following genomic regions were detected in both the F_2 $BSA_{GBS\text{-}42}$ and F_2 $BSA_{GBS\text{-}20}$ analyses: chromosome 1 (66.2 – 66.6; 71.2 – 71.9; 76.5 – 79.1 Mb), chromosome 2 (9.0 – 11.2; 20.4 – 22.2; 46.2 – 47.3; 55.2 – 58.9 Mb), chromosome 4 (25.9 – 25.9 Mb), and chromosome 6 (11.8 – 13.1 Mb) (Table 12). Composite examination of QTL intervals

across the 4 mapping methodologies consistently revealed a significant region associated with 3-DOA accumulation in sorghum pericarp tissue on chromosome 1 where the sum of the confidence intervals spanned from 71.0 - 73.1 Mb (Table 12). A second region of overlap between significant confidence intervals was discovered in 3 of the 4 mapping techniques on chromosome 2 from 8.2 -11.2 Mb (Table 12). Finally, a third region of overlap between significant confidence intervals was discovered on chromosome 4 where the sum of 3 of the 4 mapping techniques covered a range of 51.7 – 62.2 Mb (Table 12).

Table 12. Summary of genomic regions significantly associated with pericarp 3-DOA concentration from the different mapping methodologies. Details of detection techniques are listed in Table 4.

Chromosome	**F_5 QTL_{GBS}** Interval (Mb)	**F_5 BSA_{WGS}** Interval (Mb)	**F_2 BSA_{GBS-42}** Interval (Mb)	**F_2 BSA_{GBS-20}** Interval (Mb)
1	71.9 - 73.1	62.1 - 62.6 66.7 - 66.7 66.9 - 67.1 70.8 - 71.0 71.0 - 72.1	62.3 - 62.6 65.7 - 66.6 69.9 - 70.1 71.2 - 71.9 76.5 - 79.1 79.2 - 79.4	66.2 - 66.6 71.3 - 71.6 76.7 - 78.2 78.3 - 78.3
2	8.2 - 10.6		9.0 - 11.2 20.4 - 22.2 28.5 - 33.6 38.3 - 41.4 43.6 - 43.6 46.8 - 47.3 50.0 - 50.7 55.2 - 58.9	9.0 - 10.4 16.0 - 16 17.7 - 18 20.6 - 20.6 20.7 - 21.2 29.2 - 29.6 38.3 - 41.4 46.2 - 47.2 49.7 - 52.4 56.6 - 58.7
3			23.3 - 23.3	
4	52.4 - 61.4	51.7 - 52.7	25.9 - 25.9 42.3 - 42.3 47.5 - 47.5 60.9 - 62.2	25.9 - 25.9
6		39.7 - 40.9 47.4 - 48.2	12.4 - 13.1 21.0 - 21.0	8.2 - 8.7 9.3 - 9.7 11.8 - 12.9 16.8 - 17.1 19.2 - 19.3 28.7 - 30.8 33.5 - 35.7 41.9 - 42.7
7				31.7 - 31.7
9			26.8 - 27.5	31.6 - 31.6
10	56.3 - 58.6	59.5 - 60.2	28.6 - 28.6 36.4 - 36.8	26.2 - 26.2
total QTL detected per method:	4	9	24	26
population size:	224	20 per bulk	42 per bulk	20 per bulk
number of environments:	3	3	1	1
total number of markers:	1,174	2,102,988	40,942	37,767

Several other QTL were unique to the F_2 BSA_{GBS} analysis (Table 12). Considering the influence of population size, population structure, replication in phenotyping, and sequencing methods that are addressed in the discussion, these QTL may simply be false positives (Xu et al. 2008). Thus, to define a robust list of genomic loci associated with 3-DOA accumulation in pericarp, the following criteria were used: 1) QTL detected in all four mapping methods (chromosome 1, 71.0 – 73.1 Mb), 2) QTL detected in three mapping methods (chromosome 2, 8.2 -11.1 Mb), 3) Overlapping QTL found between the F_5 QTL_{GBS}, F_5 BSA_{WGS}, and F_2 BSA_{GBS-42} (chromosome 4, 51.7 – 62.2 Mb), and 4) the QTL residing less than 1 Mb apart on chromosome 10 found in the F_5 QTL_{GBS} and the F_5 BSA_{WGS} analyses (chromosome 10, 56.3 – 60.2 Mb) (Table 13).

Within the major QTLs detected on chromosome 1, 2, 4, and 10 (Table 13), a total of 2,209 genes are annotated in the sorghum reference genome (Sbicolor_454 v3.0.1). The genes within the QTL intervals mapped to a large number of GO categories that were too numerous to summarize. To parse this data, the 2,209 genes residing in QTL intervals were cross-referenced to a previously conducted transcriptome analysis between RTx3362 and BTx378 pericarp tissue exposed to UVB fluence for 10 days after anthesis (Fedenia et al. 2020). Of the 2,209 genes residing in trait loci, a total of 479 genes showed differential expression in the transcriptome (Table 13). Relevant terms revealed from Gene Ontology enrichment of the 479 DEGs included phenylalanine ammonia-lyase activity (GO:0045548), response to external stimulus (GO:0009605), response to extracellular stimulus (GO:0009991), oxidoreductase activity (GO:0016616), and defense response to fungus (GO:0050832) (Table 13). Seven

notable candidate genes were observed including a homocysteine S-methyltransferase (Sobic.001G453100), monooxygense 1 (Sobic.004G181000), flavonoid 3'-monooxygenase (Sobic.004G200833), flavodoxin protein WrbA-related (Sobic.004G185100), and three benzyl alcohol *O*-benzoyltransferases (Sobic.010G259300, Sobic.010G259500, Sobic.010G259600). Interestingly, the flavodoxin protein WrbA-related gene was one of the most highly upregulated genes observed in red pericarp in response to UVB light (upregulated over 2736-fold in BTx378) (Fedenia et al. 2020).

Table 13. Intervals, differential expression, and Gene Ontology terms of genes within major QTL detected among mapping methods.

Chromosome	Interval (Mb)	# Genes in Interval	# DEGs[a] in Interval[b]	Major Enriched Gene Ontology Terms Among DEGs
1	71.0 – 73.1	290	60	long-chain fatty acid metabolic process (GO:0000038) wax biosynthetic process (GO:0010025) cuticle development (GO:0042335) small molecule biosynthetic process (GO:0044283) organic acid biosynthetic process (GO:0016053) carboxylic acid biosynthetic process (GO:0046394)
2	8.2 -11.1	214	28	cell killing (GO:0001906) killing of cells of other organism (GO:0031640) disruption of cells of other organism (GO:0044364) modification of physiology of other organism (GO:0035821) defense response to fungus (GO:0050832) response to external stimulus (GO:0009605) response to extracellular stimulus (GO:0009991) quercetin 3-O-glucosyltransferase/7-O-gluc.activity (GO:0080043)
4	51.7 – 62.2	1199	289	cinnamic acid biosynthetic process (GO:0009800) negative regulation of flower development (GO:0009910) negative regulation of reproductive process (GO:2000242) regulation of protein modification process (GO:0031399) phenylalanine ammonia-lyase activity (GO:0045548) histone-lysine N-methyltransferase activity (GO:0018024) oxidoreductase activity (GO:0016616)
10	56.3 – 60.2	506	102	mitotic cell cycle phase transition (GO:0044772) cell cycle phase transition (GO:0044770) regulation of mitotic cell cycle (GO:0007346) lysosome (GO:0005764) lytic vacuole (GO:0000323)

[a]Differentially expressed genes
[b]DEGs detected between RTx3362 and BTx378 pericarp after 10 days exposure to UVB light (Fedenia et al. 2020)

Preliminary examination of WGS data from RTx3362 and BTx378 revealed a deletion of the coding region of Sobic.004G185100 (53,791,514-53,792,769 bp) in RTx3362 yet the full gene sequence is present in the BTx378 genome (Figure 10). PCR-based screening the genomic DNA of the phenotypic extremes of the RIL population revealed that Sobic.004G185100 was absent in 16 of the 20 highest F_5's for pericarp 3-DOA concentration in comparison to 7 of the 20 lowest F_5's for pericarp 3-DOA concentration (Figure 11). The gene belonging to the same gene family as Sobic.004G185100 but detected in both the BTx378 and RTx3362 genomes, Sobic.002G0074400, and a gene-specific PCR signal was detected in all but one RIL that accumulates low levels of 3-DOAs (Figure 11). In contrast to Sobic.004G185100, Sobic.002G0074400 was not found to be upregulated in either RTx3362 or BTx378 pericarp (Fedenia et al. 2020) and did not reside within a significant QTL interval (Table 13).

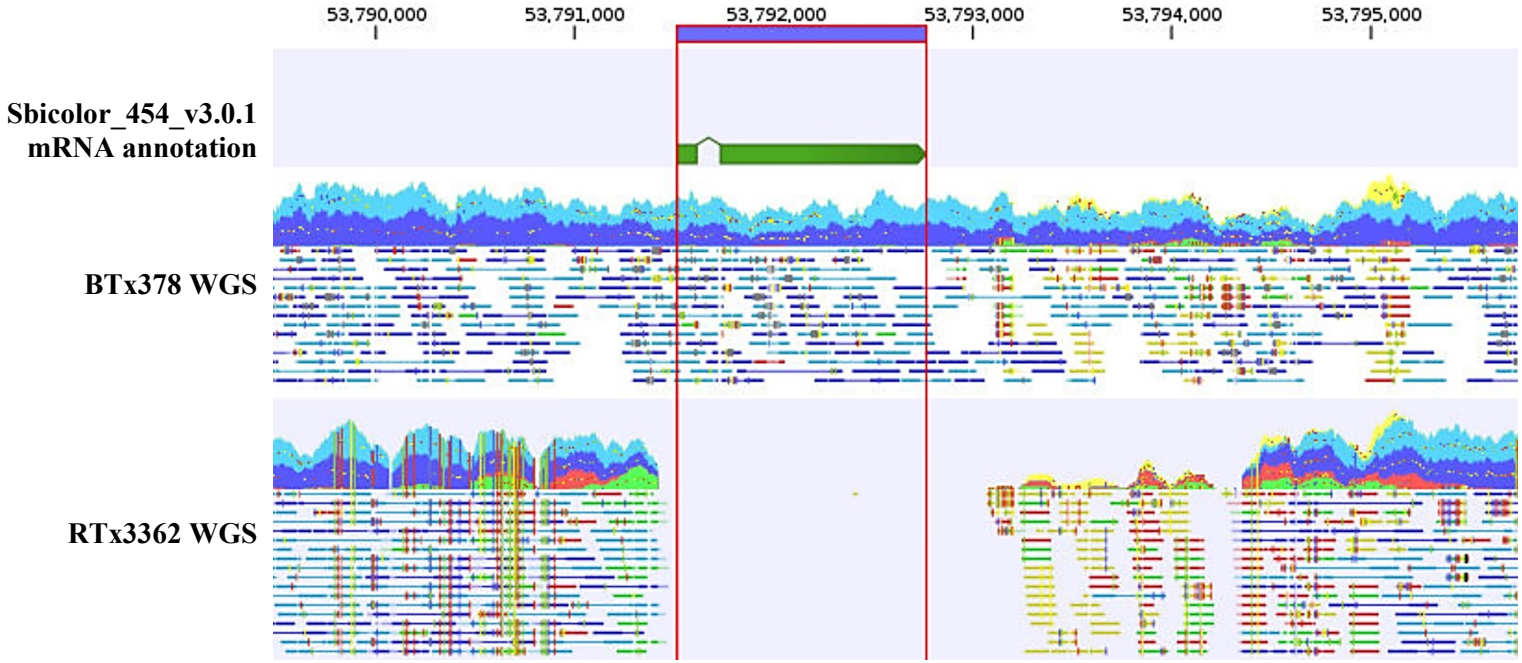


Figure 10. WGS reads from BTx378 and RTx3362 illustrating the deletion of Sobic.004G185100 gene in RTx3362 genome on chromosome 4.

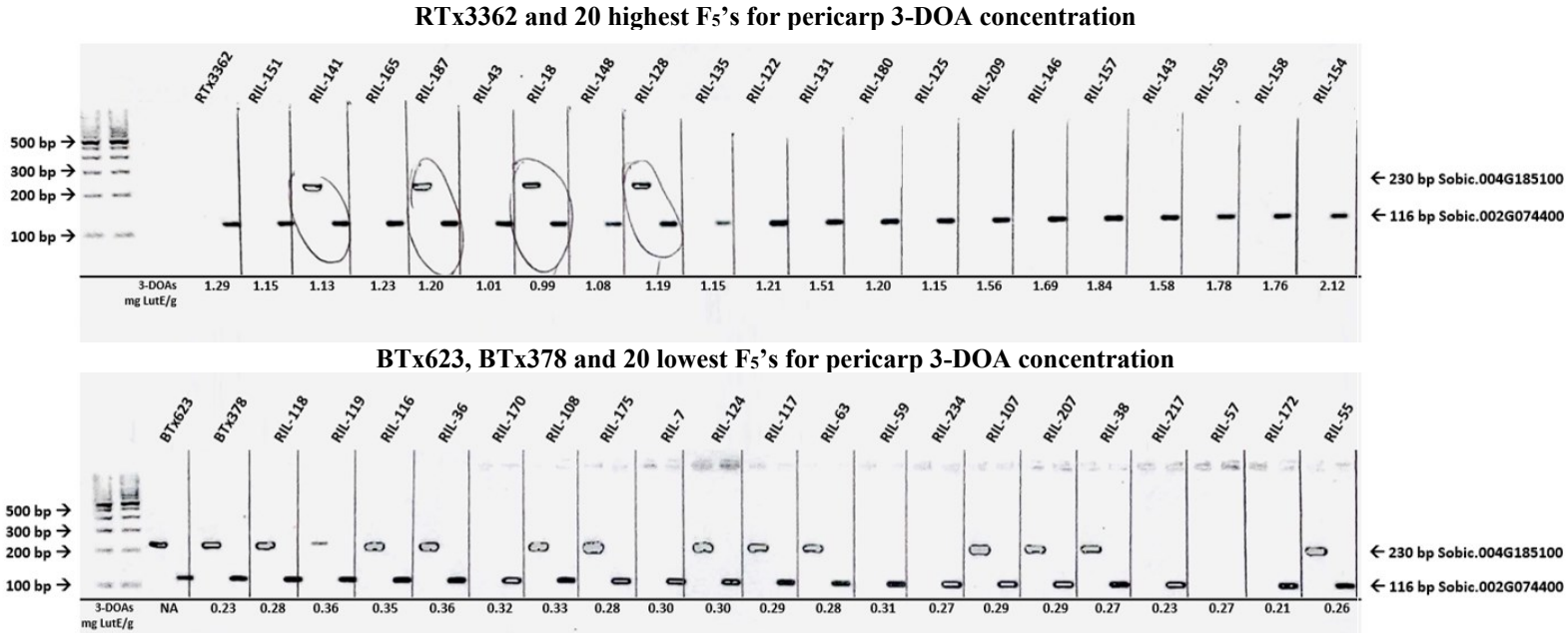


Figure 11. PCR screening for presence or absence of coding region sequences of Sobic.004G185100 (230 bp) and Sobic.002G0074400 (116 bp) genes in the 20 highest- and 20 lowest- F_5's for pericarp 3-DOA concentration. The upper panel depicts the PCR results for RTx3362 and the 20 highest F_5's for pericarp 3-DOA concentration and the lower panel depicts the results for BTx623, BTx378 and the 20 lowest F_5's for pericarp 3-DOA concentration.

Discussion

Accumulation of 3-DOAs in sorghum pericarp is a rare quantitative trait with complex inheritance and agronomic importance to the specialty grain industry (Pfeiffer and Rooney 2016). Moreover, the penetrance of this trait is highly influenced by environmental factors including the exposure to UVB light (Fedenia et al. 2020). These factors and others can hinder efforts to improve the trait by phenotypic selection while also complicating efforts to identify the genes controlling 3-DOA accumulation in sorghum pericarp. The black sorghum phenotype is, in part, a recessive trait and was previously estimated to be controlled by 2 to 12 genes (Pfeiffer and Rooney 2016). In the present research, 4 mapping approaches with different sequencing methods and population structure were used to increase detection power of loci responsible for 3-DOA accumulation in black sorghum pericarp. In addition to the mapping efforts conducted in this study, preliminary insight into the potential for epigenetic regulation of this trait was conducted by the removal of the environmental stimulus (UVB) during multiple generations of grain maturation to thereby determine if the genetic control is reversible.

Previous observations suggested that multiple generations of grain maturation under pollination bags may lessen the penetrance of the black pericarp phenotype under full sunlight (B Pfeiffer, personal communication) indicating potential reversibility of the genetic control of this trait. Epigenetic alterations influencing the phenotype via DNA methylation or histone modification are heritable and can persist in multiple generations of progeny. This phenomenon was observed in the epigenetic regulation of

phytochemical accumulation in response to UVB in high altitude maize landraces (Rius et al. 2016). Epigenetic control of UVB-induced flavonoid accumulation has also been observed in *Artemisia annua* where MYB1, MYBC, and a WRKY transcription factor were regulated by site-specific demethylation which caused overexpression of the gene encoding PAL (Pandey et al. 2019). Furthermore, epigenetic regulation of MYB domain transcription factors in response to abiotic and biotic stress in *Arabidopsis* has been previously observed (Roy 2016). The preliminary investigation into phenotype development after 8 generations of grain development in dark conditions revealed that $RTx3362_{blackout}$ grain under full sunlight was visibly black and indistinguishable from the grain of control RTx3362 panicles (Figure 7). Quantification of 3-DOA levels in mature grain from field-grown plants revealed an average 3-DOA concentration of $RTx3362_{blackout}$ grain of 1.63 mg LutE/g, which was similar to the average 3-DOA concentration (1.46 mg LutE/g) of control RTx3362 panicles. Based on these observations, multiple generations of pericarp development in the absence of the light-signal responsible for the expression of the black pericarp trait did not reverse or inhibit the penetrance of the trait under full sunlight and thus efforts were focused on the examination of differences in the DNA sequences between the black and red-seeded genotypes using the mapping methods described in this research. Although the present results do not provide evidence suggesting that the expression of the black pericarp phenotype is under epigenetic regulation, more detailed studies are required to definitively address the present conclusion.

To define high-confidence QTL that control the black pericarp trait, it was necessary to first examine the detection of what appeared to be spurious QTL. Amongst the mapping methods, the F_2 BSA_{GBS} analyses detected more QTL with less overlap/agreement when compared to the methods utilizing the F_5 population (Table 12). The statistical power in mapping QTL using GBS-BSA combined with QTLseqr is highly dependent on sufficient and/or optimized bulk pool sizes. Insufficient population sizes, despite phenotypic differences between the pools, results in a lower power to detect QTL and may only suffice to identify large-effect genes (Xu et al. 2008). Numerous studies have successfully utilized pooled DNA analyses (i.e. BSA) to detect a single major-effect gene (Barua et al. 1993; Hormaza et al. 1994; Villar et al. 1996; van Treuren 2001; Zhang et al. 2002), however, fewer reports have demonstrated the ability to detect more quantitatively-inherited trait involving more than two to three major loci (Quarrie et al. 1999). In analyses involving complex quantitative traits that are also dependent on the environment, BSA has served mainly as a complementary approach to reinforce the mapping of large-effect loci previously detected in traditional linkage map analysis (Takagi et al. 2013). In addition, the power of the BSA approach with F_2 plants is reduced by the inherent limitation associated with mapping in a F_2 population, which include: the limited number of generations of meiotic events (limited to F_1), the high level of heterozygosity, and F_2 populations cannot be phenotyped in replicated studies. Due to these limitations associated with mapping a complex trait in an F_2 population, QTL unique to the F_2 BSA mapping analysis may be false positives and deemphasized in further candidate gene investigation (Xu et al. 2008). In addition to population size and

accurate phenotyping, marker density and coverage influence BSA results (Xu et al. 2008). The use of WGS data in BSA greatly increases the density and coverage of molecular markers across the genome to aid in the refinement of trait loci where marker density is limited. Despite the smaller pool sizes used in the F_5 BSA_{WGS}, the increased marker density and coverage in combination with replicated phenotyping efforts increases the probability that at least one SNP will be polymorphic in the target gene or nearby. It appears that the increased marker density deployed in the F_5 BSA_{WGS} was more agreeable with the results obtained from the traditional linkage map analysis with less detection of QTL solely detected in the F_5 BSA_{WGS} analysis.

Although there are strengths and weaknesses to the multiple BSA analyses conducted, the additional mapping methods were generally supportive in confirming large effect loci. Four QTL intervals were detected on chromosomes 1, 2, 4, and 10 in this study (Table 13). Within these intervals, a total of 2209 genes are annotated in the sorghum BTx623 reference genome (Sbicolor_454 v3.0.1). A total of 479 genes residing in trait loci showed differential expression in black and red pericarp in previous analyses (Fedenia et al. 2020) (Table 13). Preliminary examination of WGS data from RTx3362 and BTx378 of DEGs within QTL intervals revealed a deletion of the coding region of Sobic.004G185100 (53,791,514-53,792,769 bp) in RTx3362 yet the full gene sequence is present in the BTx378 genome (Figure 10). Screening of RILs with phenotypic extreme 3-DOA levels revealed that this gene was also deleted in 16 of the 20 RILs with the highest pericarp 3-DOA content. While not absolute, the results indicate that the deletion of this gene may be an important genetic factor in controlling

the black pericarp phenotype although other genetic factors are obviously critical for full penetrance of this trait. Sobic.004G185100 is annotated as a flavodoxin protein WrbA - related, a family of proteins involved in cellular response to altered redox conditions and other stressors (Grandori and Carey, 1994). A deletion of a gene involved in plant stress response supports the previously described hypothesis that the black sorghum grain phenotype is produced in response to cellular stress (Fedenia et al. 2020). The expected physiological impact of the absence of this gene in RTx3362 would likely result in the reduced ability to detoxify accumulating reactive oxygen species (ROS) induced from abiotic stress. Surpassing a cellular concentration threshold of ROS is known to trigger flavonoid production in *Zea mays* organs (Casati and Walbot 2004) and *Arabidopsis thaliana* leaves (Fini et al. 2011). This provides evidence of a potential mechanism of induction of 3-DOA accumulation in black sorghum pericarp tissue (see Chapter IV). It is well characterized in the literature that the chemical structure of 3-DOAs confers a superior ability to detoxify free radicals due to the dihydroxy B-ring structure in comparison to the mono-hydroxy substituted ring which is common among most plant polyphenols (Fini et al. 2011). The preferential accumulation of flavonoids with a dihydroxy B-ring structure supports the notion that 3-DOAs are compounds generated as a proxy response to conduct cellular detoxification when other detoxification methods are insufficient.

In addition to the deletion of a gene critical for cellular detoxification, an important transcription factor resides near the QTL interval detected on chromosome 1 across the four mapping methods (Table 13). A previous study proposed that *Yellow*

seed 1 (*Y1*), an R2R3 MYB transcription factor that is homologous to maize *Pericarp color 1* (*P1*), is responsible for 3-DOA production in sorghum pericarp and vegetative tissue (Boddu et al 2005). The *Y1* gene resides approximately 3 Mb upstream of the major QTL detected on chromosome 1 (71.0-73.1 Mb, Table 13). It was recently reported that the recessive *y1* allele corresponds to Sobic.001G398100, and the recessive allele harbors a 3,218 bp deletion annotated in the current sorghum reference genome (Sbicolor_454 v3.0.1) (Nida et al. 2019). The physical location of this recessive allele is positioned on chromosome 1 from 68,399,401-68,400,602 bp. The suggested position of the dominant allele resides at approximately 68,398,215-68,403,823 bp, however, this has not been confirmed nor properly annotated in the published reference genome (Nida et al. 2019). Although the QTL interval on chromosome 1 did not encompass the region predicted to harbor the *Y1* gene, this could potentially be attributed to utilization of a BTx623 reference genome (Sbicolor_454 v3.0.1) missing a functional *Y1* gene annotation for alignment of the sequencing reads. To conduct an initial investigation of misalignment of the resequencing reads to the BTx623 reference genome, structural variation was examined among WGS data for both BTx378 and RTx3362. Preliminary results detected potential structural variation where the functional *Y1* gene is predicted to reside. Due to the proximity of the QTL detected on chromosome 1 and the suggested location of the functional *Y1* gene and potential misalignment occurring as a result of an incomplete reference genome, further inquiry is warranted to determine if *Y1* is a critical gene contributing to 3-DOA accumulation in sorghum pericarp tissue. Future studies involving long-read sequencing technology and whole genome assembly with annotation

may resolve this locus in the genome of black grain sorghum and assist in determining the relationship of structural variants on the control of black pericarp phenotype.

Additional candidate genes were observed including a homocysteine S-methyltransferase (Sobic.001G453100), monooxygense 1 (Sobic.004G181000), flavonoid 3'-monooxygenase (Sobic.004G200833), and three benzyl alcohol *O*-benzoyltransferases (Sobic.010G259300, Sobic.010G259500, Sobic.010G259600). Homocysteine S-methyltransferase is known to be involved in ethylene signaling and defense in response to plant stress (Zhu et al. 2018) and the QTL detection of this gene may suggest that accumulation of 3-DOAs in black sorghum pericarp involves ethylene signaling pathways. Monooxygenase and flavonoid 3'-monooxygenase determine the hydroxylation pattern of flavonoids and may relate to the unique hydroxylation pattern observed in 3-DOAs (Rhodes et al. 2014). Finally, benzyl alcohol *O*-benzoyltransferases are responsible for anthramide phytoalexin production in *Dianthus* (Yang et al. 1997) and may serve a similar purpose in the accumulation of 3-DOA phytoalexin accumulation. Further exploration and validation of the involvement of these candidate genes is necessary to determine their role in the accumulation of 3-DOAs in black sorghum pericarp tissue.

The primary objective of the mapping efforts conducted in this research was to identify genomic regions and discover candidate genes that condition 3-DOA accumulation in black sorghum grain. In summary, various QTL were discovered across the sorghum genome using a summation of mapping techniques and require further refinements to conclusively declare candidate genes in these regions. A more complete

investigation into the expression of candidate genes underlying the QTL to clarify the impact of polymorphisms detected in this research is the subject of the next chapter. Future efforts to refine the genetic linkage map, obtain a *de novo* genome assembly of a black pericarp sorghum, and validate candidate genes through various molecular techniques is warranted. The present research serves as a foundational resource to further understand the cellular and molecular mechanisms responsible for 3-DOA accumulation in sorghum grain that may ultimately contribute to marker-assisted breeding efforts to maximize the expression of 3-DOAs in sorghum pericarp tissue.

CHAPTER IV

CANDIDATE GENE NETWORK FOR ORGAN-SPECIFIC REGULATION OF 3-DOA PRODUCTION IN BLACK SORGHUM PERICARP TISSUE

Synopsis

Sorghum [*Sorghum bicolor* (L.) Moench] is a cereal of tropical origin with specialty grain-types that can serve as an important dietary source of diverse plant polyphenols (Awika 2004; Gous 1989). Sorghum grains are typically red, yellow, or white, and this coloration is associated with specific grain polyphenol profiles (Dykes 2008). Unique amongst the common grain sorghums are the black grain genotypes that are all descendants of the accession 'Black Shawaya' from the Sudan (Hahn et al. 1984). The black pigmentation of 'Black Shawaya' grain is associated with the accumulation of rare polyphenols, 3-deoxyanthocyanidins (3-DOAs) (Awika et al. 2004), and elite black grain cultivars and hybrids with high 3-DOA grain concentration have been developed for commercial production and utilization in the health food market (Rooney et al. 2013). With an increasing interest from the food and health industries to utilize 3-DOAs as nutraceutical additives (Xiong et al. 2019), development of new sorghum cultivars with high 3-DOA concentration as well as exploration of the genes and gene networks controlling this complex trait are warranted.

This research aimed to elucidate those genes and gene networks that regulate the cascade of cellular events leading to the black pericarp phenotype and associated 3-DOA accumulation in pericarp in response to UVB light. Utilizing Weighted Gene Co-

expression Network Analysis (WGCNA), a detailed gene expression analysis was performed using the transcriptomes of pericarp, leaf, and root tissue from black- and red-grain sorghum genotypes under different light regimes, and through this study a better understanding of the regulatory network governing early induction of pericarp-specific 3-DOA accumulation was obtained. From the collection of co-expressed genes correlated with early induction of 3-DOA accumulation in the black- and red-seeded parents, we propose a pathway modeling 3-DOA induction in black sorghum pericarp tissue in which UVB triggers the generation of endogenous elicitors that initiate a signal cascade resulting in the high expression of flavonoid biosynthetic genes ultimately responsible for 3-DOA synthesis. The intricate cascade of subcellular events leading to the genotypic- and tissue-specific accumulation of these phytochemicals in black-seeded sorghum may advance future efforts to expand biological production and chemical synthesis of 3-DOA for broad applications in the food and nutraceutical industry.

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] grain is a staple cereal crop worldwide and valued for its remarkable phenotypic plasticity to abiotic and biotic stressors (Mundia et al. 2019). In contrast to other cereal crops where bitter or pigmented compounds were often selected against, the superior adaptability of sorghum is often attributed to the crop's diverse polyphenol profiles which persisted throughout domestication (Olsen et al. 2013; Morris et al. 2013). Classified among these polyphenols are 3-deoxyanthocyanidins (3-DOAs), an important family of phytoalexins generated in response to pathogen and environmental stresses (Chopra et al. 2002). The

production of 3-DOAs is rare compared to other more common flavonoids and at present, sorghum represents the only dietary source of these compounds (Awika et al. 2004). The natural biosynthesis of 3-DOAs has attracted interest from the food and health industries for their unique chemical properties which confers potent antioxidant activity and stability for use in natural foods (Awika et al. 2004; Awika 2008; Shih et al. 2007). Methods to chemically synthesize 3-DOAs have been developed but are complex and low-yielding (Sweeny and Iacobucci 1981; Xiong et al. 2019) thus understanding the *in planta* production of these compounds is central to future applications of 3-DOAs in the food and nutraceutical industries.

The biosynthesis of 3-DOAs has been observed in both vegetative and reproductive tissues in sorghum, and the environmental signals and genetic mechanisms that condition 3-DOA biosynthesis is both genotype- and organ-specific. Accumulation of 3-DOAs in vegetative tissues is a common and dominant trait in sorghum with the induction of 3-DOA biosynthesis occurring in response to pathogen/insect attack or mechanical wounding (Chopra et al. 2002; Hipskind et al. 1990; Kumar et al. 2015; Viswanathan et al. 1996). In vegetative tissues, 3-DOA locally forms as non-pigmented inclusions in the cytoplasm of the cells in response to the release of fungal cell wall degrading enzymes (Nielsen et al. 2004; Snyder and Nicholson 1990). The inclusion bodies migrate to the site of attempted penetration where they are transported across the plasma membrane and adhere to the partially hydrolyzed plant cell wall (Nielsen et al. 2004). During this phase, the inclusions burst, develop pigmentation, and ultimately

cause host cell collapse that is associated with a localized hypersensitive response (Nielsen et al. 2004).

In contrast to 3-DOAs in sorghum vegetative tissues, the inheritance of 3-DOA biosynthesis in pericarp of black grain sorghum is complex (recessive with significant additive, dominance and epistatic effects), and the accumulation of pericarp 3-DOAs is in response to the environmental signal of high-fluence UVB light (Fedenia et al. 2020). Similar to stress-induced 3-DOA accumulation associated with the hypersensitive response (HR) of vegetative tissues, an exploratory transcriptome analysis revealed that prolonged (10 day) UVB exposure of black sorghum pericarp activated the expression of genes related to plant defense, ROS, and secondary metabolism, suggesting that 3-DOA accumulation is associated with several overlapping defense and stress signaling pathways (Fedenia et al. 2020). Thus, while the environmental signals regulating 3-DOA accumulation in pericarp vs. vegetative tissues may differ, the appearance of colored localized lesions on vegetative tissues and the black appearance of pericarp tissue directly exposed to UVB represent tissue-specific stress responses to different environmental stressors.

An efficient approach to unveil gene networks and regulatory associations underlying traits of interest is to use Weighted Gene Co-expression Network Analysis (WGCNA). WGCNA detects co-expressed gene networks from RNA-seq datasets that are correlated to a phenotypic trait of interest. This approach has successfully discovered co-expressed genes responsible for fruit quality traits across many species including in apricot (Zhang et al. 2019), apple (Li et al. 2020), watermelon (Umer et al.

2020), and tomato (Mandal et al. 2020). In this study, a network analysis was performed using the transcriptomes of pericarp, leaf, and root tissue from black- and red-seeded sorghum genotypes to obtain a more detailed understanding of the regulatory networks governing early induction of pericarp-specific 3-DOA accumulation. To permit a robust and informative WGCNA, 64 RNA-seq libraries were constructed from vegetative and pericarp tissues exposed to different light regimes (durations and light spectra), while also contrasting pericarp gene expression from a red- and black-seed genotype. From the collection of co-expressed genes correlated with early induction of 3-DOA accumulation between the black- and red-seed genotypes, this research proposes a pathway modeling 3-DOA induction in black sorghum pericarp tissue in which UVB triggers the generation of endogenous elicitors that initiate a signal cascade resulting in high expression of flavonoid biosynthetic genes ultimately responsible for 3-DOA accumulation.

Materials and Methods

Comparative Transcriptomics of Sorghum Pericarp, Flag leaf, and Root Samples

A transcriptome analysis was conducted on pericarp tissue of genotypes RTx3362 (black-seeded, high 3-DOA seed concentration) and BTx378 (red-seeded, low 3-DOA seed concentration) grown in controlled-environment chambers under a series of light regimes as previously detailed (Fedenia et al. 2020). In the present transcriptomics examination, RTx3362 was grown in growth chambers with panicles exposed during grain maturation to light regimes consisting of visible light (VIS), VIS supplemented with UVA, and VIS supplemented with UVA and UVB (or a no light control). For

genotypic comparison, BTx378 was grown under VIS supplemented with UVA plus UVB. Three biological replicates were obtained per each light treatment for genotype RTx3362 and two biological replicates were conducted for BTx378. The entire study was repeated although the BTx378 genotype was excluded from the second trial. For transcriptome analysis of pericarp tissue, rachis branches were removed at days 10 (T10) and 17 (T17) after the initiation of light spectral treatments, at which time the pericarp was mechanically excised and prepared for total RNA extraction following previously described methods (Fedenia et al. 2020). To examine organ-specific gene expression, tissue was collected from flag leaves and root tissue from RTx3362 and BTx378. Leaf tissue was collected using a razor blade to cut 1-inch squares from plants exposed to VIS supplemented with UVA and UVB at timepoint T10 and immediately submerged in LN_2 and stored at -80 °C. Root radicals were harvested from ~30 etiolated seedlings of RTx3362 and BTx378 (grown on germination paper, 30 °C for 4 d). Root radicals were immediately submerged in LN_2 and stored at -80 °C. Leaf and root tissues were pulverized with a mortar and pestle under LN_2 prior to RNA extraction. Each pericarp, leaf, and root sample from which RNA-seq libraries were obtained are summarized in Table 14.

Procedures for RNA isolation, sequencing, and differential gene expression analysis were followed as previously described (Fedenia et al. 2020). Confounding variables of biological experimental replication, date of extraction, and sequencing provider were controlled for during differential gene expression calculations. Differentially expressed genes (DEGs) were determined using a false discovery rate

(FDR) of $p < 0.05$ and a fold change cutoff of log2 ≥ 1 or ≤ −1 on a total of 31 pairwise comparisons summarized in Table 15. Biological function analysis was performed on DEGs using the gene ontology (GO) tool from PlantTFDB 4.0 (Jin et al. 2016) according to the *Sorghum bicolor* annotation of genotype BTx623 (Sbicolor_454 v3.0.1; www.phytozome.jgi.doe.gov). Significant GO terms were found using an FDR of $p < 0.05$. Accumulation of 3-DOAs in all samples included in the RNA-seq analysis was quantified using Ultraperformance Liquid Chromatography with Photodiode Array Detection (UPLC-PDA) following previously detailed methodology (Fedenia et al. 2020).

Table 14. Summary table of samples used for RNA-seq.

Genotype	Organ	Sampling Time	Light Treatment	Biological Reps
RTx3362	Pericarp	T10[a]	No light	6
		T17[b]	No light	6
		T10	VIS	6
		T17	VIS	6
		T10	VIS+UVA	6
		T17	VIS+UVA	6
		T10	VIS+UVA+UVB	6
		T17	VIS+UVA+UVB	6
	Flag leaf	T10	VIS+UVA+UVB	3
	Root tips	--	No light	3
BTx378	Pericarp	T10	VIS+UVA+UVB	2
		T17	VIS+UVA+UVB	2
	Flag leaf	T10	VIS+UVA+UVB	3
	Root tips	--	No light	3

[a]10 days after the initiation of light spectral treatments
[b]17 days after the initiation of light spectral treatments

Weighted Gene Co-Expression Network Analysis

WGCNA was constructed from normalized expression data from pericarp, leaf, and root samples using a step-by-step method implemented in WGCNA R package (Langfelder 2008). RNA-seq datasets utilized for WCGNA analysis included an overall

total of 64 RNA-seq datasets including 52 pericarp, 6 flag leaf, and 6 root samples (Table 14). The network was constructed using transcriptome sequencing data expressed as transcripts per kilobase million (TPM). To mathematically represent the gene co-expression network, a pairwise gene correlation matrix was constructed using Pearson correlation analysis and transformed into a weighted matrix with soft threshold (β) value of 6 and assuming networks were scale-free. A dendrogram was generated with a hierarchical clustering method from the weighted matrix (adjacency matrix). Clusters of genes with similar expression patterns were identified by a dynamic tree-cut algorithm and referred to as modules. To represent the gene expression profiles of each module, a module eigengene was defined. A module membership was then calculated for each gene as the correlations between the expression profile of that gene and the module eigengene. To capture modules related to transcriptional regulation of 3-DOA production in pericarp tissue, UPLC-PDA phenotypic data was substituted with categorical ranking to describe the treatment of interest (100 = RTx3362 pericarp exposed to VIS supplemented with UVA and UVB at timepoint T10) compared to all other pericarp treatments and tissue types (0 = all other samples). All samples were assumed to be independent of each other. The correlation between external trait data and module eigengenes were determined and gene significance for each gene within its respective module was calculated as the correlation between expression of the gene and trait data. The strongest connected genes based on intramodular connectivity generated by WGCNA for the trait of interest were visualized using Cytoscape v3.7.1. Co-

expressed genes related to the trait of interest that also possessed the largest number of connections (edges) were considered the most critical network nodes.

Conserved Motif Distribution Analysis of Hub Genes

To discover highly conserved cis regulatory elements (CRE) in genes belonging to the same co-expression module identified by network analysis, the sequences of the hub genes were extracted from the WGS data from the RTx3362 genome mapped to the current BTx623 sorghum reference genome (Sbicolor_454 v3.0.1). The promoter (~3-kb upstream from annotated 5'UTR) region was subjected to Multiple Expectation Maximization for Motif Elicitation analysis (MEME Suite; http://meme-suite.org/tools/meme; Baily et al. 2009) using a motif width setting < 50 bp and > 6 bp, maximum motif number of 20, and any number of site repetition for site distributions. Query motifs identified from the set of DNA sequences were subsequently analyzed for functional relevance using the Tomtom motif comparison tool (http://meme-suite.org/tools/tomtom; Gupta et al. 2007) against the database JASPAR (non-redundant) DNA and JASPAR CORE (2018) Plants, with a significance threshold E-value < 0.05 and Pearson correlation coefficient for the motif column comparison.

Results

Comparative Transcriptomics of Pericarp, Flag leaf, and Root Tissues

To examine the cascade of cellular events leading to the biosynthesis of 3-DOAs in black grain sorghum, changes in pericarp gene expression were quantified over a time course of treatment exposure. Pericarp was sampled on day 10 after full anthesis at which pericarp darkening was visually first apparent and on day 17 when maximum

flavonoid accumulation appears to occur during seed maturation (Fedenia et al. 2020). The series of light spectral regimes to which panicles of BTx378 and RTx3362 genotypes were exposed permitted 19 pairwise comparisons of pericarp gene expression (Table 15 & Table S2). An additional 12 pairwise comparisons permitting organ-specific gene expression were conducted using pericarp, leaf, and root tissue (Table 15 & Table S2). A total of 2.673×10^9 reads with an average of 5.35×10^7 reads per each of the 64 tissue samples were obtained after adaptor and low-quality paired-end sequences were trimmed (Table S1). An average of 45.7×10^7 clean paired-end reads per sample were aligned to the BTx623 sorghum reference genome (Sbicolor_454 v3.0.1) using the RNA-seq Analysis tool within the CLC Genomics Workbench (Table S1). Analysis of gene expression revealed a total of 164,186 DEGs across the 31 comparisons and the number of DEGs per pairwise comparison ranged from 0 to 15,155 (FDR $p < 0.05$, fold-change cutoff of $\log2 \geq 1$ or ≤ -1; Table S2). The 31 pairwise DEG analyses examined in this research are summarized in Table 15.

Table 15. RNA-seq samples compared for DEG analysis.

Sample 1				vs.	Sample 2				
Genotype	**Tissue**	**Timepoint**	**Treatment**		**Genotype**	**Tissue**	**Timepoint**	**Treatment**	**Total DEGs**
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Pericarp	T10	No light	729
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Pericarp	T10	VIS	100
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Pericarp	T10	VIS+UVA	53
RTx3362	Pericarp	T10	VIS+UVA+UVB		BTx378	Pericarp	T10	VIS+UVA+UVB	6866
RTx3362	Pericarp	T10	VIS		RTx3362	Pericarp	T10	No light	648
RTx3362	Pericarp	T10	VIS+UVA		RTx3362	Pericarp	T10	No light	897
RTx3362	Pericarp	T10	VIS+UVA		RTx3362	Pericarp	T10	VIS	2
RTx3362	Pericarp	T17	VIS+UVA+UVB		RTx3362	Pericarp	T17	No light	2835
RTx3362	Pericarp	T17	VIS+UVA+UVB		RTx3362	Pericarp	T17	VIS	391
RTx3362	Pericarp	T17	VIS+UVA+UVB		RTx3362	Pericarp	T17	VIS+UVA	176
RTx3362	Pericarp	T17	VIS+UVA+UVB		BTx378	Pericarp	T17	VIS+UVA+UVB	9050
RTx3362	Pericarp	T17	VIS		RTx3362	Pericarp	T17	No light	3467
RTx3362	Pericarp	T17	VIS+UVA		RTx3362	Pericarp	T17	No light	1701
RTx3362	Pericarp	T17	VIS+UVA		RTx3362	Pericarp	T17	VIS	0
RTx3362	Pericarp	T10	No light		RTx3362	Pericarp	T17	No light	327
RTx3362	Pericarp	T10	VIS		RTx3362	Pericarp	T17	VIS	3120
RTx3362	Pericarp	T10	VIS + UVA		RTx3362	Pericarp	T17	VIS + UVA	597
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Pericarp	T17	VIS+UVA+UVB	1840
BTx378	Pericarp	T10	VIS+UVA+UVB		BTx378	Pericarp	T17	VIS+UVA+UVB	529
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Flag leaf	T10	VIS+UVA+UVB	14004
RTx3362	Pericarp	T10	VIS+UVA+UVB		RTx3362	Root	--	No light	13844
RTx3362	Pericarp	T17	VIS+UVA+UVB		RTx3362	Flag leaf	T10	VIS+UVA+UVB	13486
RTx3362	Pericarp	T17	VIS+UVA+UVB		RTx3362	Root	--	No light	14122
RTx3362	Flag leaf	T10	VIS+UVA+UVB		RTx3362	Root	--	No light	15155
BTx378	Pericarp	T10	VIS+UVA+UVB		BTx378	Flag leaf	T10	VIS+UVA+UVB	10309
BTx378	Pericarp	T10	VIS+UVA+UVB		BTx378	Root	--	No light	11060
BTx378	Pericarp	T17	VIS+UVA+UVB		BTx378	Flag leaf	T10	VIS+UVA+UVB	11078
BTx378	Pericarp	T17	VIS+UVA+UVB		BTx378	Root	--	No light	11374
BTx378	Flag leaf	T10	VIS+UVA+UVB		BTx378	Root	--	No light	13857
RTx3362	Flag leaf	T10	VIS+UVA+UVB		BTx378	Flag leaf	T10	VIS+UVA+UVB	1513
RTx3362	Root	--	No light		BTx378	Root	--	No light	1056

Genotypes BTx378 and RTX3362 were analyzed separately for gene expression between pericarp exposed to VIS + UVA + UVB, leaf, and root tissue. A total of 5,965 and 5,541 genes were upregulated in RTx3362 pericarp treated with VIS + UVA + UVB at T10 and T17 when compared to gene expression in leaf and root tissue of RTx3362, respectively (Table 15 & Table S2). A total of 2,731 DEGs were upregulated in RTx3362 pericarp treated with VIS + UVA + UVB at T10 and T17 in comparison to gene expression in both leaf and root tissue of RTx3362 (Table 15 & Table S2). A total of 4,970 and 4,578 genes were upregulated in BTx378 pericarp treated with VIS + UVA + UVB at T10 and T17 when compared to gene expression in leaf and root tissue of BTx378, respectively (Table 15 & Table S2). A total of 2,242 DEGs were upregulated in BTx378 pericarp treated with VIS + UVA + UVB at T10 and T17 in comparison to gene expression in both leaf and root tissue of BTx378 (Table 15 & Table S2).

The DEGs in the analyses detailed above mapped to a large number of GO categories that were too numerous to summarize. To parse this expression data, the genes uniquely upregulated in pericarp vs. vegetative tissue for each genotype were compared. Of the genes uniquely upregulated in RTx3362 pericarp (vs. RTx3362 vegetative tissue) compared to those uniquely upregulated in BTx378 pericarp (vs. BTx378 vegetative tissue), 1,670 were upregulated in pericarp tissue of both genotypes. A total of 1,126 DEGs were unique to RTx3362 pericarp tissue (vs. RTx3362 vegetative tissue) and were not found differentially expressed in BTx378 pericarp tissue (vs. BTx378 vegetative tissue). The vast majority of the shared DEGs related to GO terms of biological and cellular regulation (GO:0065007; GO:0050794). In contrast, a large

proportion of the 1,126 DEGs unique to RTx3362 pericarp were related to terms response to abiotic stimulus (GO:0009628) and response to endogenous stimulus (GO:0009719).

Pericarp gene expression in response to light spectra was analyzed. In RTx3362, gene expression was examined in six pairwise comparisons: RTx3362 pericarp treated with VIS + UVA + UVB vs. No light at T10; VIS + UVA + UVB vs. VIS at T10; VIS + UVA + UVB vs. VIS + UVA at T10; VIS + UVA + UVB vs. No light at T17; VIS + UVA + UVB vs. VIS at T17; VIS + UVA + UVB vs. VIS + UVA at T17. A total of 9 genes were uniquely and consistently upregulated in RTx3362 pericarp treated with VIS + UVA + UVB in each of the six pairwise comparisons suggesting that their expression is responsive to the exposure to UVB light. These genes were unannotated in GO enrichment and encoded the following proteins: glucan endo-1,3-beta-D-glucosidase (Sobic.003G422200), tryptophan decarboxylase (Sobic.003G309700), two pathogenesis-related proteins (Sobic.005G169400; Sobic.008G182900), thaumatin-like pathogenesis-related protein 4 precursor (Sobic.008G182800), early light-inducible protein ELIP (Sobic.002G052100), glutathione S-transferase GST 29 (Sobic.003G426500), and two uncharacterized proteins (Sobic.003G339200; Sobic.002G258500). Seven of these genes (Sobic.003G422200, Sobic.003G339200, Sobic.003G309700, Sobic.005G169400, Sobic.002G258500, Sobic.008G182800, Sobic.002G052100, Sobic.008G182900, Sobic.003G426500) were also upregulated in RTx3362 pericarp treated with VIS + UVA + UVB at T10 and T17 compared to BTx378 pericarp treated with VIS + UVA + UVB at T10 and T17.

Weighted Gene Correlation Network Analysis

WGCNA was used to screen for co-expressed genes regulating 3-DOA synthesis in pericarp tissue. Out of the twenty-two modules observed, 42 genes within the *darkred* module were most associated (0.74 correlation; p = 8e-09; Table S3) with treatment RTx3362 pericarp exposed to VIS + UVA + UVB at T10. Within the *darkred* module, Gene Significance (GS) and Module Membership (MM) was moderately correlated based on examination of the p value (5.7e-05).

To perform biological function predictions, GO enrichment was conducted on the 42 co-expressed genes and revealed 24 significantly enriched terms (Table 16). Genes functioning in secondary metabolic biosynthetic process (GO:0044550) included monooxygenase (Sobic.004G181000), cytochrome P450 84A1(Sobic.001G196300), cytochrome P450 73A100 (Sobic.004G141200), and cytochrome P450 71A1 (Sobic.009G064400) (Table 16). Five genes were found to have regulatory roles (several GO terms) including pathogenesis-related protein 10a (Sobic.001G400900), pectinesterase Inhibitor 8 (Sobic.009G044100), and bowman-birk serine protease inhibitors (Sobic.003G355800, Sobic.005G215900, Sobic.005G216000). Interestingly, several of the significantly co-expressed terms discovered were members belonging to the same gene families (ABC transporters, hydroxycinnamoyltransferases, benzyl alcohol *O*-benzoyltransferase, pathogenesis-related proteins, MtN26 family proteins, cytochrome P450s, bowman-birk serine protease inhibitors). For example, three genes with activity related to response to biotic stimuli (GO:0009607) belonged to the pathogenesis-related protein family (Sobic.001G400700, Sobic.001G400800,

Sobic.001G400900) (Table 16). No GO terms were enriched for response to UV light, high light intensity, or abiotic stress.

Table 16. Gene ontology (GO) terms enriched in 42 co-expressed genes discovered in network analysis.

GO ID	Term	Count	q-value[a]	Aspect[b]	Genes
GO:0043086	negative regulation of catalytic activity	4	1.00E+00	P	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900, Sobic.009G044100
GO:0044092	negative regulation of molecular function	4	1.00E+00	P	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900, Sobic.009G044100
GO:0044550	secondary metabolite biosynthetic process	4	1.00E+00	P	Sobic.001G196300, Sobic.004G141200, Sobic.004G181000, Sobic.009G064400
GO:0051336	regulation of hydrolase activity	3	1.00E+00	P	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900
GO:0019748	secondary metabolic process	4	1.00E+00	P	Sobic.001G196300, Sobic.004G141200, Sobic.004G181000, Sobic.009G064400
GO:0065009	regulation of molecular function	4	1.00E+00	P	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900, Sobic.009G044100
GO:0010466	negative regulation of peptidase activity	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0010951	negative regulation of endopeptidase activity	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0052547	regulation of peptidase activity	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0052548	regulation of endopeptidase activity	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0044711	single-organism biosynthetic process	7	1.00E+00	P	Sobic.001G196300, Sobic.001G453100, Sobic.003G286600, Sobic.004G141200, Sobic.004G181000, Sobic.009G064400, Sobic.010G092200
GO:0009607	response to biotic stimulus	4	1.00E+00	P	Sobic.001G400700, Sobic.001G400800, Sobic.001G400900, Sobic.004G181000
GO:0051346	negative regulation of hydrolase activity	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0045861	negative regulation of proteolysis	2	1.00E+00	P	Sobic.003G355800, Sobic.005G215900
GO:0005576	extracellular region	4	1.00E+00	C	Sobic.003G244600, Sobic.003G355800, Sobic.005G215900, Sobic.005G216000
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	7	6.93E-03	F	Sobic.005G137200, Sobic.005G214400, Sobic.010G066601, Sobic.010G066700, Sobic.010G259300, Sobic.010G259500, Sobic.010G259600
GO:0004857	enzyme inhibitor activity	5	3.11E-02	F	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900, Sobic.005G216000, Sobic.009G044100
GO:0005506	iron ion binding	6	5.92E-02	F	Sobic.001G196300, Sobic.003G286600, Sobic.004G141200, Sobic.004G200833, Sobic.009G064400, Sobic.010G092200
GO:0004867	serine-type endopeptidase inhibitor activity	3	1.37E-01	F	Sobic.003G355800, Sobic.005G215900, Sobic.005G216000
GO:0030234	enzyme regulator activity	5	1.37E-01	F	Sobic.001G400900, Sobic.003G355800, Sobic.005G215900, Sobic.005G216000, Sobic.009G044100
GO:0004866	endopeptidase inhibitor activity	3	1.37E-01	F	Sobic.003G355800, Sobic.005G215900, Sobic.005G216000
GO:0030414	peptidase inhibitor activity	3	1.37E-01	F	Sobic.003G355800, Sobic.005G215900, Sobic.005G216000
GO:0004497	monooxygenase activity	4	5.42E-01	F	Sobic.001G196300, Sobic.004G141200, Sobic.004G181000, Sobic.009G064400
GO:0016705	oxidoreductase activity	4	9.53E-01	F	Sobic.001G196300, Sobic.004G141200, Sobic.004G200833, Sobic.009G064400

Gene expression within each pairwise comparison was examined for the 42 hub genes identified (Figure 12 & Table S3). The most differentially expressed gene encodes a flavonoid 3'-monooxygenase (Sobic.004G200833) and was upregulated 10,522-fold in RTx3362 pericarp treated with VIS + UVA + UVB at T17 compared to BTx378 pericarp treated with VIS + UVA + UVB at T17. Motif-based DNA sequence analysis conducted in the promoter region of RTx3362 WGS data (~3-kb upstream from annotated 5'UTR) of this gene revealed a significant query motif (E-value < 0.05) that matched the regulatory element Helix-Turn-Helix discovered utilizing MEME.

The hub genes were also compared to genes residing in QTL (Table 13; Chapter III) associated with 3-DOA accumulation in pericarp tissue conducted between the black and red parent and revealed that 6 hub genes resided in existing QTL. These genes include three benzyl alcohol *O*-benzoyltransferases (Sobic.010G259300, Sobic.010G259500, Sobic.010G259600), homocysteine S-methyltransferase (Sobic.001G453100), monooxygenase 1 (Sobic.004G181000), and flavonoid 3'-monooxygenase (Sobic.004G200833).

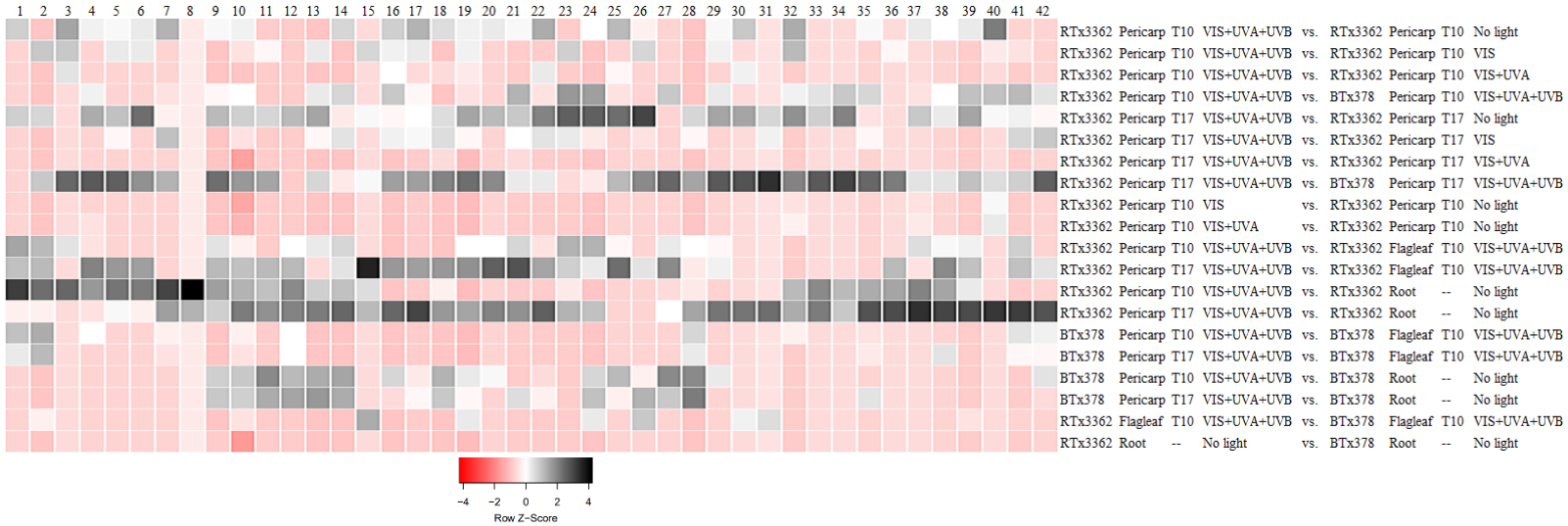


Figure 12. Heatmap showing DEGs of network genes between pairwise comparisons. Pairwise comparisons lacking any DEGs for all 42 genes were not included in the heatmap. 1 ABC transporter G family member 16; 2 Monooxygenase; 3 Benzyl alcohol *O*-benzoyltransferase; 4 Glucan endo-1,3-beta-D-glucosidase; 5 Uncharacterized; 6 Hydroxycinnamoyltransferase 4; 7 Chitinase-like; 8 Uncharacterized; 9 Arabinogalactan Peptide 22-like; 10 Benzyl alcohol *O*-benzoyltransferase; 11 WRKY34; 12 Pathogenesis-related protein 10c; 13 Uncharacterized; 14 MtN26 family protein precursor; 15 Benzyl alcohol *O*-benzoyltransferase; 16 Pathogenesis-related protein 1; 17 Permatin precursor; 18 Ethylene-responsive transcription factor 5; 19 Bowman-Birk serine protease inhibitor; 20 Lysine Histidine Transporter-like 8; 21 Cytochrome P450 84A1; 22 Uncharacterized; 23 MtN26 family protein precursor; 24 Glycerol-3-Phosphate Acyltransferase 3; 25 Hydroxycinnamoyltransferase 4; 26 Uncharacterized; 27 Pathogenesis-related protein 10a; 28 Uncharacterized; 29 Pectinesterase Inhibitor 8; 30 Alpha Carbonic Anhydrase 7; 31 4-coumarate--CoA Ligase 5; 32 Homocysteine S-methyltransferase 1; 33 Acid phosphatase-like; 34 Cytochrome P450 CYP73A100; 35 Bowman-Birk serine protease inhibitor; 36 ABC transporter B family member 11; 37 Cytochrome P450 71A1; 38 Flavonoid 3'-Monooxygenase; 39 Chalcone synthase WHP1; 40 Uncharacterized; 41 Bowman-Birk serine protease inhibitor; 42 Sphinganine C4-Monooxygenase 1. Row Z-score is the measure of distance in standard deviations from the mean.

To visualize the biological interaction of the network generated by WGCNA, highly connected genes were identified based on intramodular connectivity. Intramodular connectivity refers to the degree to which a gene is connected or co-expressed to other genes belonging to a given module. Of the 42 genes in the *darkred* module, the most highly connected genes (strongly co-expressed) included the following: pathogenesis-related proteins (Sobic.001G400700, Sobic.001G400800, Sobic.001G400900), benzyl alcohol *O*-benzoyltransferase (Sobic.010G259500, Sobic.010G066700, Sobic.010G259300), homocysteine S-methyltransferase 1 (Sobic.001G453100), hydroxycinnamolytransferase 4 (Sobic.010G066601, Sobic.010G066700), MtN26 family protein precursor (Sobic.005G166200), permatin precursor (Sobic.003G331700), ABC transporter G family member 16 (Sobic.009G022500), alpha carbonic anhydrase 7 (Sobic.007G155000), two cytochrome P450s (Sobic.001G196300, Sobic.004G141200), chitinase-like protein (Sobic.003G244600), arabinogalactan peptide 22-like protein (Sobic.003G203200) and three uncharacterized proteins (Sobic.003G204900, Sobic.003G278000, Sobic.003G278000) (Figure 13). In the present WCGNA, these highly connected genes are linked with the regulation of 3-DOA accumulation in pericarp of the black sorghum genotype in response to UVB light.

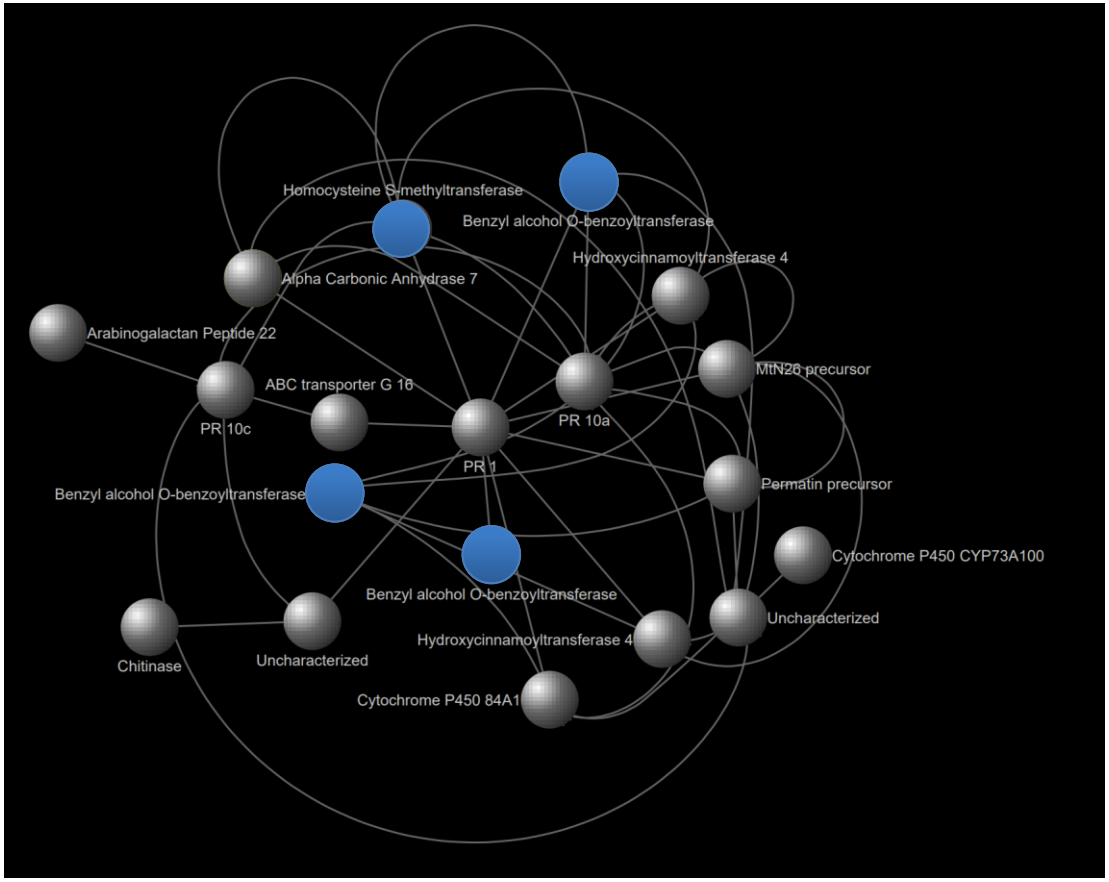
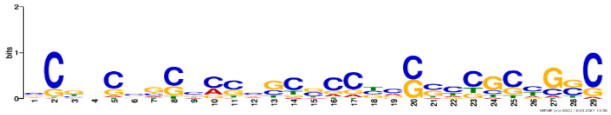
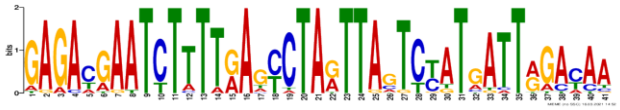
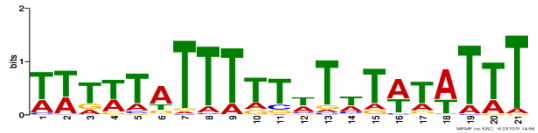
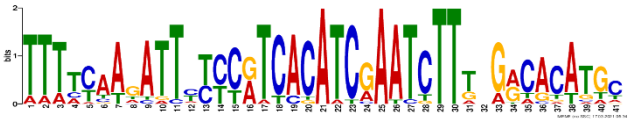
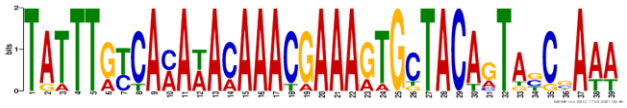


Figure 13. Weighted Gene Co-expression Network Analysis of genes (nodes, gray circles) that are most correlated with the accumulation of 3-DOAs in the black pericarp and have the highest co-expression values. Edges (gray lines) connecting one node to another indicates that those genes are significantly co-expressed. Blue nodes indicate the gene resides in a previously detected QTL interval associated with 3-DOA accumulation in black pericarp (Chapter III).

Conserved DNA Sequence Motif Distribution Analysis of Hub Genes

To investigate potential regulatory mechanisms of the 42 hub genes, a motif-based DNA sequence analysis was conducted in the promoter region of the genes (~3-kb upstream from annotated 5'UTR). Six significant query motifs (E-value < 0.05) were discovered utilizing MEME (Table 17). The query motifs most significantly matched the regulatory elements ethylene-responsive transcription factor 104 (ERF104), two *Arabidopsis* response regulators (ARR2 & ARR1), two COGWHEEL 1 motifs (COG1), and one unannotated motif (Table 17). Of the 42 hub genes, 18 genes contained an ERF104 regulatory site, 18 genes had at least one ARR regulatory site, and 33 genes had at least one COG1 regulatory site positioned upstream of the 5'UTR (Table 17). These results indicated that ERF104, ARR, and/or COG1 may bind to promoters of the hub genes to regulate 3-DOA accumulation black sorghum pericarp tissue.

Table 17. The most highly conserved DNA sequence motifs identified from the promoter region of the 42 co-expressed hub genes.

Logo	E-value	Width	Motif Annotation	Genes containing Motif
	1.3e-120	29	ERF104	Sobic.003G278000, Sobic.009G044100, Sobic.005G166200, Sobic.010G066601, Sobic.003G014400, Sobic.009G022500, Sobic.006G054900, Sobic.004G181000, Sobic.001G453100, Sobic.009G234100, Sobic.005G215900, Sobic.010G066700, Sobic.010G259500, Sobic.009G064400, Sobic.003G203200, Sobic.004G141200, Sobic.001G400900, Sobic.004G200833
	3.5e-146	41	ARR2	Sobic.009G064400, Sobic.005G137200, Sobic.009G234100, Sobic.006G178500, Sobic.003G244600, Sobic.005G214400, Sobic.001G496500, Sobic.005G215900, Sobic.001G322100, Sobic.009G044100, Sobic.010G259600, Sobic.007G145600, Sobic.004G200833, Sobic.001G453100
	2.7e-080	21	COG1	Sobic.005G214400, Sobic.010G259600, Sobic.009G064400, Sobic.006G054900, Sobic.009G234100, Sobic.001G496500, Sobic.006G178500, Sobic.003G267300, Sobic.003G244600, Sobic.009G044100, Sobic.003G355800, Sobic.003G422200, Sobic.001G400900, Sobic.010G066601, Sobic.010G259500, Sobic.001G196300, Sobic.004G296900, Sobic.003G278000, Sobic.007G145600, Sobic.001G322100, Sobic.003G014400, Sobic.001G453100, Sobic.005G137200, Sobic.007G155000, Sobic.005G215900, Sobic.003G204900, Sobic.004G181000, Sobic.010G259300, Sobic.010G092200, Sobic.003G331700, Sobic.001G400700, Sobic.003G286600, Sobic.004G141200, Sobic.005G166200, Sobic.003G203200, Sobic.010G066700, Sobic.004G200833, Sobic.003G288300, Sobic.001G400800, Sobic.009G022500
	4.1e-073	41	ARR1	Sobic.003G244600, Sobic.009G064400, Sobic.006G178500, Sobic.009G234100, Sobic.005G137200, Sobic.009G044100, Sobic.001G496500, Sobic.001G453100, Sobic.010G259600, Sobic.005G214400, Sobic.001G322100, Sobic.007G145600, Sobic.003G286600, Sobic.007G155000
	7.1e-041	21	No Annotation	Sobic.001G496500, Sobic.009G234100, Sobic.009G044100, Sobic.003G244600, Sobic.006G178500, Sobic.005G137200, Sobic.009G064400, Sobic.010G066601, Sobic.001G322100, Sobic.010G259500, Sobic.003G267300, Sobic.005G214400, Sobic.003G204900, Sobic.003G286600, Sobic.005G215900, Sobic.001G453100
	4.4e-036	39	COG1	Sobic.009G064400, Sobic.006G178500, Sobic.009G234100, Sobic.005G214400, Sobic.003G244600, Sobic.001G496500, Sobic.010G259600

Discussion

Secondary metabolites are commonly found in flowers, fruit, and vegetative tissue of most plants and play a role in the protection against an array of biotic and abiotic stresses (Winkel-Shirley 2002). Anthocyanidins are an important group of secondary metabolites containing 3-hydroxyanthocyanidins, and *O*-methylated anthocyanidins, and 3-DOAs (Khoo et al. 2017). The black pericarp trait in sorghum is associated with the rare synthesis of 3-DOAs in which there is notable interest from consumers and industry in the health properties of these secondary metabolites (Xiong et al. 2019). Numerous studies have previously examined gene expression in specific tissues to understand the cellular events and gene regulation leading to the biosynthesis of various flavonoids (Kodama et al. 2018; Zhang et al. 2019; Zhu et al. 2021), however, this is the first comprehensive organ-specific co-expression study conducted on sorghum's black pericarp and the associated biosynthesis of 3-DOAs in pericarp exposed to UVB.

While RNA-seq is useful to investigate gene-to-gene interactions, genes often function via complex interactions and it is difficult to elucidate those connections when analyzing numerous pairwise comparisons that contain a plethora of DEGs. To mine RNA-seq datasets for those interactions, co-expression networks can be utilized to ultimately deduce biological meaning from such large datasets (Grimes et al. 2019). Network analysis is particularly useful to identify key genes and regulators of traits and this approach has been previously applied to numerous crops to successfully unveil regulatory networks related to secondary metabolite synthesis (Zhang et al. 2019; Li et

al. 2020; Mandal et al. 2020). In the present study, I conducted a detailed examination of gene expression in pericarp from black seeded sorghum and contrasted that to gene expression in vegetative tissues and to pericarp from a red seeded genotype. This analysis increased the understanding of genes and gene networks controlling the accumulation of 3-DOAs that occurs in a genotypic- and tissue-specific manner in response to UVB light in sorghum.

The present study revealed that a large portion of pericarp-specific genes expressed in RTx3362 exposed to UVB were significantly associated with response to abiotic stimuli (GO:0009628) and response to endogenous stimuli (GO:0009719). The genes within these enriched GO terms indicated a change in cellular activity induced from both external and internal stimuli, however, the categories and gene annotations alone were not specific enough to infer biological meaning. To search and refine the biological significance underlying the DEGs obtained from extensive transcriptome comparisons, a network analysis was used to investigate co-expression patterns in RTx3362 pericarp tissue developing in UVB light.

A regulatory network of genes influencing pericarp-specific biosynthesis of 3-DOAs in response to UVB was discovered via the application of WGCNA to the numerous (64) transcriptome datasets in this research. Several genes discovered in the network analysis were related to inducible defense against microbial invasion including glucan endo-1,3-beta-D-glucosidase, chitinase, protease inhibitors (bowman-birk serine), benzyl alcohol *O*-benzoyltransferases, permatin, ABC transporter G family member 16, and pathogenesis related proteins (Nejat et al. 2015; Huffaker et al. 2006; Yamaguchi

and Huffaker 2011; Ryan and Pearce 2010). Moreover, certain hub genes were associated with cell-wall composition (arabinogalactan peptide 22) and degradation (glucan endo-1,3-beta-D-glucosidase). Arabinogalactan peptides are glycoproteins found in the plant cell-wall, on the plasma membrane, in the apoplastic space, and in secretions (stigma surface and wound exudates) and are thought to act as regulators of the cell surface and host-pathogen signaling (Ellis et al. 2010). In barley grain and maize seedlings, glucan endo-1,3-beta-D-glucosidase protects against microbial invasion through its ability to degrade fungal cell wall polysaccharides (Kim et al. 2000). Finally, ethylene-responsive transcription factor 5 (ERF5) positively regulates ethylene production and interacts with WRKY34 (homolog of WRKY33 (Guan et al. 2014)) to control defense pathways leading to phytoalexin synthesis in response to different pathogens (Son et al. 2012). The suite of these co-expressed genes suggests that developing black sorghum pericarp responds to UVB penetrance in a manner that is analogous to pathogen attack of pericarp cells, but specifically in black grain sorghum.

While phytoalexins are classically known to accumulate in response to both biotic and abiotic stressors, the mechanisms of induction should, in principle, slightly deviate as biotic stress directly inflicts mechanical damage to the cell while abiotic stress (i.e. UVB irradiance) does not. Local accumulation of 3-DOAs in sorghum vegetative tissue in response to pathogen invasion is a dominant and common trait that has been well characterized in the literature. During initial invasion of plant tissues, pathogens secrete elicitor molecules such as cell-wall degrading enzymes that interact with pattern recognition receptor molecules on the host cell surface (Yamaguchi and Huffaker 2011).

Microbial elicitors induce various host responses ranging from phytoalexin accumulation or induction of phytoalexin biosynthetic enzymes, accumulation of lignin, cellular necrosis, production of ethylene, electrolyte leakage, accumulation of hydroxyproline-rich glycoproteins, or accumulation of proteinase inhibitors. In contrast to biotic stress, low fluence UVB perception and signaling pathways are traditionally mediated by genes encoding UV resistance locus 8 (UVR8), constitutive photomorphogenesis 1 (COP1) and elongated hypocotyl 5 (HY5) (Jenkins 2009), none of which were differentially expressed in the samples analyzed in this study. While other nonspecific signaling pathways involving ROS, DNA damage, and wound/defense signaling molecules can also mediate an organism's response to high-fluence UVB irradiance (Jenkins 2009), the current network analysis indicates an additional layer of complexity governing the black sorghum pericarp response to high-fluence UVB irradiance.

Distinct from elicitors derived from invading pathogens, several plant species can produce and recognize endogenous host-derived elicitor molecules (Yamaguchi and Huffaker 2011). Endogenous elicitors, sometimes referred to as DAMPs, include ROS, oligosaccharide fragments, and protein fragments. The elicitors are generated in response to stress and induce the same downstream signaling pathways as those observed with a typical pathogen invasion (MAMPs) and are advantageous because they facilitate a faster, more amplified defense response (Yamaguchi and Huffaker 2011). Dixon (1986) proposed a model of endogenous elicitation in which host cells synthesize and release molecules with elicitation activity that act as signals to induce phytoalexin biosynthesis. In the model by Dixon (1986), pectic fragments from the host cell wall

were shown to function without involvement of a plant-pathogen and accompanied by a rapid induction of ethylene, hydroxyproline-rich glycoproteins, and proteinase inhibitor synthesis in certain species. Based on the collection of co-expressed genes correlated with the early induction of 3-DOA accumulation and previously characterized variation in WGS data between the black- and red-seeded parents, I propose a pathway modeling 3-DOA induction in black sorghum pericarp tissue in which UVB triggers the generation of endogenous elicitors that initiate a signal cascade resulting in the high expression of flavonoid biosynthetic genes ultimately responsible for the synthesis of 3-DOAs (Figure 14).

Induction of ROS via exposure to UVB light is well characterized in the literature in addition to several enzymes known to be critical for detoxification and protection of these species (Jenkins 2009). Some enzymatic processes are more efficient at detoxifying ROS than others such as the protein family flavodoxin-like protein WrbA which functions by responding to altered redox conditions in the cell (Carey et al. 2007). flavodoxin-like protein WrbA prevents redox cycling via its two-electron NAD(P)H:quinone oxidoreductases activity (Carey et al. 2007). In the present study, exposure of developing red-seeded sorghum to UVB light led to the high expression of the flavodoxin-related protein WrbA (Sobic.004G185100) in BTx378 at T17. This gene was not expressed in any RNA-seq library prepared from the black sorghum genotype, and examination of genomic DNA resequencing data revealed that the entirety of the coding region for this gene is not present in the black sorghum genome (Figure 10). Additionally, the region on chromosome 4 where this gene would reside coincides with a

QTL associated with 3-DOA accumulation in black pericarp tissue (Chapter III). The expression of this gene in red pericarp genotypes should enable efficient ROS detoxification, while ROS detoxification via this robust enzymatic mechanism would be lacking in black pericarp tissues. Thus, accumulation of ROS in black sorghum pericarp may act as an early endogenous elicitor that initiates cell wall disassembly resulting in pectin fragments generated via hydrolysis indicated by the hub genes Sobic.001G400900, Sobic.003G355800, and Sobic.005G215900 (Figure 12 & Table S3). Pectin fragments also possess endogenous elicitor activity; however, the fragments will lose their elicitor activity if esterified (Dixon 1986; Hahn et al. 1981). Consistent with this working model, black sorghum expresses a network hub gene that inhibits esterification of pectin fragments (Figure 12 & Table S3). Cell wall fragments also induce protease inhibitors (Figure 12) which further contribute to host defense response (Pearce et al. 1991). Notably, expression of the protease inhibitor Sobic.005215900 appears to be organ-specific and only upregulated in UVB-irradiated black sorghum pericarp in comparison to UVB-irradiated leaf tissue (Figure 12 & Table S3). Downstream signaling occurs through a cascade involving WRKY34 (Figure 12 & Table S3), mitogen-activated protein kinases (MPK3/MPK6) (not found in the network), and pathogenesis related proteins (Figure 12 & Table S3) leading to the activation of phytoalexin biosynthetic genes, particularly those identified in the network analysis including 4-coumarate--CoA ligase 5 (Sobic.007G145600), chalcone synthase WHP1 (Sobic.005G137200), and flavonoid 3'-monooxygenase (Sobic.004G200833) (Figure 12 & Table S3) (Eckardt 2011).

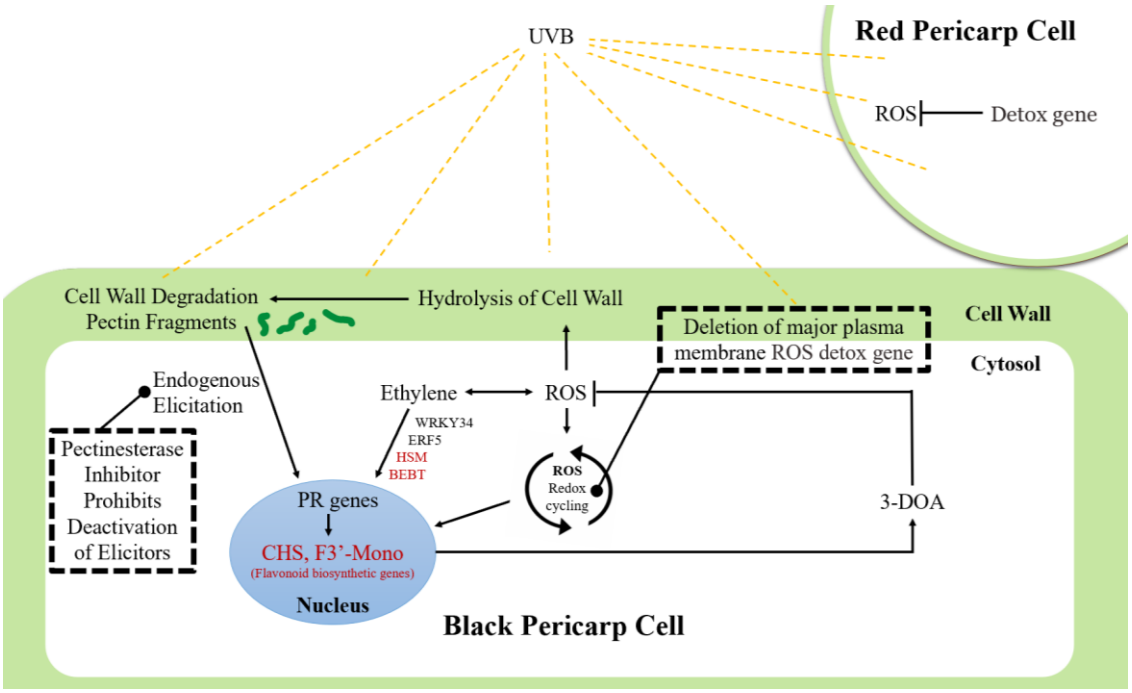


Figure 14. Model of UVB induced ROS, redox cycling, endogenous elicitor generation and signal transduction leading to 3-DOA accumulation in black pericarp tissue. Text in red indicates the gene resides in a QTL interval (Chapter III).

ERF5, a hub gene discovered in the network analysis (Figure 12), is phylogenetically closely related to ERF104, a significant highly conserved motif discovered amongst the hub genes (Table 17) (Illgen et al. 2020). ERF genes are known to be differentially regulated by ethylene, but also by additional extracellular signals from abiotic stress conditions and are essential for the expression of pathogenesis-related proteins (Fujimoto et al. 2000). Production of ethylene involves the enzymatic activity of homocysteine S-methyltransferase (Somssich and Hahlbrock 1998), a hub gene also observed in the co-expression network and found within a QTL interval (Figure 12; Chapter III). The coupled role of ethylene in phytoalexin synthesis is consistent with the model proposed by Dixon (1986).

Two additional motifs characterized as ARR regulatory sites were also highly conserved amongst the network genes (Table 17). ARR regulatory sites belong to the MYB/SANT family and are known to be involved in ethylene signaling (Hass et al. 2004). A previous study proposed that *Yellow seed 1* (*Y1*), an R2R3 MYB transcription factor, is responsible for 3-DOA production in sorghum pericarp and vegetative tissue (Boddu et al 2005), however, a recent investigation into this locus revealed large structural variants in the black sorghum genome that preclude a clear understanding of the functional significance of the *Y1* transcription factor in black pericarp at this time (Chapter III). A potential link exists between the MYB-like (ARR) binding motifs discovered in the network genes and the previous indications that 3-DOA accumulation is controlled by the *Y1* R2R3 MYB transcription factor. MYB transcription factors are a large gene family in plants that control a wide range of metabolic processes (Ibraheem et al. 2015; Jin and Martin 1999; Jin et al. 2000). Thus, MYB transcription factors beyond that of the *Y1* gene are potentially involved in the regulation of 3-DOA biosynthesis in pericarp tissues in response to UVB light. Further efforts are warranted to properly annotate the region spanning the *Y1* locus in the black sorghum genome as well as the role of additional MYB transcription factors and ethylene signaling as it pertains to 3-DOA accumulation in sorghum pericarp.

Finally, 33 genes had at least one COG1 regulatory site positioned upstream of the 5'UTR (Table 17). COG1 transcription factors, homologous to DOF transcription factors, are involved in seed coat formation and light perception in *Arabidopsis* seeds

(Bueso et al. 2016). Specifically, COG1 acts as a negative regulator of the phytochrome signaling pathway (Bueso et al. 2016) which distinguishes these transcription factors from those typically involved in the UV light signaling pathway. At present, the involvement of COG1 transcriptional regulation in UVB-dependent accumulation of 3-DOAs in black sorghum pericarp cannot be further described.

In summary, while the underlying genes controlling the inheritance of the black pericarp trait are complex, the present results implicate that pericarp-specific 3-DOA accumulation occurs in response to endogenous elicitation stimulated by higher fluences of UVB light. Higher UVB fluences activate other stress-responsive pathways causing an increase in cellular ROS. Specific to the black sorghum genotype, there appears to be a gene deletion that leads to insufficient cellular detoxification which stimulates the production of endogenous elicitors that induce the biosynthesis of 3-DOAs which have an enhanced ability to detoxify the cell. The model proposed in this research may support future efforts in both biological and synthetic elicitation of higher levels of 3-DOAs for application in the food and health industries.

CHAPTER V

CONCLUSIONS

This dissertation research aimed to address the environmental and genetic regulation in sorghum of the rare black pericarp trait and the associated accumulation of 3-DOAs. The interrelated research projects conducted in this work identified the critical light spectra required for full penetrance of the black sorghum grain traits, identified major-effect QTL for 3-DOAs accumulation, characterized organ-specific gene expression related to 3-DOA production in sorghum, and proposed a model built upon the present results that visually depicts the cascade of events that lead to the unique accumulation of 3-DOA in the black sorghum genotype.

In Chapter II, the black sorghum genotype was grown under regimes of visible light, visible light supplemented with UVA or supplemented with UVA plus UVB (or dark control). Pericarp 3-DOAs and pigmentation were maximized in the black genotype exposed to a light regime supplemented with UVB revealing that a fluence of UVB is the major determining environmental factor for trait penetrance. Through an exploratory transcriptome analysis, changes in gene expression during black pericarp development revealed that UVB light activates genes related to plant defense, reactive oxygen species, and secondary metabolism which suggested that 3-DOA accumulation is associated with activation of flavonoid biosynthesis and several overlapping defense and stress signaling pathways.

In Chapter III, multiple trait mapping methodologies were used to characterize the genomic loci associated with the black sorghum phenotype including genetic linkage

map construction with QTL analysis using an F_5 RIL population segregating for pericarp 3-DOA accumulation and BSA of opposing phenotypic extremes for 3-DOA accumulation. These mapping efforts revealed four major QTL detected on chromosomes 1, 2, 4, and 10 and several minor effect QTL associated with 3-DOA accumulation. Within the major QTL on chromosome 4, a deletion of a critical detoxification gene was discovered in the RTx3362 genome. Transcripts for this gene were found to be highly upregulated in BTx378 pericarp tissue upon exposure to UVB light. Moreover, a preliminary examination of read alignment between BTx378 and RTx3362 WGS data detected structural variants in the genomic region spanning the *Y1* gene previously characterized to regulate production of 3-DOAs in sorghum vegetative tissue. While structural variants spanning the *Y1* gene may play a role in conditioning the black pericarp phenotype, further molecular-based investigation into the genomic region flanking the *Y1* gene is necessary to ultimately determine the role *Y1* serves in 3-DOA accumulation in sorghum pericarp tissue. The additional genes underlying significant QTL discovered in this research, particularly the seven genes that were also identified as network hub genes, require further investigation to elucidate their involvement in black pericarp accumulation of 3-DOAs. Finally, a preliminary investigation on the potential reversibility of the genetic control of this trait was conducted, and it was determined that multiple generations of grain development in the absence of light could not reverse the UVB-induced black pericarp phenotype. Together, these results provided the necessary framework to uncover a potential gene

regulatory network leading to the tissue-specific accumulation of these phytochemicals in black sorghum grain described in Chapter IV.

In Chapter IV of this dissertation, a detailed transcriptome and co-expression analysis was performed on pericarp, leaf, and root tissue from black- and red-grain sorghum genotypes under different light regimes which revealed a potential pathway modeling 3-DOA induction in black sorghum pericarp tissue. In this model, UVB triggers the generation of endogenous elicitors that initiate a signal cascade resulting in the high expression of flavonoid biosynthetic genes ultimately responsible for 3-DOA synthesis. The deletion of the detoxification gene in RTx3362 discovered in Chapter III plays a critical role in this model in that the expression and action of this gene product in pericarp tissue from BTx378 ultimately protects and prevents the ROS-induced signaling cascade from occurring. In theory, the red-sorghum pericarp would not require the metabolically expensive production of 3-DOAs for ROS detoxification and cellular protection as is necessary in RTx3362 pericarp in response to UVB exposure.

Collectively, the work achieved in this research has addressed important questions surrounding the unique environmental and genetic control of black pericarp traits in sorghum. This research has also generated future objectives to clarify the genomic regions surrounding the Y1 MYB transcription factor, to validate variants observed in genes underlying QTL regions, and to confirm the model proposed to be controlling the genotype-specific accumulation of 3-DOAs in black sorghum grain.

REFERENCES

Abdel-Aal ES, Young JC, Rabalski I (2006) Anthocyanin composition in black, blue, pink, purple, and red cereal grains. Journal of Agricultural and Food Chemistry 54:4696-704. https://doi.org/10.1021/jf0606609

Awika JM, Rooney LW, Waniska RD (2004) Properties of 3-deoxyanthocyanins from sorghum. Journal of Agricultural and Food Chemistry 52:4388−4394. https://doi.org/10.1021/jf049653f

Awika JM. Behavior of 3-deoxyanthocyanidins in the presence of phenolic copigments. Food Research International 41:532-8. https://doi.org/10.1016/j.foodres.2008.03.002

Aharoni A, De Vos CR, Wein M, Sun Z, Greco R, Kroon A, Mol JN, O'Connell AP (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. The Plant Journal 28:319-32. https://doi.org/10.1046/j.1365-313X.2001.01154.x

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE (2014) A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. The Plant Cell 26:962-80. https://doi.org/10.1105/tpc.113.122069

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. Physiology and Molecular Biology of Plants 19:307-21. https://doi.org/10.1007/s12298-013-0179-1

Andrews S (2010) A quality control tool for high throughput sequence data. Babraham Bioinformatics: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ayyangar GR, Vijiaraghavan C, Ayyar MS, Rao VP (1933a) Inheritance of characters in sorghum. III. Grain colours red, yellow, white. Indian Journal of Agricultural Science 3:594-604.

Ayyangar GR, Vijiaraghavan C, Pillai VG, Ayyar MA (1933b) Inheritance of characters in sorghum–the great millet. II. Purple pigmentation on leaf sheath and glume. Indian Journal of Agricultural Sciences 3:589-604.

Ayyangar GR, Rao VP, Ponnaiya BW (1938) The occurrence and inheritance of purple-tipped grains in sorghum. Proceedings of the Indian Academy of Sciences-Section B 8:396-398.

Azuma A, Yakushiji H, Koshita Y, Kobayashi S (2012) Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. Planta 236:1067-80. https://doi.org/10.1007/s00425-012-1650-x

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37(suppl_2):W202-8. https://doi.org/10.1093/nar/gkp335

Barua UM, Chalmers KJ, Hackett CA, Thomas WT, Powell W, Waugh R (1993) Identification of RAPD markers linked to a *Rhynchosporium secalis* resistance locus in barley using near-isogenic lines and bulked segregant analysis. Heredity 71:177–184. https://doi.org/10.1007/BF01184917

Baskar V, Venkatesh R, Ramalingam S (2018) Flavonoids (antioxidants systems) in higher plants and their response to stresses. Antioxidants and Antioxidant Enzymes in Higher Plants. Springer, pp 253-268.

Bentley N, Grauke LJ, Klein PE (2019) Genotyping by sequencing (GBS) and SNP marker analysis of diverse accessions of pecan (*Carya illinoinensis*). Tree Genetics & Genomes 15.1:8. https://doi.org/10.1007/s11295-018-1314-5

Blakely ME, Rooney LW, Sullins RD, Miller FR (1979) Microscopy of the pericarp and the testa of different genotypes of sorghum. Crop Science 19:837-42. https://doi.org/10.2135/cropsci1979.0011183X001900060023x

Boddu J, Jiang C, Sangar V, Olson T, Peterson T, Chopra S (2006) Comparative structural and functional characterization of sorghum and maize duplications containing orthologous myb transcription regulators of 3-deoxyflavonoid biosynthesis. Plant Molecular Biology 60:185-99. https://doi.org/10.1007/s11103-005-3568-1

Boddu J, Svabek C, Ibraheem F, Jones AD, Chopra S (2005) Characterization of a deletion allele of a sorghum myb gene yellow seed1 showing loss of 3-deoxyflavonoids. Plant Science 169.3:542-52. https://doi.org/10.1016/j.plantsci.2005.05.007

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114-20. https://doi.org/10.1093/bioinformatics/btu170

Bratt K, Sunnerheim K, Bryngelsson S, Fagerlund A, Engman L, Andersson RE, Dimberg LH (2003) Avenanthramides in oats (*Avena sativa* L.) and structure−antioxidant activity relationships. Journal of Agricultural and Food Chemistry 51:594-600. https://doi.org/10.1021/jf020544f

Braun J, Tevini M (1993) Regulation of uv-protective pigment synthesis in the epidermal layer of rye seedlings. Photochemistry and Photobiology 57:318-323. https://doi.org/10.1111/j.1751-1097.1993.tb02294.x

Branham SE, Farnham MW (2019) Identification of heat tolerance loci in broccoli through bulked segregant analysis using whole genome resequencing. Euphytica 215:34. https://doi.org/10.1007/s10681-018-2334-9

Branham SE, Wechter WP, Lambel S, Massey L, Ma M, Fauve J, Farnham MW, Levi A (2018) QTL-seq and marker development for resistance to *Fusarium oxysporum* f. sp. *niveum* race 1 in cultivated watermelon. Molecular Breeding 38.11:1-9. https://doi.org/10.1007/s11032-018-0896-9

Brederode FT, Linthorst HJ, Bol JF (1991) Differential induction of acquired resistance and pr gene expression in tobacco by virus infection, ethephon treatment, uv light and wounding. Plant Molecular Biology 17:1117-25.

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19.7:889-90. https://doi.org/10.1093/bioinformatics/btg112

Brosché M, Strid Å (2003) Molecular events following perception of ultraviolet-b radiation by plants. Physiologia Plantarum 117:1-0. https://doi.org/10.1034/j.1399-3054.2003.1170101.x

Bueso E, Muñoz-Bertomeu J, Campos F, Martínez C, Tello C, Martínez-Almonacid I, Ballester P, Simón-Moya M, Brunaud V, Yenush L, Ferrándiz C (2016) *Arabidopsis* cogwheel 1 links light perception and gibberellins with seed tolerance to deterioration. The Plant Journal 87:583-96. https://doi.org/10.1111/tpj.13220

Burchard P, Bilger W, Weissenböck G (2000) Contribution of hydroxycinnamates and flavonoids to epidermal shielding of uv-a and uv-b radiation in developing rye primary leaves as assessed by ultraviolet-induced chlorophyll fluorescence measurements. Plant, Cell & Environment 23:1373-80. https://doi.org/10.1046/j.1365-3040.2000.00633.x

Carey J, Brynda J, Wolfová J, Grandori R, Gustavsson T, Ettrich R, Smatanová IK (2007) WrbA bridges bacterial flavodoxins and eukaryotic nad(p)h:quinone oxidoreductases. Protein Science 16:2301-5. https://doi.org/10.1110/ps.073018907

Casañal A, Zander U, Muñoz C, Dupeux F, Luque I, Botella MA, Schwab W, Valpuesta V, Marquez JA (2013) The strawberry pathogenesis-related 10 (pr-10) fra a proteins control flavonoid biosynthesis by binding to metabolic intermediates. Journal of Biological Chemistry 288:35322-32. https://doi.org/10.1074/jbc.M113.501528

Casati P, Walbot V (2004) Rapid transcriptome responses of maize (*Zea mays*) to uv-b in irradiated and shielded tissues. Genome Biology 5.3:1-9. https://doi.org/10.1186/gb-2004-5-3-r16

Casati P, Walbot V (2005) Differential accumulation of maysin and rhamnosylisoorientin in leaves of high-altitude landraces of maize after uv-b exposure. Plant, Cell & Environment 28:788-99. https://doi.org/10.1111/j.1365-3040.2005.01329.x

Chopra S, Gevens A, Svabek C, Wood KV, Peterson T, Nicholson R (2002) Excision of the candystripe1 transposon from a hyper-mutable y1-cs allele shows that the sorghum y1 gene controls the biosynthesis of both 3-deoxyanthocyanidin phytoalexins and phlobaphene pigments. Physiological and Molecular Plant Pathology 60:321-30. https://doi.org/10.1006/pmpp.2002.0411

Cone KC (2007) Anthocyanin synthesis in maize aleurone tissue. Endosperm: developmental and molecular biology. Springer Berlin Heidelberg, pp 121-139

Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN, Walker AR, Robinson SP, Bogs J (2009) The grapevine r2r3-myb transcription factor vvmybf1 regulates flavonol synthesis in developing grape berries. Plant Physiology 151:1513-30. https://doi.org/10.1104/pp.109.142059

Davis CB, Markey CE, Busch MA, Busch KW (2007) Determination of capsaicinoids in habanero peppers by chemometric analysis of uv spectral data. Journal of Agricultural and Food Chemistry 55:5925-33. https://doi.org/10.1021/jf070413k

Dixon RA (1986) The phytoalexin response: elicitation, signalling and control of host gene expression. Biological Reviews 61:239-91. https://doi.org/10.1111/j.1469-185X.1986.tb00719.x

Doyle J (1991) DNA protocols for plants. Molecular Techniques in Taxonomy. Springer Berlin Heidelberg, pp 283-293. https://doi.org/10.1007/978-3-642-83962-7_18

Dressel A, Hemleben V (2009) Transparent testa glabra 1 (ttg1) and ttg1-like genes in *Matthiola incana* r. br. and related brassicaceae and mutation in the wd-40 motif. Plant Biology 11:204-212. https://doi.org/10.1111/j.1438-8677.2008.00099.x

Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E, Debeaujon I, Routaboul JM, Alboresi A, Weisshaar B, Lepiniec L (2008) Mybl2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. The Plant Journal 55:940-53. https://doi.org/10.1111/j.1365-313X.2008.03564.x

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) Myb transcription factors in *Arabidopsis*. Trends in Plant Science 15:573-81. https://doi.org/10.1016/j.tplants.2010.06.005

Dykes L, Rooney LW, Waniska RD, Rooney WL (2005) Phenolic compounds and antioxidant activity of sorghum grains of varying genotypes. Journal of Agricultural and Food Chemistry 53.17:6813-6818. https://doi.org/10.1021/jf050419e

Dykes L, Rooney LW (2006) Sorghum and millet phenols and antioxidants. Journal of Cereal Science 44.3:236-251. https://doi.org/10.1016/j.jcs.2006.06.007

Dykes L, Seitz LM, Rooney WL, Rooney LW (2009) Flavonoid composition of red sorghum genotypes. Food Chemistry 116:313− 317. https://doi.org/10.1016/j.foodchem.2009.02.052

Dykes L (2008) Flavonoid composition and antioxidant activity of pigmented sorghums of varying genotypes. Dissertation, Texas A&M University

Dykes L, Peterson GC, Rooney WL, Rooney LW (2011) Flavonoid composition of lemon-yellow sorghum genotypes. Food Chemistry 128.1:173-179. https://doi.org/10.1016/j.foodchem.2011.03.020

Dykes L, Rooney WL, Rooney LW (2013) Evaluation of phenolics and antioxidant activity of black sorghum hybrids. Journal of Cereal Science 58:278−283. https://doi.org/10.1016/j.jcs.2013.06.006

Eckardt NA (2011) Induction of phytoalexin biosynthesis: wrky33 is a target of mapk signaling. The Plant Cell 23:1190. https://doi.org/10.1105/tpc.111.230413

Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-proteins: key regulators at the cell surface. Plant Physiology 153:403-19. https://doi.org/10.1104/pp.110.156000

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007) Red colouration in apple fruit is due to the activity of the myb transcription factor, mdmyb10. The Plant Journal 49:414-27. https://doi.org/10.1111/j.1365-313X.2006.02964.x

FAO (2019) Food and Agricultural Organization of the United Nations. Rome, Italy. FAOSTAT Database. http://faostat3.fao.org Accessed 8 Feb. 2021

Fedenia L, Klein RR, Dykes L, Rooney WL, Klein PE (2020) Phenotypic, phytochemical, and transcriptomic analysis of black sorghum (*Sorghum bicolor* L.)

pericarp in response to light quality. Journal of Agricultural and Food Chemistry 68.37:9917-9929. https://doi.org/10.1021/acs.jafc.0c02657

Ficco DB, De Simone V, De Leonardis AM, Giovanniello V, Del Nobile MA, Padalino L, Lecce L, Borrelli GM, De Vita P (2016) Use of purple durum wheat to produce naturally functional fresh and dry pasta. Food Chemistry 205:187-95. https://doi.org/10.1016/j.foodchem.2016.03.014

Fini A, Brunetti C, Di Ferdinando M, Ferrini F, Tattini M (2011) Stress-induced flavonoid biosynthesis and the antioxidant machinery of plants. Plant Signaling & Behavior 6.5:709-11. https://doi.org/10.4161/psb.6.5.15069

Frohnmeyer H, Staiger D (2003) Ultraviolet-b radiation-mediated responses in plants. Plant Physiology 133:1420-8. https://doi.org/10.1104/pp.103.030049

Fuleki T, Francis FJ (1968) Quantitative methods for anthocyanins. Journal of Food Science 33.1:72-77. https://doi.org/10.1111/j.1365-2621.1968.tb00887.x

Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000) Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of gcc box–mediated gene expression. The Plant Cell 12:393-404. https://doi.org/10.1105/tpc.12.3.393

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosome intervals using dna pools from existing mapping populations. Nucleic Acids Research 19:6553-6558. https://doi.org/10.1093/nar/19.23.6553

Goff SA, Cone KC, Chandler VL (1992) Functional analysis of the transcriptional activator encoded by the maize b gene: evidence for a direct functional interaction between two classes of regulatory proteins. Genes & Development 6:864-75. https://doi.org/10.1101/gad.6.5.864

Görlach A, Bertram K, Hudecova S, Krizanova O (2015) Calcium and ros: a mutual interplay. Redox Biology 6:260-71. https://doi.org/10.1016/j.redox.2015.08.010

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

Grandori R, Carey J (1994) Six new candidate members of the α/β twisted open-sheet family detected by sequence similarity to flavodoxin. Protein Science 12:2185-93. https://doi.org/10.1002/pro.5560031204

Greenberg BM, Wilson MI, Huang XD, Duxbury CL, Gerhardt KE, Gensemer RW (1997) The effects of ultraviolet-b radiation on higher plants. Plants for Environmental Studies 2, pp 1-35

Green R, Fluhr R (1995) Uv-b-induced pr-1 accumulation is mediated by active oxygen species. The Plant Cell 7:203-12. https://doi.org/10.1105/tpc.7.2.203

Grimes T, Potter SS, Datta S (2019) Integrating gene regulatory pathways into differential network analysis of gene expression data. Scientific Reports 9:1-2. https://doi.org/10.1038/s41598-019-41918-3

Guan Y, Meng X, Khanna R, LaMontagne E, Liu Y, Zhang S (2014) Phosphorylation of a wrky transcription factor by mapks is required for pollen development and function in arabidopsis. PLoS Genetics 10:e1004384. https://doi.org/10.1371/journal.pgen.1004384

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. Genome Biology 8:1-9. https://doi.org/10.1186/gb-2007-8-2-r24

Hahn DH, McDonough C, Rosenow D, Miller F, Rooney L, Meckenstoc D (1984) Shawaya: A "black" sorghum from the western sudan. Sorghum Newsletter:51.

Hahn MG, Darvill AG, Albersheim P (1981) Host-pathogen interactions: the endogenous elicitor, a fragment of a plant cell wall polysaccharide that elicits phytoalexin accumulation in soybeans. Plant Physiology 68:1161-9. https://doi.org/10.1104/pp.68.5.1161

Hahn DH, Rooney LW (1986) Effect of genotype on tannins and phenols of sorghum. Cereal Chemistry 63:4-8.

Halbwirth H, Martens S, Wienand U, Forkmann G, Stich K (2003) Biochemical formation of anthocyanins in silk tissue of *Zea mays*. Plant Science 164:489-95. https://doi.org/10.1016/S0168-9452(02)00433-8

Hass C, Lohrmann J, Albrecht V, Sweere U, Hummel F, Yoo SD, Hwang I, Zhu T, Schäfer E, Kudla J, Harter K (2004) The response regulator 2 mediates ethylene signalling and hormone signal integration in *Arabidopsis*. The EMBO Journal 23:3290-302. https://doi.org/10.1038/sj.emboj.7600337

Hayes C (2012) Genetic and environmental effects on phenolic composition and agronomic performacne in black sorghum (*Sorghum bicolor* L.) hybrids. Thesis, Texas A&M University

Heim KE, Tagliaferro AR, Bobilya DJ (2002) Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. The Journal of Nutritional Biochemistry 13:572-84. https://doi.org/10.1016/S0955-2863(02)00208-5

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Molecular Biology and Evolution 20:735-47. https://doi.org/10.1093/molbev/msg088

Heredia JB, Cisneros-Zevallos L (2009) The effect of exogenous ethylene and methyl jasmonate on pal activity, phenolic profiles and antioxidant capacity of carrots (*Daucus carota*) under different wounding intensities. Postharvest Biology and Technology 51:242-9. https://doi.org/10.1016/j.postharvbio.2008.07.001

Hill KM (2014) Inheritance of black pericarp trait in sorghum. Dissertation, Texas A&M University

Hipskind JD, Hanau R, Leite B, Nicholson RL (1990) Phytoalexin accumulation in sorghum: identification of an apigeninidin acyl ester. Physiological and Molecular Plant Pathology 36:381-96. https://doi.org/10.1016/0885-5765(90)90067-8

Hormaza JI, Dollo L, Polito VS (1994) Identification of a rapd marker linked to sex determination in *Pistacia vera* using bulked segregant analysis. Theoretical Applied Genetics 89:9–13 https://doi.org/10.1007/BF00226975

hu Dong Y, Mitra D, Kootstra A, Lister C, Lancaster J (1995) Postharvest stimulation of skin color in royal gala apple. Journal of the American Society for Horticultural Science 120:95-100. https://doi.org/10.21273/JASHS.120.1.95

Huffaker A, Pearce G, Ryan CA (2006) An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. Proceedings of the National Academy of Sciences 103:10098-103. https://doi.org/10.1073/pnas.0603727103

Ibraheem F, Gaffoor I, Chopra S (2010) Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional yellow seed1 in *Sorghum bicolor*. Genetics 184:915-26. https://doi.org/10.1534/genetics.109.111831

Ibraheem F, Gaffoor I, Tan Q, Shyu CR, Chopra S (2015) A sorghum myb transcription factor induces 3-deoxyanthocyanidins and enhances resistance against leaf blights in maize. Molecules 20:2388-404. https://doi.org/10.3390/molecules20022388

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by qtl-seq. Theoretical and Applied Genetics 128:1329–1342. https://doi.org/10.1007/s00122-015-2509-x

Illgen S, Zintl S, Zuther E, Hincha DK, Schmülling T (2020) Characterisation of the erf102 to erf105 genes of *Arabidopsis thaliana* and their role in the response to cold stress. Plant Molecular Biology 18:1-8. https://doi.org/10.1007/s11103-020-00993-1

Jenkins GI (2009) Signal transduction in responses to uv-b radiation. Annual Review of Plant Biology 60:407-31. https://doi.org/10.1146/annurev.arplant.59.032607.092953

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C (2000) Transcriptional repression by atmyb4 controls production of uv-protecting sunscreens in *Arabidopsis*. The EMBO Journal 19:6150-61. https://doi.org/10.1093/emboj/19.22.6150

Jin H, Martin C (1999) Multifunctionality and diversity within the plant myb-gene family. Plant Molecular Biology 41:577-85. https://doi.org/10.1023/A:1006319732410

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2016) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research. https://doi.org/10.1093/nar/gkw982

Kadomura-Ishikawa Y, Miyawaki K, Noji S, Takahashi A (2013) Phototropin 2 is involved in blue light-induced anthocyanin accumulation in *Fragaria* x *ananassa* fruits. Journal of Plant Research 126:847-57. https://doi.org/10.1007/s10265-013-0582-2

Kawahigashi H, Kasuga S, Sawada Y, Yonemaru JI, Ando T, Kanamori H, Wu J, Mizuno H, Momma M, Fujimoto Z, Hirai MY (2016a) The sorghum gene for leaf color changes upon wounding (p) encodes a flavanone 4-reductase in the 3-deoxyanthocyanidin biosynthesis pathway. G3: Genes, Genomes, Genetics 6:1439-47. https://doi.org/10.1534/g3.115.026104

Kawahigashi H, Nonaka E, Mizuno H, Kasuga S, Okuizumi H (2016b) Classification of genotypes of leaf phenotype (p/tan) and seed phenotype (y1 and tan1) in tan sorghum (*Sorghum bicolor*). Plant Breeding 135:683-90. https://doi.org/10.1111/pbr.12426

Khalil A, Baltenweck-Guyot R, Ocampo-Torres R, Albrecht P (2010) A novel symmetrical pyrano-3-deoxyanthocyanidin from sorghum species. Phytochemistry Letters 3:93-5. https://doi.org/10.1016/j.phytol.2010.02.003

Khoo HE, Azlan A, Tang ST, Lim SM (2017) Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food & Nutrition Research 61:1361779. https://doi.org/10.1080/16546628.2017.1361779

Kim JB, Olek AT, Carpita NC (2000) Cell wall and membrane-associated exo-β-d-glucanases from developing maize seedlings. Plant Physiology 123:471-86. https://doi.org/10.1104/pp.123.2.471

Kodama M, Brinch-Pedersen H, Sharma S, Holme IB, Joernsgaard B, Dzhanfezova T, Amby DB, Vieira FG, Liu S, Gilbert MT (2018) Identification of transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using rna-seq. BMC Genomics 19:1-3. https://doi.org/10.1186/s12864-018-5135-6

Koyama K, Ikeda H, Poudel PR, Goto-Yamamoto N (2012) Light quality affects flavonoid biosynthesis in young berries of cabernet sauvignon grape. Phytochemistry 78:54-64. https://doi.org/10.1016/j.phytochem.2012.02.026

Kuhn DN, Chappell J, Boudet A, Hahlbrock K (1984) Induction of phenylalanine ammonia-lyase and 4-coumarate: coa ligase mrnas in cultured plant cells by uv light or fungal elicitor. Proceedings of the National Academy of Sciences 81:1102-6. https://doi.org/10.1073/pnas.81.4.1102

Kumar VG, Viswanathan R, Malathi P, Nandakumar M, Sundar AR (2015) Differential induction of 3-deoxyanthocyanidin phytoalexins in relation to *Colletotrichum falcatum* resistance in sugarcane. Sugar Technology 17:314-21. https://doi.org/10.1007/s12355-014-0334-1

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:1-3. https://doi.org/10.1186/1471-2105-9-559

Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows–wheeler transform. Bioinformatics 26.5:589-95. https://doi.org/10.1093/bioinformatics/btp698

Li H, Han M, Yu L, Wang S, Zhang J, Tian J, Yao Y (2020) Transcriptome analysis identifies two ethylene response factors that regulate proanthocyanidin biosynthesis during *Malus* crabapple fruit development. Frontiers in Plant Science 11:76. https://doi.org/10.3389/fpls.2020.00076

Li S (2014) Transcriptional control of flavonoid biosynthesis: fine-tuning of the myb-bhlh-wd40 (mbw) complex. Plant Signaling & Behavior 9:e27522. https://doi.org/10.4161/psb.27522

Lin-Wang KU, Micheletti D, Palmer J, Volz R, Lozano L, Espley R, Hellens RP, Chagne D, Rowan DD, Troggio M, Iglesias I (2011) High temperature reduces apple

fruit colour via modulation of the anthocyanin regulatory complex. Plant, Cell & Environment 34:1176-90. https://doi.org/10.1111/j.1365-3040.2011.02316.x

Liu H, Du Y, Chu H, Shih CH, Wong YW, Wang M, Chu IK, Tao Y, Lo C (2010) Molecular dissection of the pathogen-inducible 3-deoxyanthocyanidin biosynthesis pathway in sorghum. Plant and Cell Physiology 51:1173-85. https://doi.org/10.1093/pcp/pcq080

Liu Z, Liu Y, Pu Z, Wang J, Zheng Y, Li Y, Wei Y (2013) Regulation, evolution, and functionality of flavonoids in cereal crops. Biotechnology Letters 35:1765-80. https://doi.org/10.1007/s10529-013-1277-4

Lo SC, Nicholson RL (1998) Reduction of light-induced anthocyanin accumulation in inoculated sorghum mesocotyls: implications for a compensatory role in the defense response. Plant Physiology 116:979-89. https://doi.org/10.1104/pp.116.3.979

Lo SC, de Verdier KA, Nicholson RL (1999) Accumulation of 3-deoxyanthocyanidin phytoalexins and resistance to *Colletotrichum sublineolum* in sorghum. Physiological and Molecular Plant Pathology 55:263-73. https://doi.org/10.1006/pmpp.1999.0231

Logemann E, Tavernaro A, Schulz W, Somssich IE, Hahlbrock K (2000) Uv light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. Proceedings of the National Academy of Sciences 97:1903-7. https://doi.org/10.1073/pnas.97.4.1903

Lower WR, Gorsuch JW, Hughes J (1997) Plants for Environmental Studies. CRC Press.

Ludwig SR, Habera LF, Dellaporta SL, Wessler SR (1989) Lc, a member of the maize r gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proceedings of the National Academy of Sciences 86:7092-6. https://doi.org/10.1073/pnas.86.18.7092

Mandal S, Ji W, McKnight TD (2020) Candidate gene networks for acylsugar metabolism and plant defense in wild tomato *Solanum pennellii*. The Plant Cell 32:81-99. https://doi.org/10.1105/tpc.19.00552

Mano H, Ogasawara F, Sato K, Higo H, Minobe Y (2007) Isolation of a regulatory gene of anthocyanin biosynthesis in tuberous roots of purple-fleshed sweet potato. Plant Physiology 43:1252-68. https://doi.org/10.1104/pp.106.094425

Mansfeld BN, Grumet R (2018) QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. The Plant Genome. https://doi.org/10.3835/plantgenome2018.01.0006

Matsui K, Umemura Y, Ohme-Takagi M (2008) Atmybl2, a protein with a single myb domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. The Plant Journal 55:954-67. https://doi.org/10.1111/j.1365-313X.2008.03565.x

Mazza G, Brouillard R (1987) Color stability and structural transformations of cyanidin 3, 5-diglucoside and four 3-deoxyanthocyanins in aqueous solutions. Journal of Agricultural and Food Chemistry 35:422-6.

McCormick RF (2017) High-throughput genotyping analyses and image-based phenotyping in *Sorghum bicolor*. Dissertation, Texas A&M University

Mellway RD, Tran LT, Prouse MB, Campbell MM, Constabel CP (2009) The wound-, pathogen-, and ultraviolet b-responsive myb134 gene encodes an r2r3 myb transcription factor that regulates proanthocyanidin synthesis in poplar. Plant Physiology 150:924-41. https://doi.org/10.1104/pp.109.139071

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America 88:9828-9832. https://doi.org/10.1073/pnas.88.21.9828

Miranda CL, Maier CS, Stevens JF (2012) Flavonoids. eLS. https://doi.org/10.1002/9780470015902.a0003068.pub2

Mizuno H, Kawahigashi H, Kawahara Y, Kanamori H, Ogata J, Minami H, Itoh T, Matsumoto T (2012) Global transcriptome analysis reveals distinct expression among duplicated genes during sorghum-*Bipolaris sorghicola* interaction. BMC plant biology 12:1-5. https://doi.org/10.1186/1471-2229-12-121

Mizuno H, Yazawa T, Kasuga S, Sawada Y, Kanamori H, Ogo Y, Hirai MY, Matsumoto T, Kawahigashi H (2016) Expression of flavone synthase II and flavonoid 3′-hydroxylase is associated with color variation in tan-colored injured leaves of sorghum. Frontiers in Plant Science 7:1718. https://doi.org/10.3389/fpls.2016.01718

Mizuno H, Yazawa T, Kasuga S, Sawada Y, Ogata J, Ando T, Kanamori H, Yonemaru JI, Wu J, Hirai MY, Matsumoto T (2014) Expression level of a flavonoid 3′-hydroxylase gene determines pathogen-induced color variation in sorghum. BMC research notes 7:1-2. https://doi.org/10.1186/1756-0500-7-761

Morishige DT, Klein PE, Hilley JL, Sahraeian SM, Sharma A, Mullet JE (2013) Digital genotyping of sorghum - a diverse plant species with a large repeat-rich genome. BMC Genomics 14:448. https://doi.org/10.1186/1471-2164-14-448

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PF, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences of the United States of America 110:453– 458. https://doi.org/10.1073/pnas.1215985110

Mundia CW, Secchi S, Akamani K, Wang G (2019) A regional comparison of factors affecting global sorghum production. Sustainability 11.7:2135 https://doi.org/10.3390/su11072135

Nakatsuka T, Nishihara M (2010) Udp-glucose: 3-deoxyanthocyanidin 5-O-glucosyltransferase from *Sinningia cardinalis*. Planta 232:383-92. https://doi.org/10.1007/s00425-010-1175-0

Nejat N, Cahill DM, Vadamalai G, Ziemann M, Rookes J, Naderali N (2015) Transcriptomics-based analysis using rna-seq of the coconut (*Cocos nucifera*) leaf in response to yellow decline phytoplasma infection. Molecular Genetics and Genomics 290:1899-910. https://doi.org/10.1007/s00438-015-1046-2

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The tt8 gene encodes a basic helix-loop-helix domain protein required for expression of dfr and ban genes in *Arabidopsis* siliques. The Plant Cell 12:1863-78. https://doi.org/10.1105/tpc.12.10.1863

Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L (2001) The *Arabidopsis* tt2 gene encodes an r2r3 myb domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. The Plant Cell 13:2099-114. https://doi.org/10.1105/TPC.010098

Nida H, Girma G, Mekonen M, Lee S, Seyoum A, Dessalegn K, Tadesse T, Ayana G, Senbetay T, Tesso T, Ejeta G (2019) Identification of sorghum grain mold resistance loci through genome wide association mapping. Journal of Cereal Science 85:295-304. https://doi.org/10.1016/j.jcs.2018.12.016

Nielsen KA, Gotfredsen CH, Buch-Pedersen MJ, Ammitzbøll H, Mattsson O, Duus JØ, Nicholson RL (2004) Inclusions of flavonoid 3-deoxyanthocyanidins in *Sorghum bicolor* self-organize into spherical structures. Physiological and Molecular Plant Pathology 65:187-96. https://doi.org/10.1016/j.pmpp.2005.02.001

Nguyen NH, Jeong CY, Kang GH, Yoo SD, Hong SW, Lee H (2015) Mybd employed by hy 5 increases anthocyanin accumulation via repression of mybl 2 in *Arabidopsis*. The Plant Journal 84:1192-205. https://doi.org/10.1111/tpj.13077

Olsen KM, Wendel JF (2013) Crop plants as models for understanding plant adaptation and diversification. Frontiers in Plant Science 4:290. https://doi.org/10.3389/fpls.2013.00290

Olson A, Klein RR, Dugas DV, Lu Z, Regulski M, Klein PE, Ware D (2014) Expanding and vetting *Sorghum bicolor* gene annotations through transcriptome and methylome sequencing. The Plant Genome 7:plantgenome2013-08. https://doi.org/10.3835/plantgenome2013.08.0025

Ormrod DP, Landry LG, Conklin PL (1995) Short-term uv-b radiation and ozone exposure effects on aromatic secondary metabolite accumulation and shoot growth of flavonoid-deficient *Arabidopsis* mutants. Physiologia Plantarum 93:602-10. https://doi.org/10.1111/j.1399-3054.1995.tb05106.x

Pandey N, Pandey-Rai S (2014) Short term uv-b radiation-mediated transcriptional responses and altered secondary metabolism of in vitro propagated plantlets of *Artemisia annua* L. Plant Cell, Tissue and Organ Culture 116:371-85. https://doi.org/10.1007/s11240-013-0413-0

Pandey N, Goswami N, Tripathi D, Rai KK, Rai SK, Singh S, Pandey-Rai S (2019) Epigenetic control of uv-b-induced flavonoid accumulation in *Artemisia annua* L. Planta 249:497-514. https://doi.org/10.1007/s00425-018-3022-7

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457.7229:551-556. https://doi.org/10.1038/nature07723

Paulsmeyer M, Chatham L, Becker T, West M, West L, Juvik J (2017) Survey of anthocyanin composition and concentration in diverse maize germplasms. Journal of Agricultural and Food Chemistry 65:4341-50. https://doi.org/10.1021/acs.jafc.7b00771

Petti C, Kushwaha R, Tateno M, Harman-Ware AE, Crocker M, Awika J, DeBolt S (2014) Mutagenesis breeding for increased 3-deoxyanthocyanidin accumulation in leaves of *Sorghum bicolor* (L.). Journal of Agricultural and Food Chemistry 62:1227-32. https://doi.org/10.1021/jf405324j

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science 253:895-7. https://doi.org/10.1126/science.253.5022.895

Pfeiffer BK (2014) Genetic and environmental influences on the inheritance of sorghum with a black pericarp. Thesis, Texas A&M University

Pfeiffer BK (2017) The improvement of grain sorghum productivity, black pericarp color, and protein digestibility. Dissertation, Texas A&M University

Pfeiffer BK, Rooney WL (2015) Sunlight induces black color and increases flavonoid levels in the grain of sorghum line Tx3362. Crop Science 55:1703-11. https://doi.org/10.2135/cropsci2014.11.0757

Pfeiffer BK, Rooney WL (2016) Inheritance of pericarp color, nutritional quality, and grain composition traits in black sorghum. Crop Science 56.1:164-172. https://doi.org/10.2135/cropsci2015.04.0224

Portu J, López R, Santamaría P, Garde-Cerdán T (2018) Methyl jasmonate treatment to increase grape and wine phenolic content in tempranillo and graciano varieties during two growing seasons. Scientia Horticulturae 240:378-86. https://doi.org/10.1016/j.scienta.2018.06.019

Quarrie SA, Lazić-Jančić V, Kovačević D, Steed A, Pekić S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. Journal of Experimental Botany 50:1299–1306. https://doi.org/10.1093/jxb/50.337.1299

Ravaglia D, Espley RV, Henry-Kirk RA, Andreotti C, Ziosi V, Hellens RP, Costa G, Allan AC (2013)Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of r2r3 myb transcription factors. BMC Plant Biology 13:1-4. https://doi.org/10.1186/1471-2229-13-68

Rhodes DH, Hoffmann Jr L, Rooney WL, Ramu P, Morris GP, Kresovich S (2014) Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. Journal of Agricultural and Food Chemistry 62:10916-27. https://doi.org/10.1021/jf503651t

Rooney WL, Rooney LW, Awika J, Dykes L (2013) Registration of Tx3362 sorghum germplasm. Journal of Plant Registration 7:104−107. https://doi.org/10.3198/jpr2012.04.0262crg

Rooney LW, Miller FR, Mertin JV (1981) Variation in the structure and kernel characteristics of sorghum. Proceedings of the International Symposium on Sorghum Grain Quality 28:143-162.

Roy S (2016) Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. Plant Signaling & Behavior 11:e1117723. https://doi.org/10.1080/15592324.2015.1117723

Smith CW, Frederiksen RA (2000) Sorghum: origin, history, technology, and production. John Wiley & Sons (Vol. 2). Hoboken, NJ, pp 272

Rausher MD (2006) The evolution of flavonoids and their genes. The Science of Flavonoids. Springer, New York, NY, pp 175-211

Reddy VS, Goud KV, Sharma R, Reddy AR (1994) Ultraviolet-b-responsive anthocyanin production in a rice cultivar is associated with a specific phase of phenylalanine ammonia lyase biosynthesis. Plant Physiology 105:1059-66. https://doi.org/10.1104/pp.105.4.1059

Rius SP, Emiliani J, Casati P (2016) P1 epigenetic regulation in leaves of high altitude maize landraces: effect of uv-b radiation. Frontiers in Plant Science 7:523. https://doi.org/10.3389/fpls.2016.00523

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of rna-seq data. Genome Biology 11:1-9. https://doi.org/10.1186/gb-2010-11-3-r25

Rosinski JA, Atchley WR (1998) Molecular evolution of the myb family of transcription factors: evidence for polyphyletic origin. Journal of Molecular Evolution 46:74-83. https://doi.org/10.1007/PL00006285

Ruiz-García Y, Gómez-Plaza E (2013) Elicitors: a tool for improving fruit phenolic content. Agriculture 3:33-52. https://doi.org/10.3390/agriculture3010033

Ryan KG, Swinny EE, Markham KR, Winefield C (2002) Flavonoid gene expression and uv photoprotection in transgenic and mutant petunia leaves. Phytochemistry 59:23-32. https://doi.org/10.1016/S0031-9422(01)00404-6

Ryan CA, Pearce G (2010) Peptide hormones for defense, growth, development and reproduction. Plant Hormones. Springer, Dordrecht, pp 700-716

Sainz MB, Grotewold E, Chandler VL (1997) Evidence for direct activation of an anthocyanin promoter by the maize c1 protein and comparison of dna binding by related myb domain proteins. The Plant Cell 9:611-25. https://doi.org/10.1105/tpc.9.4.611

Saewan N, Jimtaisong A (2013) Photoprotection of natural flavonoids. Journal of Applied Pharmaceutical Science 3:129-41. https://doi.org/10.7324/JAPS.2013.3923

Sak K (2014) Cytotoxicity of dietary flavonoids on different human cancer types. Pharmacognosy Reviews 8:122. https://doi.org/10.4103%2F0973-7847.134247

Samanta A, Das G, Das SK (2011) Roles of flavonoids in plants. Carbon 100:12-35.

Serna L (2007) Bhlh proteins know when to make a stoma. Trends in Plant Science 12:483-485. https://doi.org/10.1016/j.tplants.2007.08.016

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498-504. http://www.genome.org/cgi/doi/10.1101/gr.1239303

Sharma M, Chai C, Morohashi K, Grotewold E, Snook ME, Chopra S (2012) Expression of flavonoid 3'-hydroxylase is controlled by p1, the regulator of 3-deoxyflavonoid biosynthesis in maize. BMC Plant Biology 12:1-3. https://doi.org/10.1186/1471-2229-12-196

Shih CH, Chu IK, Yip WK, Lo C (2006) Differential expression of two flavonoid 3′-hydroxylase cdnas involved in biosynthesis of anthocyanin pigments and 3-deoxyanthocyanidin phytoalexins in sorghum. Plant and Cell Physiology 47:1412-9. https://doi.org/10.1093/pcp/pcl003

Shih CH, Siu SO, Ng R, Wong E, Chiu LC, Chu IK, Lo C (2007) Quantitative analysis of anticancer 3-deoxyanthocyanidins in infected sorghum seedlings. Journal of Agricultural and Food Chemistry 55:254-9. https://doi.org/10.1021/jf062516t

Snyder BA, Nicholson RL (1990) Synthesis of phytoalexins in sorghum as a site-specific response to fungal ingress. Science 248:1637-9. https://doi.org/10.1126/science.248.4963.1637

Son GH, Wan J, Kim HJ, Nguyen XC, Chung WS, Hong JC, Stacey G (2012) Ethylene-responsive element-binding factor 5, erf5, is involved in chitin-induced innate immunity response. Molecular Plant-Microbe Interactions 25:48-60. https://doi.org/10.1094/MPMI-06-11-0165

Somssich IE, Hahlbrock K (1998) Pathogen defence in plants—a paradigm of biological complexity. Trends in Plant Science 3:86-90. https://doi.org/10.1016/S1360-1385(98)01199-6

Stafford HA (1994) Anthocyanins and betalains: evolution of the mutually exclusive pathways. Plant Science 101:91-8. https://doi.org/10.1016/0168-9452(94)90244-5

Stephens JC, Karper RE (1965) Release of breeding stocks of male sterilized grain sorghum lines. Texas Agricultural Experimental Station. Miscellaneous Publication MP-758.

Stich K, Forkmann G (1988) Biosynthesis of 3-deoxyanthocyanins with flower extracts from *Sinningia cardinalis*. Phytochemistry 27:785-9. https://doi.org/10.1016/0031-9422(88)84093-7

Sweeny J, Iacobucci G (1981) Synthesis of anthocyanidins-III1: total synthesis of apigeninidin and luteolinidin chlorides. Tetrahedron 37:1481-1483. https://doi.org/10.1016/S0040-4020(01)92086-1

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H (2013) Qtl-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of dna from two bulked populations. The Plant Journal 74.1:174-83. https://doi.org/10.1111/tpj.12105

Taleon V, Dykes L, Rooney WL, Rooney LW (2012) Effect of genotype and environment on flavonoid concentration and profile of black sorghum grains. Journal of Cereal Science 56:470-5. https://doi.org/10.1016/j.jcs.2012.05.001

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verries C, Cheynier V, Romieu C (2009) Ectopic expression of vvmybpa2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiology 149:1028-41. https://doi.org/10.1104/pp.108.131862

Tilbrook K, Arongaus AB, Binkert M, Heijde M, Yin R, Ulm R (2013) The uvr8 uv-b photoreceptor: perception, signaling and response. The Arabidopsis Book/American Society of Plant Biologists 11. https://doi.org/10.1199/tab.0164

Toledo-Ortiz G, Huq E, Quail PH (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. The Plant Cell 15:1749-70. https://doi.org/10.1105/tpc.013839

Tsao R (2010) Chemistry and biochemistry of dietary polyphenols. Nutrients 2:1231-1246. https://doi.org/10.3390/nu2121231

Turnbull JJ, Sobey WJ, Aplin RT, Hassan A, Firmin JL, Schofield CJ, Prescott AG (2000) Are anthocyanidins the immediate products of anthocyanidin synthase. Chemical Communications 2000:2473-4. https://doi.org/10.1039/B007594I

Ulm R, Nagy F (2005) Signalling and gene regulation in response to ultraviolet light. Current Opinion in Plant Biology 8:477-82. https://doi.org/10.1016/j.pbi.2005.07.004

Umer MJ, Safdar LB, Gebremeskel H, Zhao S, Yuan P, Zhu H, Kaseb MO, Anees M, Lu X, He N, Gong C (2020) Identification of key gene networks controlling organic acid and sugar metabolism during watermelon fruit development by integrating metabolic phenotypes and gene expression profiles. Horticulture Research 7:1-3. https://doi.org/10.1038/s41438-020-00416-8

Ubi BE, Honda C, Bessho H, Kondo S, Wada M, Kobayashi S, Moriguchi T (2006) Expression analysis of anthocyanin biosynthetic genes in apple skin: effect of uv-b and temperature. Plant Science 170:571-8. https://doi.org/10.1016/j.plantsci.2005.10.009

Valencia RC, Rooney WL. Genetic control of sorghum grain color. International Sorghum and Millet Collaborative Research Support Program (INTSORMIL-CENTA) Presentations. 2009.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E (2013) From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics 43.1:11-0. https://doi.org/10.1002/0471250953.bi1110s43

Vogel G, LaPlant KE, Mazourek M, Gore MA, Smart CD (2020) A combined bsa-seq and linkage mapping approach identifies genomic regions associated with *Phytophthora* root and crown rot resistance in squash. Theoretical and Applied Genetics 14:1-7. https://doi.org/10.1007/s00122-020-03747-1

Villar M, Lefevre F, Bradshaw HD Jr, du-Cros ET (1996) Molecular genetics of rust resistance in poplars (*Melampsora larici-populina* Kleb/*Populus* sp.) by bulked segregant analysis in a 2 × 2 factorial mating design. Genetics 143:531–536

Van Ooijen J (2018) JoinMap® 5, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen.

van Treuren R (2001) Efficiency of reduced primer selectivity and bulked dna analysis for the rapid detection of aflp polymorphisms in a range of crop species. Euphytica 117:27–37. https://doi.org/10.1023/A:1004003121622

Vinall HN, Cron AB (1921) Improvement of sorghums by hybridization. Journal of Heredity 12:435-43. https://doi.org/10.1093/oxfordjournals.jhered.a102043

Viswanathan R, Mohanraj D, Padmanaban P, Alexander KC (1996) Accumulation of 3-deoxyanthocyanidin phytoalexins luteolinidin and apigeninidin in sugarcane in relation to red rot disease. Indian Phytopathology 49:174-5.

Wang J, Qian J, Yao L, Lu Y (2015) Enhanced production of flavonoids by methyl jasmonate elicitation in cell suspension culture of *Hypericum perforatum*. Bioresources and Bioprocessing 2:1-9. https://doi.org/10.1186/s40643-014-0033-5

Waniska RD, Rooney LW (2000) Structure and chemistry of the sorghum caryopsis. Sorghum: origin, history, technology, and production 2:649-79.

Weiergang I, Hipskind JD, Nicholson RL (1996) Synthesis of 3-deoxyanthocyanidin phytoalexins in sorghum occurs independent of light. Physiological and Molecular Plant Pathology 49:377-88. https://doi.org/10.1006/pmpp.1996.0060

Winkel-Shirley B (2001) Flavonoid biosynthesis a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology 126:485-93. https://doi.org/10.1104/pp.126.2.485

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Current Opinion in Plant Biology 5:218-23. https://doi.org/10.1016/S1369-5266(02)00256-X

Winefield CS, Lewis DH, Swinny EE, Zhang H, Arathoon HS, Fischer TC, Halbwirth H, Stich K, Gosch C, Forkmann G, Davies KM (2005) Investigation of the biosynthesis of 3-deoxyanthocyanins in *Sinningia cardinalis*. Physiologia Plantarum 124:419-30. https://doi.org/10.1111/j.1399-3054.2005.00531.x

Wu G, Bornman JF, Bennett SJ, Clarke MW, Fang Z, Johnson SK (2017) Individual polyphenolic profiles and antioxidant activity in sorghum grains are influenced by very low and high solar uv radiation and genotype. Journal of Cereal Science 77:17-23. https://doi.org/10.1016/j.jcs.2017.07.014

Wu Y, Li X, Xiang W, Zhu C, Lin Z, Wu Y, Li J, Pandravada S, Ridder DD, Bai G, Wang ML (2012) Presence of tannins in sorghum grains is conditioned by different natural alleles of tannin1. Proceedings of the National Academy of Sciences 109:10281-6. https://doi.org/10.1073/pnas.1201700109

Xiong, Y, Zhang P, Warner RD, Fang Z (2019) 3-deoxyanthocyanidin colorant: nature, health, synthesis, and food applications. Comprehensive Review in Food Science and Food Safety 18:1533−1549. https://doi.org/10.1111/1541-4337.12476

Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by myb–bhlh–wdr complexes. Trends in Plant Science. 2015 Mar 1;20(3):176-85. https://doi.org/10.1016/j.tplants.2014.12.001

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Science 48:391-407. https://doi.org/10.2135/cropsci2007.04.0191

Yamaguchi Y, Huffaker A (2011) Endogenous peptide elicitors in higher plants. Current Opinion in Plant Biology 14:351-7. https://doi.org/10.1016/j.pbi.2011.05.001

Yang WX, Jie XL (2007) Sunlight effect on pigment content of purple wheat. Journal of Triticeae Crops 2:1101-5.

Yang Q, Reinhard K, Schiltz E, Matern U (1997) Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA: anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. Plant Molecular Biology 35:777-89. https://doi.org/10.1023/A:1005878622437

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R (2010) The origin and evolution of plant flavonoid metabolism. Frontiers in Plant Science 10:943. https://doi.org/10.3389/fpls.2019.00943

Zhang J, Xu Y, Wu X, Zhu L (2002) A bentazon and sulfonylurea sensitive mutant: breeding, genetics and potential application in seed production of hybrid rice. Theoretical Applied Genetics 105:16–22. https://doi.org/10.1007/s00122-002-0874-8

Zhang L, Zhang Q, Li W, Zhang S, Xi W (2019) Identification of key genes and regulators associated with carotenoid metabolism in apricot (*Prunus armeniaca*) fruit using weighted gene coexpression network analysis. BMC Genomics 20:1-5. https://doi.org/10.1186/s12864-019-6261-5

Zhu L, Zhou Y, Li X, Zhao J, Guo N, Xing H (2018) Metabolomics analysis of soybean hypocotyls in response to *Phytophthora sojae* infection. Frontiers in Plant Science 9:1530. https://doi.org/10.3389/fpls.2018.01530

Zhu J, Zhao W, Li R, Guo D, Li H, Wang Y, Mei W, Peng S (2021) Identification and characterization of chalcone isomerase genes involved in flavonoid production in *Dracaena cambodiana*. Frontiers in Plant Science 12:226. https://doi.org/10.3389/fpls.2021.616396

Zhu F (2018) Anthocyanins in cereals: composition and health effects. Food Research International 109:232-49. https://doi.org/10.1016/j.foodres.2018.04.015

Zoratti L, Karppinen K, Luengo Escobar A, Häggman H, Jaakola L (2014) Light-controlled flavonoid biosynthesis in fruits. Frontiers in Plant Science 5:534. https://doi.org/10.3389/fpls.2014.00534

APPENDIX

SUPPLEMNTARY MATERIAL

Supplementary Figures

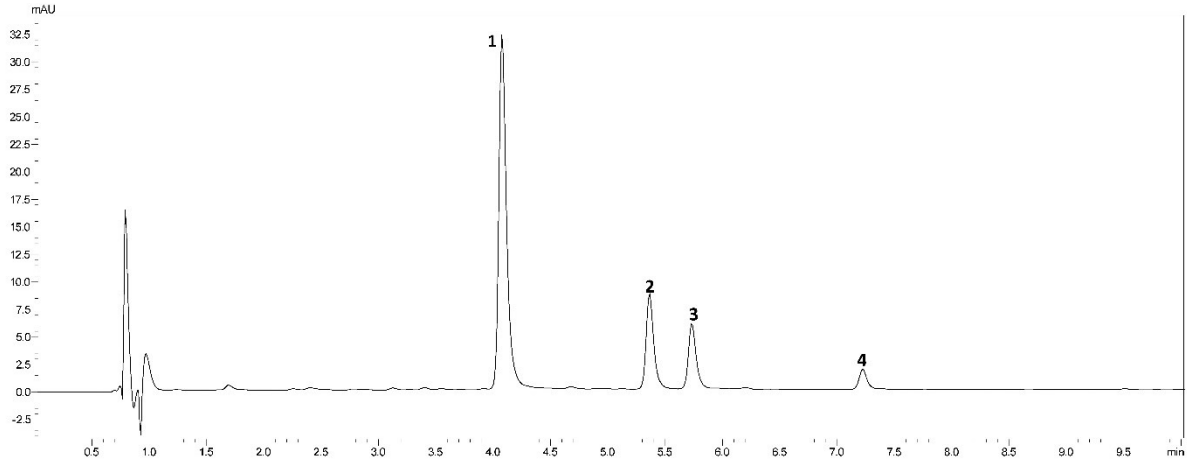


Figure S1. Ultra-Performance Liquid Chromatography with Photodiode Array Detection (UPLC-PDA) chromatogram of mature seed from RTx3362 exposed to VIS + UVA + UVB light showing representative peaks of (1) luteolinidin; (2) apigeninidin; (3) 5-methoxy-luteolinidin; and (4) 7-methoxy-apigeninidin. PDA = 485 nm.

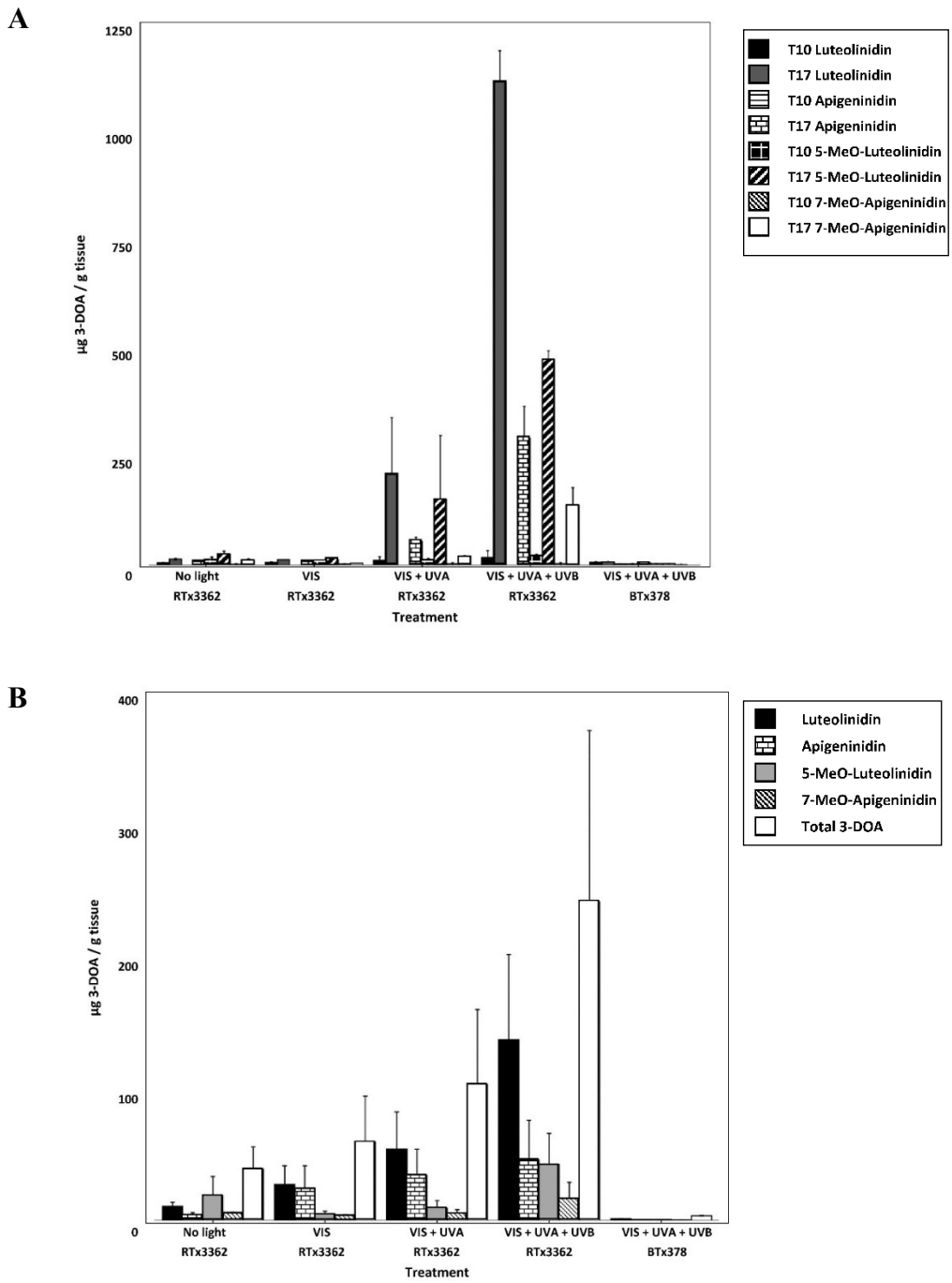


Figure S2. A) Concentration of 3-deoxyanthocyanidins (3-DOA) luteolinidin, apigeninidin, 5-methoxy-luteolinidin (5-MeO-Luteolinidin), and 7-methoxy-apigeninidin (7-MeO-Apigeninidin) during sorghum pericarp development at 10 days after treatment (T10) and 17 days after treatment (T17) in black (RTx3362) and red (BTx378) sorghum genotypes under different light treatments. Error bars represent standard deviation from the average of the replications. Statistical differences for the accumulation of each compound across treatments are summarized in Table 3. B) Concentration of individual and total 3-DOAs in mature sorghum grain in black (RTx3362) and red (BTx378) sorghum genotypes under different light treatments. Error bars represent standard deviation from the average of the replications. Statistical differences for the accumulation of each compound across treatments are summarized in Table 5.

Supplementary Tables

Table S1. A summary of the mapping statistics from the 64 RNA-seq datasets including 52 pericarp, 6 flag leaf, and 6 root samples.

Genotype	Tissue	Time Point	Treatment	Rep	Total reads (M)	Reads mapped in pairs (M)	Reads Mapped (%)	Mapped to intergenic region (%)	Mapped to genes (%)
RTx3362	Pericarp	T10	No light	1	59.71	51.22	85.77	8.73	91.27
				2	53.39	49.25	92.25	8.33	91.67
				3	62.31	56.22	90.22	9.18	90.82
				4	41.91	33.7	80.4	4.39	95.61
				5	41.13	33.65	81.82	4.23	95.77
				6	47.26	38.47	81.41	3.96	96.04
RTx3362	Pericarp	T10	VIS	1	65.2	59.67	91.53	10.45	89.55
				2	52.18	44.35	84.99	7.77	92.23
				3	58.46	54.81	93.77	9.25	90.75
				4	48.09	39.31	81.74	4.23	95.77
				5	50.38	41.33	82.04	4.7	95.3
				6	43.63	35.08	80.39	4.3	95.7
RTx3362	Pericarp	T10	VIS + UVA	1	42.03	36.29	86.36	8.66	91.34
				2	51.59	43.97	85.22	8.73	91.27
				3	69.44	62.66	90.24	9.05	90.95
				4	46.18	37.7	81.63	4.47	95.53
				5	47.57	39.21	82.43	4.61	95.39
				6	45.36	37.02	81.61	3.79	96.21
RTx3362	Pericarp	T10	VIS + UVA + UVB	1	60.88	55.09	90.49	9.65	90.35
				2	47.45	41.38	87.2	7.7	92.3
				3	73.37	65.96	89.89	9.52	90.48
				4	46.17	37.95	82.19	4.63	95.37
				5	40.76	33.57	82.37	4.7	95.3
				6	39.06	32.22	82.48	4.19	95.81
BTx378	Pericarp	T10	VIS + UVA + UVB	1	48.89	41.51	84.9	5.41	94.59
				2	44.71	36.74	82.17	6.78	93.22
RTx3362	Pericarp	T17	No light	1	59.75	53.86	90.14	8.97	91.03
				2	68.55	58.81	85.79	10.03	89.97
				3	51.69	43.68	84.49	9.67	90.33
				4	44.69	37.06	82.94	8.43	91.57
				5	42.67	34.67	81.24	3.79	96.21
				6	50.40	41.39	82.11	3.64	96.36
RTx3362	Pericarp	T17	VIS	1	52.32	43.94	83.99	4.86	95.14
				2	40.79	35.87	87.92	5.6	94.4
				3	70.11	62.26	88.80	8.09	91.91
				4	50.87	41.19	80.97	4.8	95.2
				5	56.65	45.53	80.37	3.81	96.19
				6	48.61	39.34	80.92	4.03	95.97
RTx3362	Pericarp	T17	VIS + UVA	1	48.50	41.59	85.75	8.94	91.06
				2	55.18	46.89	84.98	9.68	90.32
				3	53.39	47.54	89.04	8.63	91.37
				4	41.93	34.32	81.85	4.64	95.36
				5	47.30	38.26	80.89	4.09	95.91
				6	53.18	43.57	81.94	3.72	96.28
RTx3362	Pericarp	T17	VIS + UVA + UVB	1	48.53	41.24	84.98	8.63	91.37
				2	51.64	46.66	90.36	8.39	91.61
				3	42.69	37.22	87.19	8.81	91.19
				4	39.11	32.14	82.19	3.96	96.04
				5	53.13	42.88	80.70	4.16	95.84
				6	50.10	40.87	81.57	6.28	93.72
BTx378	Pericarp	T17	VIS + UVA + UVB	1	62.98	55.45	88.04	6.58	93.42
				2	61.48	50.03	81.38	9.48	90.52
RTx3362	Flag leaf	T10	VIS + UVA + UVB	1	73.7	63.1	85.68	3.8	96.2
				2	56.7	48.9	86.28	3.69	96.31
				3	58.6	51	87.16	3.57	96.43
RTx3362	Root	--	No light	1	54	47.4	87.74	3.36	96.64
		--		2	57.2	50.2	87.96	3.04	96.96
		--		3	66.2	58.2	87.84	3.3	96.7
BTx378	Flag leaf	T10	VIS + UVA + UVB	1	58.1	49.5	85.2	3.79	96.21
				2	54.4	46.9	86.3	3.83	96.17
				3	75.8	65.8	86.82	3.85	96.15
BTx378	Root	--	No light	1	60.1	52.1	86.6	4.11	95.89
		--		2	66.8	58.9	88.08	3.57	96.43
		--		3	67.8	59.6	87.96	3.53	96.47

Table S2. Pairwise differential gene expression analysis between all pericarp, flag leaf, and root treatments.

RNAseq_DEGS.xlsx

Table S3. Sobic gene identifiers (Sbicolor_454 v3.0.1) to 42 hub genes discovered in the network analysis.

Gene	Annotation
Sobic.009G022500	ABC transporter G family member 16
Sobic.004G181000	Monooxygenase
Sobic.010G259300	Benzyl alcohol *O*-benzoyltransferase
Sobic.003G422200	Glucan endo-1,3-beta-D-glucosidase
Sobic.006G054900	Uncharacterized
Sobic.010G066700	Hydroxycinnamoyltransferase 4
Sobic.003G244600	Chitinase-like
Sobic.003G278000	Uncharacterized
Sobic.003G203200	Arabinogalactan Peptide 22-like
Sobic.010G259500	Benzyl alcohol *O*-benzoyltransferase
Sobic.009G234100	WRKY34
Sobic.001G400800	Pathogenesis-related protein 10c
Sobic.003G288300	Uncharacterized
Sobic.001G496500	MTN26L2 - MtN26 family protein precursor
Sobic.010G259600	Benzyl alcohol *O*-benzoyltransferase
Sobic.001G400700	Pathogenesis-related protein 1
Sobic.003G331700	Permatin precursor
Sobic.004G296900	Ethylene-responsive transcription factor 5
Sobic.003G355800	Bowman-Birk serine protease inhibitor
Sobic.006G178500	Lysine Histidine Transporter-like 8
Sobic.001G196300	Cytochrome P450 84A1
Sobic.003G014400	Uncharacterized
Sobic.005G166200	MTN26L5 - MtN26 family protein precursor
Sobic.005G214400	Glycerol-3-Phosphate Acyltransferase 3
Sobic.010G066601	Hydroxycinnamoyltransferase 4
Sobic.001G322100	Uncharacterized
Sobic.001G400900	similar to Pathogenesis-related protein 10a
Sobic.003G204900	Uncharacterized
Sobic.009G044100	Pectinesterase Inhibitor 8
Sobic.007G155000	Alpha Carbonic Anhydrase 7
Sobic.007G145600	4-coumarate--CoA Ligase 5
Sobic.001G453100	Homocysteine S-methyltransferase 1
Sobic.003G286600	Acid phosphatase-like
Sobic.004G141200	Cytochrome P450 CYP73A100
Sobic.005G215900	Putative Bowman-Birk serine protease inhibitor
Sobic.003G267300	ABC transporter B family member 11
Sobic.009G064400	Cytochrome P450 71A1
Sobic.004G200833	Flavonoid 3'-Monooxygenase
Sobic.005G137200	Chalcone synthase WHP1
Sobic.001G401800	Uncharacterized
Sobic.005G216000	Bowman-Birk serine protease inhibitor
Sobic.010G092200	Sphinganine C4-Monooxygenase 1