GENETIC RESISTANCE TO FUNGAL PATHOGENS IN SORGHUM [*SORGHUM BICOLOR* (L.) MOENCH]

by

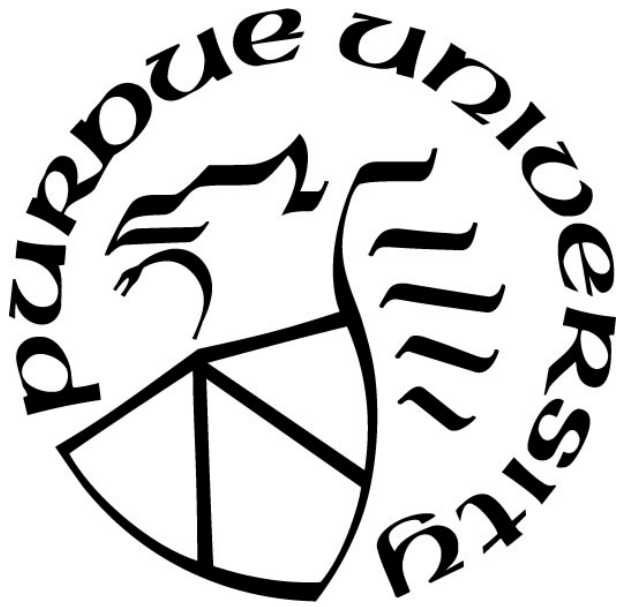
Habte Nida Chikssa

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Botany and Plant Pathology

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Tesfaye Mengiste, Chair
Department of Botany and Plant Pathology

Dr. Gebisa Ejeta
Department of Agronomy

Dr. Guri Johal
Department of Botany and Plant Pathology

Dr. Jianxin Ma
Department of Agronomy

Approved by:
Dr. Christopher J Staiger

Dedicated to my father Nida Chikssa and brother Beyene Nida

ACKNOWLEDGMENTS

I would like to express my deep and special gratitude to my advisors Prof Tesfaye Mengiste and Prof Gebisa Ejeta for their overwhelming humbleness, inspiration and continued support throughout my study. I was so lucky to have these two great mentors with diverse experience and kind personality. I am really thankful to them. I would also thank my graduate committee members Prof Guri Johal and Prof Jianxin Ma for their valuable advice and encouragement that have been additional energy to make significant progress in my graduate research. I thank Prof Tesfaye Tesso for his technical support and encouragement. I am very much grateful to Dr Sanghun Lee for his exceptional kindness and support in molecular techniques. He was always there to help in times when I was stuck with some of the laboratory procedures that I was not familiar to. A special thank also goes to Dr Gezahegn Girma for his unreserved support in bioinformatics without which the association and mapping studies wouldn't be possible. I am very much thankful to Dr Demeke Mewa and Dr Adeyanju Adedayo (Dayo) for their valuable support in bioinformatics as well as in various laboratory and greenhouse activities. I would also thank Dr Chao-Jan Liao, Sara Gebremeskel, Dr Carol Bvindi and Dr Namrata Jaiswal for their technical support and valuable advice throughout my study. I am also thankful to all the other previous and current members of Mengiste and Gebisa's lab who helped me with various technical issues during my study. My sincere thanks go to the Ethiopian community in West Lafayette and Lafayette for their encouragement and support during my stay at Purdue. I have never felt alone nor missed home and family because of their kindness and hospitality.

A special thank goes to Dr Alemu Tirfessa for his continued support in every aspect both in my research career and during this study. I would never been in such level without his effort and trust in me from the very beginning that he offered me an opportunity to join the sorghum program in 2013. I am very much thankful to the sorghum research staff members at Melkassa and Chiro: Gash Tese (Tesema Woldu) and Gash Tile (Tilahun Hadis), Moges Mekonen, Amare Seyoum, Amare Nega, Adane G/Yohannes, Tamrat Bejiga, Dr Solomon Zewdu, Kinde Nouh, Mesfin Bekele, Mohamod Salah, Sewmehon Siraw, Belayneh Wendim, Daniel Nadew, Ligaba Ayele, Temesegen Teressa, Zigale Semahegn, Kidanemariam Wagaw, Temesgen Begna, Chalachew Endalamaw, Hilemariam Solomon, Wakjira Chifra, Fantahun Abebe, Sintayehu Hailu, Teshome, Etetu, Shito, Genet, Damenu and Eyerus for their technical support during my study and

for the excellent tradition of team work and social set up that makes working in the team so enjoyable. I am very much grateful to our driver Girma Mamo for his exceptional kindness and support. I am also thankful to the drivers Girma Akalewold, Tesfaye Terefe, Daniel Bekele, Bedo Dugo, Fikadu Lemi and a number of others. I am grateful to the sorghum research team in collaborating centers at Bako and Jimma: Kebede Desalegn, Chemeda Birhanu, Tsegau Senbetay, Dr Dagnachew Lule and technical staff members for their technical support during field study. It wouldn't be possible to do the field work without their support. I am also grateful to Dr Getachew Ayana for his technical and administrative support as a SMIL coordinator. I am thankful to Dr Bedru Beshir and Dr Mohamed Yusuf for their support in administrative issues. I thank Dr Mandefro Nigussie, Dr Diriba Geleti, Dr Taye Tadesse, Dr Taye Tessema as well as both MARC and EIAR management and administrative staff for their support with various administrative issues during my study.

I am thankful for the financial support by the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through American People provided to the United States Agency for International Development (USAID).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	analysis of variance
AOS	allene oxide synthase
ARG4	ANTHRACNOSE RESISTANCE GENE 4
ARG5	ANTHRACNOSE RESISTANCE GENE 5
bHLH	basic helix–loop–helix
BLINK	bayesian-information and linkage-disequilibrium iteratively nested keyway
BSA	bulked-segregant analysis
cAMP	cyclic adenosine monophosphate
CC	coiled coil
CMLM	compressed mixed linear model
DEGs	differentially expressed genes
DTF	days to flowering
FarmCPU	fixed and random model circulating probability unification
FDR	false discovery rate
FGMR	field (plot) grain mold rating
FHB	fusarium head blight
GAPIT	genome association and prediction integrated tool
GBS	genotyping by sequencing
GLM	general linear model
GO	gene ontology
GRAB1	geminivirus rep a-binding 1
GWAS	genome wide association study
HCAAs	hydroxycinnamic acid amides
Hpi	hours post inoculation
IGV	integrative genomics viewer
KEGG	kyoto encyclopedia of genes and genomes
LEA	late embryogenesis abundant
LogFC	log fold change
LRR	leucine-rich-repeat

MAF	minor allele frequency
MAPKs	mitogen-activated protein kinases
MBW	MYB–bHLH–WDR
MLM	mixed linear model
MYB	myeloblastosis
NBS	nucleotide-binding-site
PCA	principal components analysis
PGMR	panicle grain mold rating
PHT	plant height
PR	pathogenesis-related
PRRs	pattern recognition receptors
PTI	PAMP triggered immunity
Q-Q	quantile-quantile plot
qRT-PCR	quantitative real-time PCR
RILs	recombinant inbred lines
RLCK	receptor like cytoplasmic kinases
SNP	single nucleotide polymorphism
TAG	triacylglycerol
TBT	tags by taxa
TGMR	threshed grain mold rating
TOPM	tags on physical map
UTR	untranslated region
WebMeV	multiple experiment viewer

ABSTRACT

Sorghum [*Sorghum bicolor* (L.) Moench] is the fifth most widely grown cereal crop in the world that serves as a staple food for millions of people. Grain mold of sorghum, caused by a consortium of fungal pathogens, is a leading constraint to sorghum production. A second sorghum disease with significant economic impact is anthracnose caused by the ascomycete fungus *Colletotrichum sublineolum (Cs).* Grain mold causes yield reduction and is highly detrimental to food quality due to contamination by toxigenic fungi and mycotoxins while anthracnose results in significant yield reduction in susceptible cultivars. Genetic resistance is considered the only effective and sustainable way to control both diseases, but the genetic control of these diseases are not well understood. In this project, we implemented genetic, genomic and molecular approaches to identify loci and/or genes underlying resistance to the two diseases. The results presented in Chapters 2 to 5 provide new insights to the genetic and genomic architecture of resistance to grain mold and anthracnose. Chapter 1 provides background information and review of the literature on the pathology of the two diseases, the contrasting and shared mechanisms of genetic resistance and approaches to QTL and gene identification. Chapter 2 and Chapter 3 describe genome wide association studies (GWAS) conducted on sorghum landrace accessions from Ethiopia. Results of both sets of GWAS were recently published (Nida et al., 2019, Journal of Cereal Science 85, 295-304; Nida et al., 2021, Theoretical Applied Genetics, https://doi.org/10.1007/s00122-020-03762-2). Chapter 4 describes global transcriptome profiles of early stage of the developing grain from resistant and susceptible sorghum genotypes which uncovered process that correlate with resistance or susceptibility to grain mold. Finally, Chapter 5 summarizes two anthracnose resistance genes identified through whole genome resequencing and genetic mapping.

In Chapter 2, genomic regions associated with grain mold resistance were identified through GWAS conducted using sorghum landraces. A major grain mold resistance locus containing tightly linked and sequence related MYB transcription factor genes were identified based on association between SNPs and grain mold resistance scores of 1425 accessions. The locus contains *YELLOW SEED1* (*Y1,* Sobic.001G398100), a likely non-functional pseudo gene (*Y2*, Sobic.001G398200), and *YELLOW SEED3* (*Y3*, Sobic.001G397900). SNPs and other sequence polymorphisms that alter the *Y1* and *Y3* genes correlated with susceptibility to grain mold and provided a strong genetic evidence. Although *Y1* has long been known as a regulator of kernel

color and the biosynthesis of 3-deoxyanthocynidin phytoalexins, it was not annotated in the sorghum genome. The data suggest that the MYB genes and their grain and glume specific expressions determine responses to molding fungi.

Chapter 3 focuses on GWAS conducted on a subset of early to medium flowering accessions to identify grain mold resistance loci. In addition, because of the caveats associated with grain flavonoid mediated mold resistance, we specifically aimed to identify resistance loci independent of grain flavonoids. A multi-environment grain mold phenotypic data and 173,666 SNPs were used to conduct GWAS using 635 accessions and a subset of non-pigmented accessions, potentially producing no tannins and/or phenols. A novel sorghum *KAFIRIN* gene encoding a seed storage protein, and *LATE EMBRYOGENESIS ABUNDANT* 3 (*LEA3*) gene encoding a protein with differential accumulation in seeds were identified. The *KAFIRIN* and *LEA3* loci were also grain mold resistance factors in accessions with non-pigmented grains. Moreover, the known SNP (S4_62316425) in *TAN1* gene, a regulator of tannin accumulation in sorghum grain was significantly associated with grain mold resistance. These data suggest the critical role of loci harboring seed protein genes for resistance to sorghum grain mold.

In Chapter 4, global transcriptome profiles of developing grain of resistant and susceptible sorghum genotypes were studied. The developing kernels of grain mold resistant RTx2911 and susceptible RTx430 sorghum genotypes were inoculated with a mixture of fungal pathogens mimicking the species complexity of the diseases under natural infestation. Global transcriptome changes corresponding to multiple molecular and cellular processes, and biological functions including defense, secondary metabolism, and flavonoid biosynthesis were observed with differential regulation in the two genotypes. Genes encoding pattern recognition receptors (PRRs), regulators of growth and defense homeostasis, antimicrobial peptides, pathogenesis-related proteins, zein seed storage proteins, and phytoalexins showed increased expression correlating with resistance. The data suggest a pathogen inducible defense system in the developing grain of sorghum that involves the chitin PRR, MAPKs, key transcription factors, downstream components regulating immune gene expression and accumulation of defense molecules.

Finally, Chapter 5 deals with anthracnose resistance loci and subsequent genetic mapping and identification of two resistance genes. The sorghum line SAP135 was previously described for its broad-spectrum resistance to anthracnose. To identify the specific resistance gene, a mapping population was generated by crossing SAP135 with the susceptible line TAM428. Bulked-

segregant analysis (BSA) combined with whole genome re-sequencing of resistant and susceptible pools (BSA-seq) of the mapping population defined a single major peak on chromosome 8 for resistance to the Cs strain *Csgrg* which was designated as *ANTHRACNOSE RESISTANCE GENE 4* (*ARG4*). *ARG4* was co-localized with a locus identified in a parallel but an independent mapping study conducted using the sorghum line P9830 against another Cs strain *Csgl1*. Fine mapping revealed that the resistance loci from the two populations delineated two tightly linked loci, the latter locus designated as *ANTHRACNOSE RESISTANCE GENE 5* (*ARG5*). *ARG4* (*Sobic.008G166400*) and *ARG5* (*Sobic.008G177900*) encode canonical NBS-LRR proteins widely known intracellular immune receptors. Interestingly, SAP135 carries a functional *ARG4* but lacks *ARG5* whereas P9830 harbors a functional *ARG5* and lacks *ARG4* and both show sequence homology to wheat rust resistance genes. *Csgrg* and *Csgl1* are both virulent on sorghum lines TAM428 and BTx623, thus both lines carry susceptible alleles of *ARG4* and *ARG5*. Supplemental information for the unpublished chapters are presented in the appendix section of this thesis.

CHAPTER 1. LITRATURE REVIEW

1.1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench], native to Africa is among the major cereals grown for food, feed, biofuels and alcoholic beverages. It is drought tolerant and has a unique potential to thrive in marginal areas and nutrient deficient soils compared to most agronomic crops. Globally, about 59.3 million tons of sorghum grain was produced from an area of 42.1 million hectares during 2018 [1]. Sorghum is a multi-purpose crop in developing countries where its grain is used as a staple human food while its stalk and leaves are used as livestock feed and construction materials. Sorghum Stover after grain harvest serves as a major livestock feed during dry periods in regions of developing countries where it is predominantly grown as a dryland crop. Besides being gluten free alternative to wheat, sorghum grain has health benefits because of its antioxidant property and potential in reducing obesity [2].

The major biotic and abiotic stresses that limit the productivity of sorghum include drought, fungal pathogens, the parasitic weed Striga, and a number of insect pests and birds. Although sorghum is inherently drought tolerant, the erratic and limited rainfall in most sorghum growing ecologies affect sorghum's development and yield potential, which under extreme condition causes a complete yield loss. Fungal diseases are most destructive diseases in sorghum, which result in significant losses to quantity and quality of grain [3]. Grain mold, anthracnose, stalk rot, head smut, ergot, downy mildew are among the major fungal diseases of sorghum [4-7]. The parasitic weed Striga, also called witchweed is a major constraint to sorghum production in semiarid tropical Africa and Asia [8-10]. Over 150 insect species are known to infest sorghum [11] of which sorghum midge, shoot fly, sugarcane aphid, fall army worm, stem borer are among the most destructive to the crop [12-17]. Moreover, the red-billed quelea, which often forms colonies of large number, is a threat to sorghum production in Africa. A recent study on prevalence of quelea and tannin sorghum indicates parallel distribution of the two suggesting role of tannins in deterring quelea [18].

Sorghum is a resilient crop that has evolved diverse resistance mechanisms against major abiotic and biotic stresses. This chapter provides a review of advances towards understanding

sorghum's immune response to fungal pathogens that cause major threat to its production. It provides background and rationale for undertaking the current research aimed at identification of genes associated with resistance to the two major fungal diseases of the crop, grain mold and anthracnose. It highlights pathology of the two diseases, contrasting and shared mechanisms of host resistance against the diseases and methods of QTL (quantitative trait loci) and gene identification.

1.2 Terminology and definitions

Although the term "grain mold" is being exclusively used in most recent publications to describe fungal deterioration of sorghum grain, other terms including grain moulds, head mould, seed moulds, grain weathering, grain deterioration have been used in older literatures as reviewed by [19, 20]. Grain mold may be defined differently by various authors based on grain maturity stage and development of the disease. Two concepts have been described in defining grain mold of sorghum [20]. The first concept describes fungal infection and colonization of grains between anthesis and harvest, in which case, grain mold is broadly defined as a fungal component of pre-harvest grain deterioration due to either parasitic and/or saprophytic interaction with the plant. The second concept limits grain mold to infection and colonization of spikelet tissue prior to grain maturity and this definition does not include fungal damage after physiological maturity. The post-physiological maturity or postharvest grain deterioration constitute component of weathering based on the second concept. The definition based on the first concept is used in this thesis as it explains the damage due to grain mold better than the second concept from practical perspective. Therefore, in this review the more generalized definition of grain mold, which includes all fungal associations with sorghum spikelet tissue that occurs between anthesis and harvest is used.

Although it is a widely known fungal disease of a range of plants spices [21-23], it appears that a definition for the term anthracnose is not available in common literature. However, based on glossary of plant pathology by the American Phytopahological society (APS), anthracnose is referred as a disease caused by acervuli-forming fungi (archaic order Melanconiales) and characterized by sunken lesions and necrosis [24]. Anthracnose of sorghum is recognized in three forms, which are foliar anthracnose, anthracnose stalk rot and panicle and grain anthracnose [4].

1.3 Symptoms

Symptoms associated with grain mold vary with the species of fungi causing the disease, time and level of infection [19]. Early symptoms appear as discoloration of spikelet tissues such as lemma, palea, glumes, and lodicules [20, 25, 26]. Depending on the type of fungi involved, discoloration of grain is a common symptom but lightly infected grains may appear completely normal [19, 20, 25]. Symptoms of anthracnose in sorghum includes chlorotic flecks, acervuli formation and necrotic lesions on leaves and death of leaves [27, 28].

1.4 Causal fungi and their infection strategy

More than 40 genera of pathogenic and opportunistic fungi are associated with molded grains [26]. Some of the major grain mold causing fungal genera includes *Fusarium*, *Curvularia*, *Alternaria*, *Phoma*, *Bipolaris*, *Exserohilum*, *Aspergillus*, and *Penicillium.* Among these, several *Fusarium spp*, *Curvularia lunata, Alternaria alternata, Phoma sorghina, Bipolaris australiensis* and *Exserohilum turcicum* are widely known grain mold causing species [25]. Based on earlier reviews on studies across India, Africa and USA, the predominant species causing the disease are *Fusarium* spp. and *Curvularia* spp. of which *F. moniliforme* and *Curvularia lunata* appears to have worldwide significance [19, 26]. The *Fusarium* species *F. andiyazi, F. proliferatum, F. sacchari, F. verticillioides, F. thapsinum, F. nygamai, F. pseudonygamai* which were formerly included in *F. moniliforme* are among the major mold causing pathogens [25]. Based on a survey on *Fusarium* species and moniliformin occurrence in Argentina, *F. verticillioides*, *F. thapsinum* and *F. andiyazi* were the most frequently recovered from sorghum grains followed by *F. proliferatum* and *F. subglutinans* [29]. *F. verticillioides* is also a major pathogen of maize that causes ear and stalk rot diseases, and produces the mycotoxin fumonisin [30-33]. The related fungal pathogen *F. graminearum* causes *Fusarium* head blight of wheat and barley [34].

ColIetotrichum graminicola has long been considered to be the causal fungus for anthracnose on cereals including maize, sorghum and grasses [35]. However, morphological and genetic analysis of isolates from maize and sorghum revealed that the isolates infecting the two crops belong to distinct species [36-38]. Therefore, isolates from maize are now identified as *C. graminicola* while those from sorghum are considered *C. sublineolum* [38]. Currently, *C. sublineolum* is used in almost all recent studies as the causal agent of sorghum anthracnose [27,

28, 39-43]. *C. sublineolum* has a hemi-biotrophic infection based on its mode of nutrition [44]. *C. graminicola* has three recognizable infection phases that include formation of melanized appressoria on host surface before penetration, intracellular colonization of living host cells which corresponds to the biotrophic phase and a final necrotrophic phase with extensive host cell death and appearance of symptoms [45]. The particular infection strategy is characterized by successive expression of different sets of fungal genes associated to pathogenic transitions involving induction of effectors and secondary metabolites before penetration and during biotrophy and upregulation of hydrolases and transporters during the switch to necrotrophy [46]. The same infection stages are likely to be observed in *C. sublineolum*.

1.5 Importance

Grain mold is a complex disease caused simultaneous infection by fungal pathogens with the impacts on yield and grain quality. Yield reductions are due to caryopsis abortion, reduced seed filling and lower grain density while seed quality and market values are affected due to surface discoloration, embryo and endosperm deterioration and contamination by mycotoxigenic fungi and their mycotoxins [47, 48]. The disease remains to be a major problem to the crop with losses reaching up to 100% in highly susceptible cultivars and annual global losses estimated at $130 million [49]. Besides yield losses, contamination of grain by toxigenic fungi and their mycotoxin could be most important even if the levels may be higher or lower than the standards depending on cultivar and growing conditions. In a recent analysis of multimycotoxin levels of grain samples collected from farmers' stores in South, East and Northwest Ethiopia [50], all tested samples were contaminated by *Fusarium* and *Aspergillus* spp though the prevalence of the major mycotoxins was lower than 15% except zearalenone. Zearalenone occurred in one third of the samples at average level of 44 mg/kg. This study also reported that the concentration of aflatoxins B1 and G1 were higher than the European standards. Since the samples were collected from storage, there is a possibility that the contaminations may be aggravated due to the farmers' storage condition. Zearalenone was considered as a major mycotoxin danger throughout tropical areas of sorghum production since the 1970s [19]. A recent study on mycotoxin contamination of sorghum grain from four sub-Saharan African (SSA) countries (Burkina Faso, Ethiopia, Mali and Sudan) indicated that 33% of the analyzed grain samples were contaminated by at least one of the

mycotoxins: aflatoxins, fumonisins, sterigmatocystin, Alternaria toxins, ochratoxin A (OTA) and zearalenone [51]. This study also indicated that mycotoxins from *Aspergillus* spp. and *Alternaria* spp. could be concerns in SSA grain sorghum with possible health risks. Anthracnose can cause yield losses of over 50% in susceptible lines [52].

1.6 Host resistance to major grain and leaf diseases in sorghum

1.6.1 Concepts of disease resistance in plants

Resistance to pathogens in plants are broadly grouped into non-host or host resistance. Non-host resistance refers a general resistance against majority of pathogens by all plant species and includes physical barriers and antimicrobial compounds [53, 54]. It is a form of resistance to non-adapted pathogens [55-58]. It is the most effective and durable form of resistance [59]. Conversely, host resistance is expressed by certain genotypes of a susceptible host [54].

Upon infection by certain pathogens, plants respond through two main immune pathways [60]. In the first pathway, plants recognize and respond to conserved pathogen or microbial associated molecular patterns (PAMP), resulting a PAMP-triggered immunity (PTI) while in the second branch, plants respond to pathogen virulence effectors, that result in effecter-triggered immunity (ETI). PTI is also termed basal resistance [61-63]. ETI is an accelerated and intensified PTI response which usually involves a hypersensitive cell death [60].

Physical or chemical defense strategies that contribute to disease resistance in plants include preformed barriers such as preformed components of the plant cell wall, antimicrobial enzymes,secondary metabolites, and inducible structural barriers [59]. Saponins which are glycosylated plant secondary metabolites are examples of preformed chemical barriers as demonstrated using saponin-deficient (sad) mutants of oat species [64]. An example of inducible barrier is that of non-race specific mlo resistance to powdery mildew in barley [65]. General elicitors released by pathogens during infection activate production of antimicrobial compounds such as phytoalexins which are inducible antimicrobial secondary metabolites [66-68].

1.6.2 Resistance to grain mold in sorghum

Physical and chemical kernel properties are widely known components of sorghum's resistance against grain mold. Testa pigmentation, concentration of phenolic compounds, pericarp color and kernel hardness were identified as major factors associated with resistance to grain mold in the crop [69-72]. Higher tannin level in the testa layers of pigmented testa was considered to be the most important trait conferring grain mold resistance [69]. Cultivars with harder grain, higher levels of phenols and colored kernel have better resistance to grain mold [70, 72]. However, the correlations between grain mold resistance and seed color, seed flavan-4-ol content, glume phenol and flavan-4-ol contents, and glume cover were weaker and inconsistent [72]. The other group of phenolic compounds and secondary metabolites other than tannins that are associated with grain mold resistance include the anthocyanins (3-deoxyanthocyanidins) and flavan-4-ols [70, 71]. These suggest a positive role for the major 3-deoxyanthocyanidins (apigeninidin and luteolinidin) and flavan-4-ols in grain mold resistance although the results were not conclusive.

Studies on genetics of grain mold resistance in sorghum has almost the same age as studies on grain phenols and physical properties. The earlier genetic studies include those where the authors studied the association of genes controlling caryopsis traits with grain mold resistance [69]. This study indicated that testa pigmentation (B_1-B_2-), a red pericarp (R-Y-) which are among the major traits conferring grain mold resistance are dominantly inherited. Four closely located QTLs were detected on linkage group F for grain quality traits and mold during germination [73]. Latter, the inheritance of grain mold resistance was determined in non-pigmented testa sorghum using F1, F2 and backcross generations of the cross between the resistant Sureno and susceptible RTx430 [74]. The generation means analysis mainly detected both additive and dominance effects almost in all environments with broad and narrow sense heritability values ranging from 0.46 to 0.82 and 0.39 to 0.59, respectively, which is fairly high to enable improvement through breeding. Moreover, at least 4 to 10 genes were estimated by the study that contribute to grain mold resistance.

Recent QTL mapping studies identified multiple small to moderate effect QTLs conferring grain mold resistance. Five QTLs were detected for grain mold resistance in a study on 125 F5 recombinant inbred lines derived from a cross between Sureno and RTx430 [49]. The QTLs accounted for 10 to 23% of the phenotypic variance and the detection was dependent on

environment. The QTLs were found on chromosome 4 to 7 and 9. Those on chromosome 4 and 5 were stronger than the rest and the two accounted for 29-32% of the phenotypic variation [49]. Moreover, the two linkage groups contain closely located QTLs for plant stature traits (height, peduncle length) and grain mold which may be due to pleiotropic effects of a single locus and/or closely located loci [49]. In a different association mapping study on the sorghum mini core accession, two loci (on chromosome 2 and 8) were detected that are linked to grain mold resistance [75]. More recently, availability of dense SNP markers through genotyping by sequencing (GBS) approach enabled implementation of a number of genome wide association studies (GWAS). GWAS in the US sorghum association panel (SAP) identified resistant accessions and two genetic loci associated with low seed deterioration and one locus implicated in emergence rate [76]. A separate GWAS for grain mold resistance in SAP involving individual or combined inoculation with selected grain mold causing species resulted in significant SNPs associated with resistance to the disease [77].

1.6.3 Resistance to anthracnose in sorghum

The first set of 11 lines that are resistant to anthracnose were identified from the Texas Agricultural Experiment Station and USDA-ARS sorghum conversion program [78], [79] which laid a foundation for study of inheritance of genetic resistance to the disease. Inheritance study on these lines indicated that resistance was dominant in some lines and recessive in others while one resistant line (SC-748-5) showed resistance in all test environments [80]. A 3:1 ratio of resistant to susceptible phenotypes in F2 populations of a cross between SC-748-5 and the susceptible BTx623 suggested a single dominant gene conferring resistance in SC-748-5 [80]. The resistance gene carried by SC-748-5, named as *Cg1*, was mapped to the distal end of chromosome 5, co-segregating with an AFLP marker, *Xtxa6227* and SSR marker, *Xtxp549* [81]. Latter, through sequencing of SC-748-5 and mapping of the anthracnose resistance locus carried by this line using recombinant lines of a cross between BTx623 and SC-748-5, a major QTL was consistently detected on chromosome 5, but the result suggested that the resistance in SC748-5 may not be under the control of single gene [82].

In order to define pathotypes of *C. sublineolum*, a set of 18 sorghum lines were established as differentials [83]. These include sorghum lines SC283, Brandes, SC112-14, Theis, BTx623,

SC748-5, BTx378, Tx2536, SC326-6, BTx398, QL3, TAM428, SC414-12E, SC328C, PI570841, PI570726, PI569979, and IS18760. Three of the lines (SC748-5, SC326-6 and SC414-12E) are among the 11 originally identified as resistant to the anthracnose [79]. These germplasm served as source of resistance for breeding and mapping resistance genes in recent studies. The resistance gene carried by SC112-14 was again mapped to the distal end of chromosome 5 but appears to be independent of *Cg1* [84]. Additional major QTLs on chromosome 5 and chromosome 9 were mapped using the sorghum lines SC414-12E and SC155-14E, respectively [85].

More recently, genome wide association mapping (GWAS) and diversity studies that involve multi-parent populations re-detected the QTLs identified previously by bi-parental approaches, and identified a few new loci and additional sources of resistance for anthracnose [28, 86, 87]. The four consistently detected regions, three closely located on distal end of chromosome 5, and one on chromosome 9 span large genomic regions with candidate genes related to multiple defense response mechanisms [28, 87]. Based on GWAS, the three loci at the distal end of chromosome 5 are located at about 65.2, 66.5 and 71.6 Mbp [87] while that on chromosome 9 is located at about 1.2 Mbp [28].

The sorghum 3-deoxyanthocyanidin phytoalexins accumulate at the site of *C. sublineolum* infection and resistance through this mechanism requires a functional *Yellow seed1 (Y1)* [88]. Maize, with an orthologues gene, *Pericarp color1 (P1)* and flavonoid structural genes doesn't produce the 3-deoxyanthocyanidin phytoalexins in response to infection, but transgenic maize lines expressing the sorghum *Y1* gene were able to synthesize the 3-deoxyanthocyanidin phytoalexins in response to *C.graminicola* infection and showed resistance response [89]. The sorghum 3-deoxyanthocyanidins that include apigeninidin and luteolinidin accumulate at the sites of infection and prevent proliferation of the fungus [90, 91]. Luteolinidin is more fungitoxic than apigeninidin [91]. This observation indicates that the 3-deoxyanthocyanidin phytoalexin dependent resistance is a shared resistance mechanism against both grain mold and anthracnose.

Transgenic sorghum lines expressing the wheat *Lr34* multipathogen resistance gene not only exhibited resistance against anthracnose and rust, but also showed increased level of 3-deoxyanthocyanidin metabolites and elevated expression of flavonoid phytoalexin biosynthesis

genes associated with *Lr34* expression [40]. This may be an evidence for an association between R gene mediated resistances with flavonoid biosynthesis pathways.

Mapping studies aimed at identification of genomic regions associated with resistance to the grain mold and anthracnose in sorghum followed the traditional QTL mapping approaches. Recombinant inbred lines or segregating generations resulting from crossing of two lines contrasting for the two diseases and simple sequence repeat (SSR) and Amplification fragment length polymorphism (AFLP) markers were used to construct linkage maps and identify QTLs associated with resistance to the diseases [49, 81].

As in many other plant species, advances in sequencing technologies provided new opportunities to identify genomic regions associated with the two diseases at a higher resolution due to the use of larger number of single nucleotide polymorphism (SNP) markers [28, 39, 76, 77, 86]. The advances in sequencing technologies also brought new capabilities to conduct QTL mapping at a multi-parent level that has particularly been useful to identify resistance loci carried by multiple parental lines. Among the first set of mapping studies that utilized large SNP markers and association panels of sorghum include application of 14,739 SNPs on sorghum mini core of 242 accessions that was used to conduct association mapping to identify loci linked to rust and grain mold resistance [75] and a small (n = 142) or large (n = 336) association panels and a bi-parental inbred lines (n = 263) genotyped at 265,487 SNPs used to identify loss of function phenotypes in sorghum flavonoid pigmentation traits [92]. Since QTL analysis could be time-consuming and labor-intensive due to the requirement of marker development and selection, BSA-seq (also termed QTL-seq) was proposed as a rapid approach for QTL identification that involves whole-genome resequencing of DNAs from two populations each composed of a pool of individuals showing extreme phenotypes [93]. We adapted the term BSA-seq as this can be applied to qualitative traits. Effectiveness of the BSA-seq approach was validated by a few studies that identified loci associated with important agronomic traits in barley, chickpea and millet [94-97] is also getting popularity in recent years [98-103].

Population structure and kinship are the two major factors that can affect results of genome wide association studies and need to be accounted for [104]. Population structure may be referred to as a large-scale systematic variation in ancestry or groups of individuals with a shared ancestry

than that expected in a random-mating population, of this shared ancestry corresponds to kinship while patterns of kinship among groups of individuals refers population structure [105]. False positives that arise due to population structure and kinship is a major problem in GWAS and a number of statistical models developed to control such detections [106, 107]. The widely used models that correct for both or either of population structure and kinship includes GLM, MLM [108], CMLM [109], FarmCPU [110], and BLINK [111] each with both advantages and limitations. Population structure and kinship are often represented by Q and K matrixes, respectively [110]. General Linear Model (GLM) reduces false positive in GWAS by fitting population structure as covariates [112], that means GLM only accounts for population structure. Mixed Linear Model (MLM) accounts for both population structure and kinship [108] where MLM with Q + K model was proposed as a better model than the Q or K models alone. Although MLM has been one of the most widely used models, it was found computing intensive and disadvantageous in that real associations may barely detected by MLM when traits are associated with population structure [110]. The limitations associated with MLM were then resolved to some extent by the next generations of models (CMLM, FarmCPU and BLINK) but none of these new models may be without limitations. Improving statistical power and reducing computing time have been the objectives of the evolutions of new models while addressing the issues of confounding due to population structure and kinship [110]. The latest model termed "Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway" (BLINK) proposed as the next efficient model that improves statistical power than its predecessor (FarmCPU) and reduce computational time [111]. Moreover, FarmCPU uses bin method with assumption that quantitative trait nucleotides (QTNs) are evenly distributed across the genome, which is one of the requirement for the model [110]. In BLINK, this requirement is eliminated by replacing the bin method with linkage disequilibrium information [111].

1.7 Rationale and objectives

Host responses to pathogenic fungal pathogens that cause major diseases in sorghum, grain mold and anthracnose are poorly understood. Despite the identification of few resistance loci by previous studies, resistance genes and their underlying mechanisms associated with such loci are not known. Sorghum grain, being distinct from the commonly studied leaf tissue and containing largely the storage compounds such as starch and protein poses unique challenges to genetic and

molecular studies. Moreover, complexity of the fungal pathogens involved in causing grain mold, effect of environment, limitations in phenotyping and nature of resistance to the disease are some of the reasons for limited success in identification of resistance loci. Some of the known resistance mechanisms against grain mold such as tannins and phenols in grain have nutritional drawbacks that hinders combining resistance with grain quality traits although some phenols have health benefits [2]. Complexity of anthracnose resistance loci which often span large genomic regions and harboring cluster of resistance genes with large unexplained phenotyping variation present different set of challenges. Genetic resistance to anthracnose appears to be conferred primarily through major R genes and this is known to be mostly race specific. However, there was no specific R gene identified with either host specific or broad-spectrum resistance to the disease to date. This is despite more than 340 of NBs-LRR class of R genes present in the sorghum genome [113]. In addition, host response to both diseases is partly known to involve an active pathogen inducible system leading to accumulation of key antimicrobial compounds. However, the regulation of such inducible defense systems in both grain and leaf tissues is not well known.

The overall goal of this research is to get insight of the genetic, genomic and molecular architecture of sorghum's immune response to major fungal pathogens. We conducted a serious of genome wide association studies, global transcriptome analysis and bulk segregant analysis using diverse natural populations and unique sorghum variant with broad spectrum resistance to the diseases.

The specific objectives are:

- Identify genomic regions and candidate genes associated with grain mold resistance in large collection of natural populations
- Determine global changes in gene expression, molecular and cellular functions, and metabolic pathways that are reprogrammed early during fungal infection of the developing grain
- Identify loci and candidate genes associated with resistance to anthracnose in a sorghum line SAP135 which has a broad spectrum resistance to anthracnose and rust
- Fine map and identify candidate genes in a recently identified anthracnose resistance loci mapped using a sorghum line P9830

CHAPTER 2. IDENTIFICATION OF SORGHUM GRAIN MOLD RESISTANE LOCI THROUGH GENOME WIDE ASSOCIATION MAPPING

A version of this chapter was previously published by *Journal of Cereal Science* [114], https://doi.org/10.1016/j.jcs.2018.12.016

2.1 Abstract

Grain mold, caused by a consortium of pathogenic fungal species, is the most important disease of sorghum [*Sorghum bicolor* (L.) Moench]. Genome wide association study on 1425 diverse Ethiopian sorghum landraces identified a major grain mold resistance locus containing tightly linked and sequence related MYB transcription factor genes. The locus contains *YELLOW SEED1* (*Y1*), a likely non-functional pseudo gene (*Y2*), and *YELLOW SEED3* (*Y3*). SNPs and other sequence polymorphisms that alter the *Y1* and *Y3* genes correlated with susceptibility to grain mold and provided a strong genetic evidence. Accordingly, the expression of both *Y1* and *Y3* genes in the developing grain and glumes of a widely known susceptible sorghum line, RTx430, were severely reduced but significantly increased in the resistant line, RTx2911. In addition, the expression of flavonoid biosynthesis genes such as *DIHYDROFLAVONOL 4-REDUCTASE 3* (*DFR3*) was significantly induced in the resistant line in response to inoculation by a mixture of spores from different molding fungi while the susceptible line displayed reduced expression. The data suggest that the MYB genes and their grain and glume specific expressions may determine the differential regulation of the flavonoid biosynthesis pathway genes, the synthesis of 3-deoxyanthocynidins and ultimately responses to molding fungi. The study also indicated that resistance to grain mold may be negatively associated with grain functionality traits such as '*injera*' making quality of sorghum.

2.2 Introduction

Sorghum, *Sorghum bicolor* (L.) Moench, is among the world's most important cereal crops used for food, feed, and bio-fuels. The crop is known for its adaptation to arid and semi-arid agro-ecologies where other crops do not thrive well. Grain mold remains one of the most widespread and major diseases affecting the sorghum grain and its quality especially in regions with high humidity and temperature during grain development. Broadly, grain mold is defined as a fungal component of pre-harvest grain deterioration due to either parasitic and/or saprophytic colonization of the plant (Forbes et al., 1992). An alternative definition limits grain mold to infection and colonization of the spikelet tissue prior to grain maturity excluding fungal damage after physiological maturity. All fungal associations with sorghum spikelet tissue that occur between anthesis and harvest affect the quality and quantity of the grain, thus a laxer definition may better explain the damage from practical point. Regardless, grain mold is a result of infection by multiple fungal pathogens which reduces yield and grain quality. Yield reductions are due to caryopsis abortion, reduced seed filling and lower grain density while seed quality and market values are affected due to surface discoloration, embryo and endosperm deterioration and contamination by toxigenic fungi and their mycotoxins [47, 48]. Mold affected grains have significantly reduced processing qualities and cannot be used for food [115]. Grain mold is highly detrimental to the quality of food produced from sorghum, and thus resistance is a major consideration.

The pathology of grain mold, the virulence strategies of the fungal species during their interaction with the sorghum tissue has not been well studied. Grain mold infection in sorghum usually start at or after flowering depending on environmental conditions but the symptoms may not be visible until grain maturity [19]. The disease is caused by a complex of fungal species in the genera *Fusarium*, *Curvularia*, *Alternaria*, *Phoma*, *Bipolaris*, *Exserohilum*, *Aspergillus*, *Colletotrichum*, and *Penicillium.* Multiple *Fusarium spp*, *Curvularia lunata, Alternaria alternata, Phoma sorghina, Bipolaris australiensis* and *Exserohilum turcicum* are known to cause grain mold (Thakur et al., 2006). *Gibberella zeae* was identified as the dominant species causing grain mold in sorghum followed by *F. verticillioides*. *F. graminearum* is the causal agent of fusarium head blight (FHB) in wheat which is a major constraint in the production of small grains [116, 117]. Plant diseases that affect the reproductive tissues may provide a paradigm for studies in grain mold resistance. However, grain mold in sorghum is more complex involving many fungal species. Corn

ear molds are also caused by different fungal species. In all these cases, the diseases are caused by non-obligate facultative fungi with necrotrophic mode of nutrition that have the ability to colonize, degrade and utilize dead or dying plant materials such as senescing tissues. Finding genetic resistance to diseases caused by these host unspecific fungi have been challenging. For FHB, although many resistance QTLs have been identified, the specific genes and their functions are largely unknown. Efforts to breed genetic resistance to *F. graminearum* and *F. verticillioides* and other molding fungi have been limited. Knowledge on resistance genes underlying the QTL and mechanisms is important for breeding purposes.

Research on sorghum grain mold focused on identification of physical and chemical kernel properties that are associated with resistance to the disease. Testa pigmentation, concentration of phenolic compounds, pericarp color and kernel hardness were identified as major factors associated with resistance [69-72]. Some of the previous QTL mapping studies identified small to moderate effect QTLs conferring resistance [49]. However, much of the identified QTLs were associated with traits that modulate resistance indirectly and major effect QTLs or genes have not been identified. The complex nature of the fungal species involved, limitations in phenotyping methods, quantitative nature of genetic resistance, and significant effect of environmental factors on the disease are among the major reasons for limited success in identification of major-effect grain mold resistant loci.

Here, we describe the identification of sorghum grain mold resistance locus enabled by mold resistance scores and genotyping by sequencing (GBS) of diverse Ethiopian sorghum landrace collection. Significant genetic variation was observed that likely represent different mechanisms of grain mold resistance in the population. Importantly, genome wide association analyses identified SNPs with significant association with grain mold resistance defining a narrow genomic region carrying two genes encoding putative R2R3 MYB transcription factors. Furthermore, we show that the two R2R3 MYB transcription factors, *YELLOW SEED1* and a sequence related R2R3 MYB, designated *YELLOW SEED3*, are two candidate genes that correlated with resistance. Sorghum natural variants with sequence polymorphisms at these two candidate genes were studied for grain mold resistance and expression of flavonoid pathway genes, supporting the role of *YELLOW SEED1* and *YELLOW SEED3* in grain mold resistance. *YELLOW SEED1* has been previously implicated in resistance to the fungal pathogen *Colletotrichum sublineolum* and regulates the biosynthesis of the 3-deoxyanthocyaidin phytoalexins. By using a

large population of sorghum natural variants from the center of origin and diversity, we delineate two regulatory genes and allelic variants for this complex trait. In sum, our genetic data implicates two related R2R3 MYB proteins, as important regulators of resistance to a consortium of fungal pathogens.

2.3 Materials and Methods

2.3.1 Phenotyping

The current study is part of an effort to describe a large-scale Ethiopian sorghum germplasm collection, next generation sequencing, and association mapping for various traits under USAID funded Sorghum and Millet Innovation Lab (SMIL) project. The germplasm collections were sampled from over 9000 sorghum accessions maintained at the Ethiopian Biodiversity Institute (EBI) and the national and regional agricultural research centers in Ethiopia. The total number of germplasms studied for grain mold and other traits were 2010. Among these, 1940 were Ethiopian Landrace accessions assembled from the various institutions in the country mainly from EBI (about 1550) and the remaining were different breeding lines and released varieties. Due to missing genotype data, only 1425 of the landraces were used for the association analysis described in the current study. Field evaluation for grain mold resistance was conducted during 2015 and 2016 seasons at Bako Agricultural Research Center, Ethiopia. Each accession was planted in single row plot of 3m length with 20 cm between plants. Genotypes were randomized in each test year. Five plants representing each accession were tagged and scored for mold resistance using a rating scale of 1-5, where 1 represents highly resistant, and 5 is complete seed deterioration indicating extreme susceptibility. To study the relationship between kernel color and grain mold resistance, additional data from a separate phenotyping study conducted in 2017 on a subset of 700 accessions at two locations, Bako and Jimma were used. Bako and Jimma are suitable for grain mold study because of high rainfall and relative humidity.

2.3.2 DNA extraction, genotyping, SNP calling and quality control

DNA was extracted from a lyophilized young tissue using acetyl trimethylammonium bromide (CTAB) protocol modified for 96-well plates (Mace et al., 2003). A total of eighteen 96-plex GBS libraries were constructed and genotyped at the University of Wisconsin, Biotechnology

center. Some of the accessions were excluded from genotyping due to either missing significant amount of phenotype in multiple years or missing tissue during DNA extraction. The genotyping by sequencing (GBS) procedure [118] was implemented using the *Ape*KI restriction enzyme (recognition site: G|CWCG). Single nucleotide polymorphism (SNP) calling was performed with the TASSEL GBS pipeline v5 [119] through aligning reads to the most recent version of the sorghum reference genome, *Sorghum bicolor* v3.1.1 [120]. SNPs were filtered by excluding genotypes with > 20% missing individuals, markers with > 40% missing and minor allele frequency (MAF) values > 0.05.

2.3.3 Genome wide association analysis

A total of 1425 landrace accessions, representing a subset of the collection, and have robust SNP markers that passed quality control were used for GWAS analysis. There were missing mold scores for 13 and 11 landraces during 2015 and 2016, respectively, which were different between the years. Therefore, the GWAS for grain mold included 1412 and 1414 landraces in 2015 and 2016, respectively. Population structure and kinship, known to result in spurious associations [108] were adequately accounted for in the GWAS analysis. ADMIXTURE analysis [121] was implemented for different K (number of sub-populations) to study existence of population structure. Using 10-fold cross validation (CV) for K=2 to K=20, the ADMIXTURE analysis indicated a steep decrease in CV error values until K=11, indicating presence of population structure and 11 as optimum number of sub-populations (clusters). The clustering pattern was also found to correspond well with a pair-wise distance based hierarchal clustering generated by calculating genetic distance based on the SNP markers identified. Because of the existence of population structure, Compressed Mixed Linear Model (CMLM) [109] implemented in GAPIT package [122] in R software [123] was used as the most appropriate model to conduct GWAS analysis. *Quantile-quantile* (Q-Q) plots of p-values were examined to determine how well the model accounted for population structure and familial relatedness. Significant associations were determined using a false discovery rate- adjusted p-value of < 0.05 as implemented in GAPIT. Manhattan and Q-Q plots were visualized using the R package qqman [124].

2.3.4 Visualization of grain mold resistance locus in re-sequenced sorghum lines

The polymorphism pattern of grain mold resistance locus in diverse sorghum lines was observed using sequence data of all the re-sequenced lines. A BAM file of the locus encompassing about 60 kb genomic region of all the re-sequenced lines was generated and visualized using Integrative Genomics Viewer (IGV 2.3.97) [125, 126].

2.3.5 Isolation, multiplication and inoculation of grain mold fungi

To study pathogen induced gene expression, a mixture of spore suspension from five *Fusarium* and one *Alternaria* species were spray inoculated on to panicles at 20 days after anthesis, which represent the soft dough stages of the developing grains. The *Fusarium* and *Alternaria* species were isolated from infected grain collected from Purdue University research field, West Lafayette, Indiana. Isolation and maintenance of the fungal culture, inoculum preparation and plant inoculation were conducted as described with minor modifications [25]. In addition to using grain from susceptible lines for inoculum production, oat meal agar and 70% potato dextrose agar (PDA) was used for isolation and maintaining fungal cultures. Fungal cultures produced sufficient spores of the various grain mold fungi within a week to about 10 days on 70 % PDA. Oat meal agar (100%) was used to grow fungal culture to harvest mycelium for DNA isolation. For PCR amplification of internal transcribed spacer of the fungal isolates, ITS1 (5′-TCCGTAGGTGAACCTTGCGG-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) primers were used. The specific species of *Fusarium* were determined by sequencing of the internal transcribe spacer using the ITS1 primer and alignment with nucleotide sequence databases [127] and they were found to be *F. proliferatum, F. graminearum, F. thapsinum*, *F. verticillioides and F. oxysporum*. The five major *Fusarium* species isolated from the sorghum grain were grown on plates and conidial suspension of equal proportions of the different species were used for inoculation.

Plants were grown in the greenhouse under optimum condition for sorghum and later transferred to a humidity chamber for inoculation. During each gene expression experiments 8-10 plants were grown for each genotype and 5 plants from each of the test genotypes that flowered at the same time were selected and inoculated at the same time. Tissue samples were taken from at least 3 plants for each genotype during each experiment. To confirm the gene expression results,

experiments were repeated for 3 to 4 times. Inoculation and disease establishment was done in a humidity chamber equipped with humidifier (Herrmidifier 707 Atomizing, Trion Indoor Air Quality) that generate mist. The humidifier has a humidistat adjustable to the required level and controls the amount of moisture added to the surrounding air. This technique enabled us to regulate the humidity to 85 to 90% inside the chamber which has been effective to infect plants. Plants prepared for inoculation were moved to the humidity chamber a day before inoculation and moved back to a greenhouse at 48 hours after inoculation and maintained in the greenhouse under overhead mister until maturity.

2.3.6 Validation of candidate genes using gene expression analysis and PCR based characterization

Total RNA was extracted as described with minor modifications [128]. The Acid-Phenol:Chloroform, pH 4.5 (5:1) from Ambion instead of the citrate buffer saturated phenol (pH 4.3):chloroform (1:1) was used. This has been effective to extract high quality RNA from developing sorghum grain (20 days after flowering) while RNA from glume can also be extracted using the regular TRIzol method (Invitrogen). Following DNase treatment (Promega), cDNA was synthesized from 2 μg total RNA using the AMV reverse transcriptase (NEB). Quantitative PCR analyses were performed on the CFX Connect real-time system (Bio-Rad) using a SYBR Green Supermix (Bio-Rad). Sorghum *Actin* gene was used as an endogenous control for normalization. A minimum of three technical and three biological replicates were used for the quantitative real-time PCR (qRT-PCR) analysis for each sample. Expression levels were calculated by the comparative cycle threshold (Ct) method. Primers used for qRT-PCR for the MYB genes are: YS1-qPCR-F1 (5'-AGACCGATCAGACCACAACC-3') and YS1-qPCR-R1 (5'-CACGTAGTCATGGCGAGCTA-3') for *YELLOW SEED1* (*Y1*); YS1-L-qPCR-F1 (5'-GGAGCAGATCGACCAGAGC-3') and YS1-L-qPCR-R1 (5'-GAGGGAAGCCGTTAACATAGC-3') for *YELLOW SEED3* (*Y3*). The primers used in qRT-PCR for flavonoid biosynthesis genes are taken from [129] and include *CHALCONE ISOMERASE* (*CHI*), forward (5'-AAGTTCAAGGAGGCGTTCAA-3') and reverse (5'-CGACTGGCTGGTTCTCTTTC-3'); *DIHYDROFLAVONOL 4-REDUCTASE3* (*DFR3*); forward (5'-CGGATGTGACGATTGTTTGA-3') and reverse (5'-GGGCATATTGGTTTGGAACTT-3'). Expression of genes in flavonoid biosynthesis pathways was studied at different grain developmental stages from flowering to the hard dough stages. The widely known RTx430 and

RTx2911 genotypes representing susceptible and resistant lines, respectively, were used to study expression of different genes. The two lines were challenged by a mixture of spore suspension of *Fusarium* and *Alternaria* species. Developing grain and glume tissues were sampled at 0, 24 and 48 hours after inoculation.

2.3.7 Analysis of association between grain mold resistance and grain functionality traits

Correlation and descriptive analysis were conducted to understand the relationship between grain mold resistance, kernel color and grain functionality traits. Additional sensory and tannin data available from a separate study on '*injera*' (a pan cake like traditional Ethiopian recipe) making property of Ethiopian landraces have been used [130]. Landraces that have kernel color, grain mold score, sensory score and tannin data were used for the association analysis. Correlation coefficients (r) (Pearson) and the corresponding R^2 values were calculated using Excel spreadsheet.

2.4 Results

2.4.1 Genetic variation for grain mold resistance in Ethiopian sorghums landraces

Genetic variation is a key pre-requisite to study genetic control of traits and underlying mechanisms. We have summarized the variation in grain mold resistance using frequency distribution of grain mold score taken during 2016 at Bako, Ethiopia (Figure 2.1). A large number of the sorghum landraces used in this study were found to be highly resistant while more than half were rated as resistant to moderately resistant and some were susceptible to grain mold. A similar trend was observed during 2015 season (data not shown). This indicates existence of good level of genetic variation for the trait in the collection and suitability for association mapping.

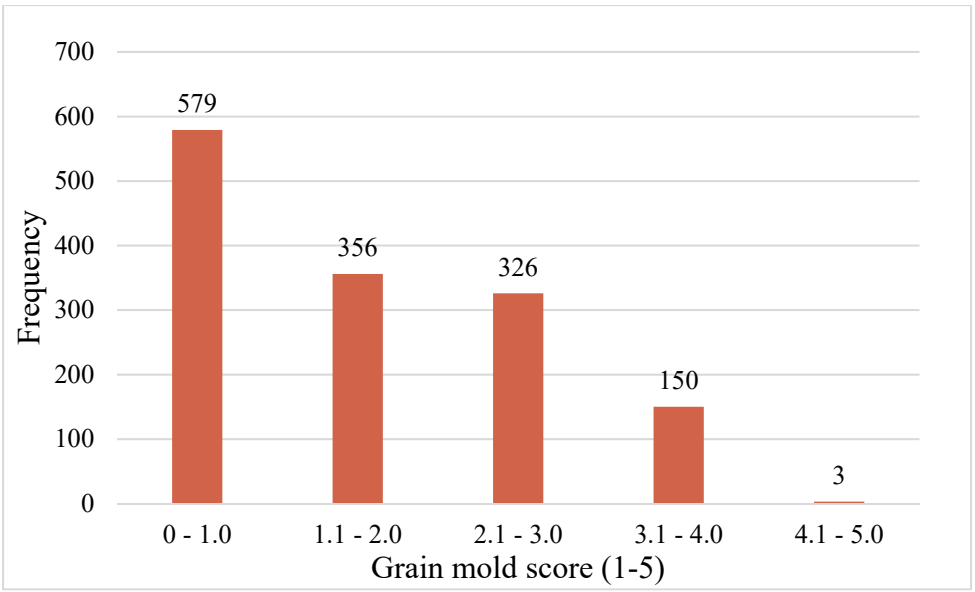


Figure 2.1. Frequency distributions of grain mold scores in Ethiopian landrace sorghum collections. Data are from field trial conducted in 2016 at Bako, Ethiopia. Scores for each landraces are average of five plants. Grain mold was scored on a 1-5 scale where 1 = no mold or highly resistant, 2 = resistant with minor mold infection, 3 = moderate infection, 4 = susceptible, 5 = highly susceptible).

2.4.2 Genome wide association analysis for grain mold resistance

The quality control analysis on initially discovered SNPs of about 879,407 produced 72,190 robust SNP markers. Genome wide association analysis based on the 72,190 SNPs and grain mold rating data of 1414 landraces taken at Bako during 2016 season, identified a significant peak on chromosome 1 at position 68.3 MB (Figure 2.2A). The *quantile-quantile* (Q-Q) plot (Figure 2.2B) which indicates the plots of the observed and expected -log (*P*-values) for most SNPs showed that the GWAS model used for the analysis has accounted for population characteristics. The significant SNPs (FDR < 0.05) are S1_68388126, S1_68364163, S1_68364119, S1_68364116 and S1_68365986 which are all located at the classical sorghum kernel color locus (Table 2.1). The S1_68388126 was the most significant SNP located 23.9 Kb away from the other four tightly linked SNPs. S1_68388126, was localized to a non-genic region while the other four SNPs are all inside Sobic.001G397900 which encodes an R2R3 MYB transcription factor annotated as "*SIMILAR TO YELLOW SEED1*", related to the widely known *YELLOW SEED1* (*Y1*) R2R3 MYB transcription factor. However, *Y1* has not been annotated in sorghum genome. Initially, we thought Sobic.001G397900 could be the *Y1* gene, but, after blast

searches using the published sequence of *Y1* [131] against the current version of sorghum genome (*Sorghum bicolor* v3.1.1), we realized that *Y1* is different from Sobic.001G397900. Previous reports might have mistaken the position of the actual *Y1* gene in the sorghum genome because of a very high sequence similarity between *Y1* and Sobic.001G397900 and also because part of the *Y1* gene is deleted from the genome of the reference line (BTx623).

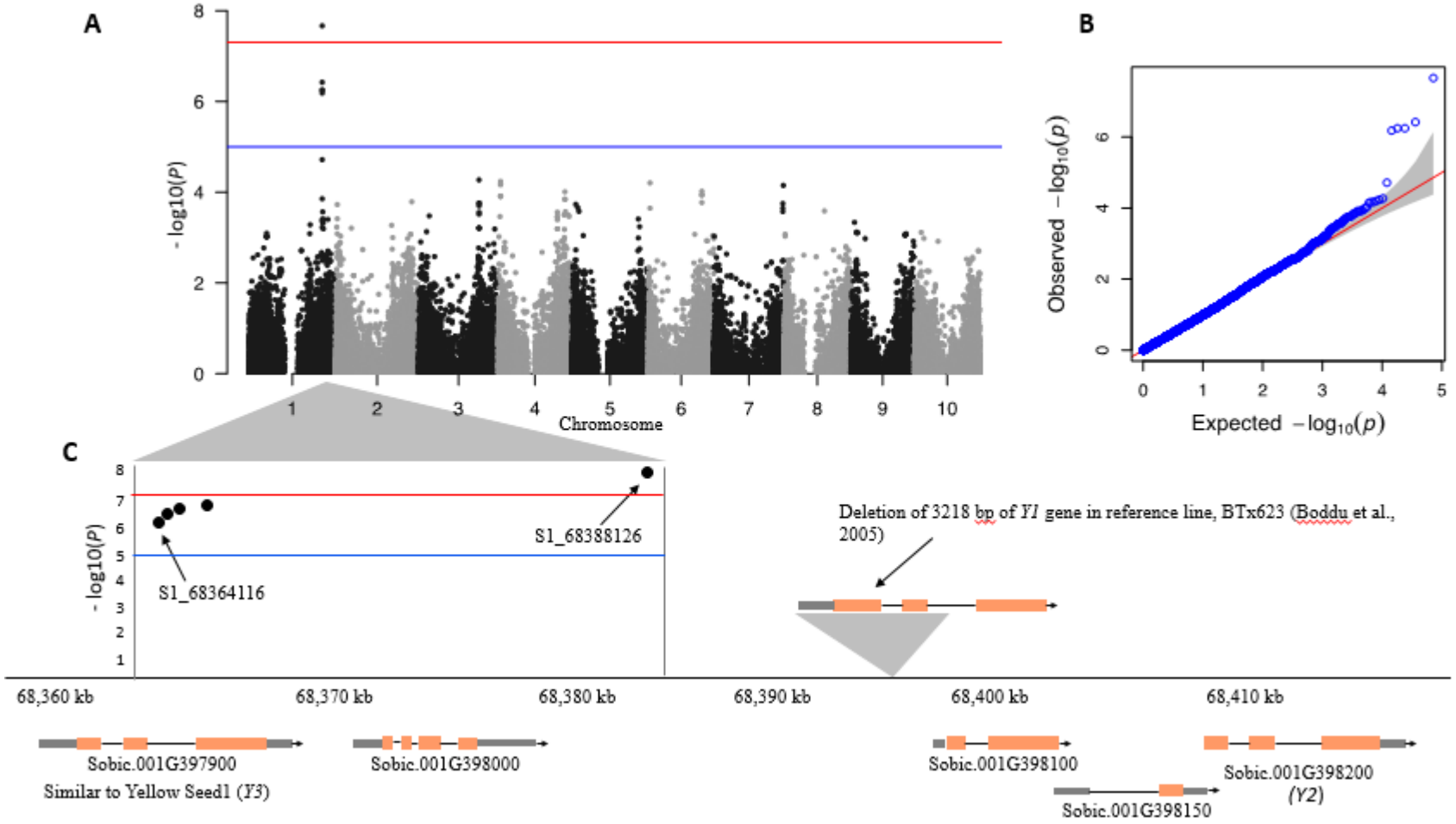


Figure 2.2. Genome wide association study of grain mold resistance. A) Manhattan plot indicating significant false discovery rate (FDR)-adjusted P-value for marker-trait association using compressed mixed linear model (CMLM). B) Quantile-Quantile (Q-Q) plot of CMLM. C) Detailed view around *Y1* and *Y3* MYB genes on chromosome 1. The horizontal red and blue lines on the Manhattan plot represent FDR adjusted p-values < 0.01 and < 0.05, respectively.

The Y*ELLOW SEED1* (*Y1*) locus also shows association with kernel color [132]. Y*ELLOW SEED1* (*Y1*) regulates the accumulation of 3-deoxyflavonoid pigments and phlobaphenes in pericarp, glumes and leaves of sorghum [131]. However, its association with grain mold resistance was not detected in any previous mapping efforts and the genomic organization of the locus has also not been studied in detail. Previous mapping studies for kernel color only indicated a rough position of *Y1* locus over a larger genomic region and details of the locus and the underlying gene are not known in natural variants [92]. *YELLOW SEED1* gene was identified using a stock line, RR-30, that carries a functional Y1-rr allele of the *Y1* locus. RR-30 originated from a spontaneous excision of a Candystripe1 (Cs1) transposable element from a line with the Y1-cs allele [131]. The authors also described a tightly linked and perhaps a non-functional R2R3 MYB (*Y2*) separated from *Y1* by an intergenic region of 9.084 Kb in head to tail orientation [133]. Our current genetic mapping using natural variants of sorghum and further characterization of the genomic region identified a significant grain mold resistance locus that includes *Y1*, the tightly linked pseudo gene (*Y2*), and another third MYB gene (Sobic.001G397900), which we designated as *YELLOW SEED3* (*Y3*) (Figure 2C).

Table 2.1. Significant SNPs located in the sorghum kernel color locus

SNP	**Nucleotide variation**	**P.value**	**FDR Adjusted P-values**	**R^2 value**
S1_68388126	A/G	2.16E-08	0.001561	0.241125
S1_68364163	G/A	3.78E-07	0.009439	0.238007
S1_68364116	A/T	5.62E-07	0.009439	0.237577
S1_68364119	G/A	5.62E-07	0.009439	0.237577
S1_68365986	T/C	6.54E-07	0.009439	0.237413

Alignment of sequences covering the *Y1* locus and blast search for position of *Y1* in the sorghum genome enabled us to describe the relative positions of all the three copies of the gene in the reference genome, BTx623 (Figure 2.2C). BTx623 has a 3218 bp deletion in *Y1* gene including 5' non-coding, promotor, exon1, intron1, exon2 and part of intron2 sequences which results in a non-functional gene [131]. Analyses of the partial sequence of *Y1* found in BTx623 showed that the deletion occurred at 68,397,090 bp in the current version of sorghum genome (*Sorghum bicolor* v3.1.1). The partial sequence of *Y1* available in BTx623 has been predicted as a MYB gene (Sobic.001G398100) containing two exons, a small exon1 and a large second exon which

corresponds to exon3 of the actual *Y1*. However, Sobic.001G398100 has not been annotated as *Y1*. Therefore, we propose that the corresponding *Y1* gene in sorghum genome is Sobic.001G398100. We further characterized some of the widely known sorghum lines for variations in the *Y1* gene using PCR markers. Sorghum lines IS9830, Tetron and ZZZ carry deletions similar to BTx623 while the gene is intact in RTx2911, RTx430, SQR, BTx378, SC35, SC283 and 555, although some of these lines contain sequence variations in other parts of the gene. Sequence variation at the *Y3* and *Y1* genes and relative grain mold resistance score in sorghum lines RTx430, RTx2911 and BTx623 (Figure 2.3) correlate with glume, grain color of the lines (Figure 2.4).

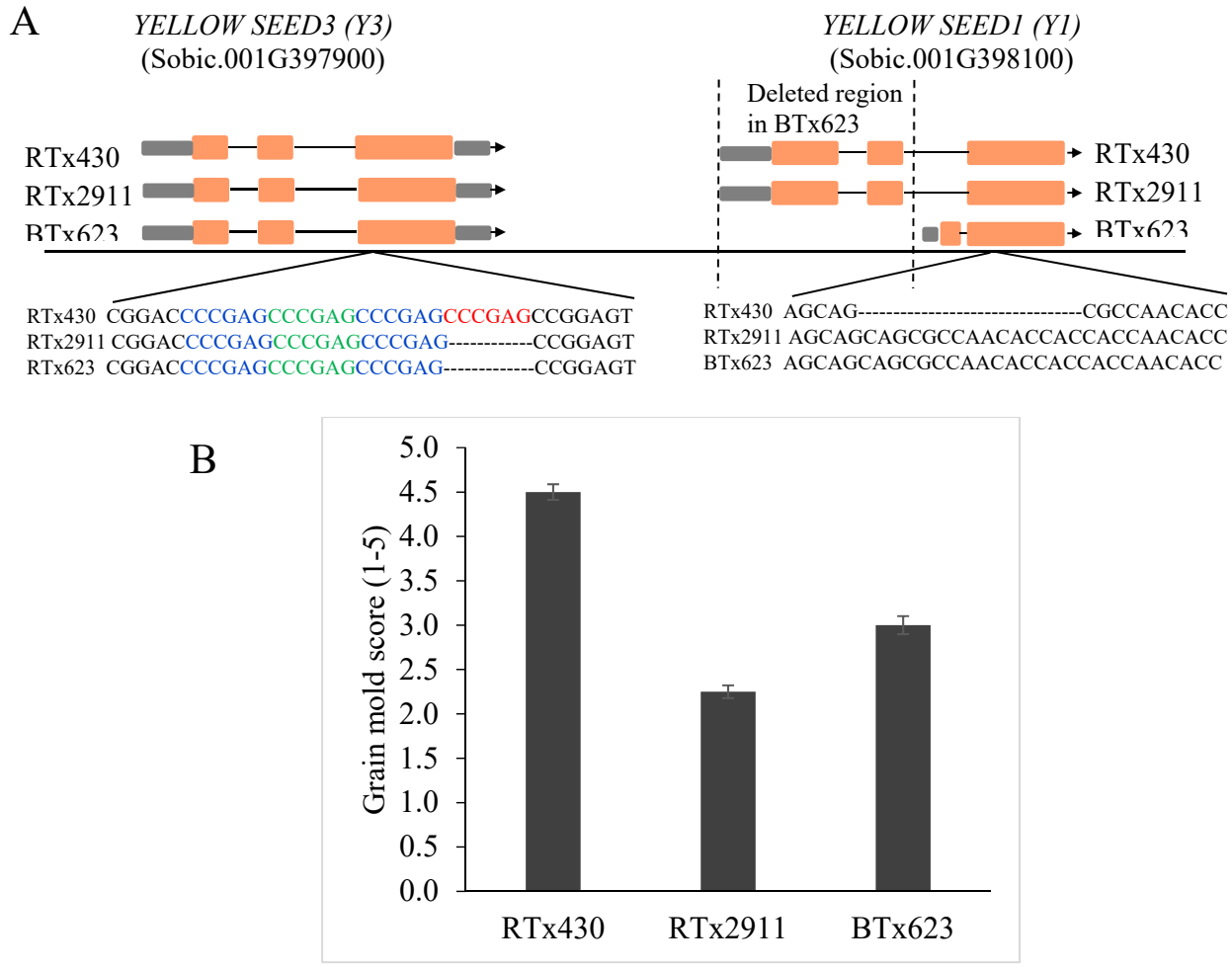


Figure 2.3. Genomic organization of *Y1* and *Y3* genes and grain mold resistance of variants. A) Genomic structure and partial illustration of sequence variation between sorghum genotypes at the grain mold resistance locus harboring *Y3* and *Y1* genes. B) Relative grain mold resistance of sorghum genotypes. Mold resistance was scored on a 1-5 scale, as described in Figure 2.1 under greenhouse condition. Data represent mean ± SE of four plants per line.

Sequencing of different parts of both *Y1* and *Y3* as well as visualization of the locus in the re-sequenced lines using integrative genomic viewer (IGV) indicated that the grain mold susceptible line, RTx430, has sequence rearrangements in both *Y1* and *Y3*, which could result in changes in amino acid sequences and premature stop codons (Figure 2.3A). The relative grain mold resistance level of the three genotypes based on our greenhouse experiments indicated that BTx623 is moderately susceptible to grain mold compared to the highly susceptible line, RTx430 (Figure 2.3B). The glume and grain colors of the sorghum lines (Figure 2.4) that are polymorphic for the locus provide further evidence about the functions of *Y3* and *Y1*. The grain and glume of RTx430 line do not appear to have any pigmentation. Whereas, BTx623 has red glume color but white grain. The resistant genotype, RTx2911, is red colored in both grain and glume. Additional studies are required to understand the structure and function of both *Y1* and *Y3* in diverse and distinct sorghum lines. These two genes appear to be functional in the mold resistant line, RTx2911. Previous reports suggest that the reference line, BTx623, produces apigeninidin, a 3-deoxyflavonoid pigment [129, 134, 135].

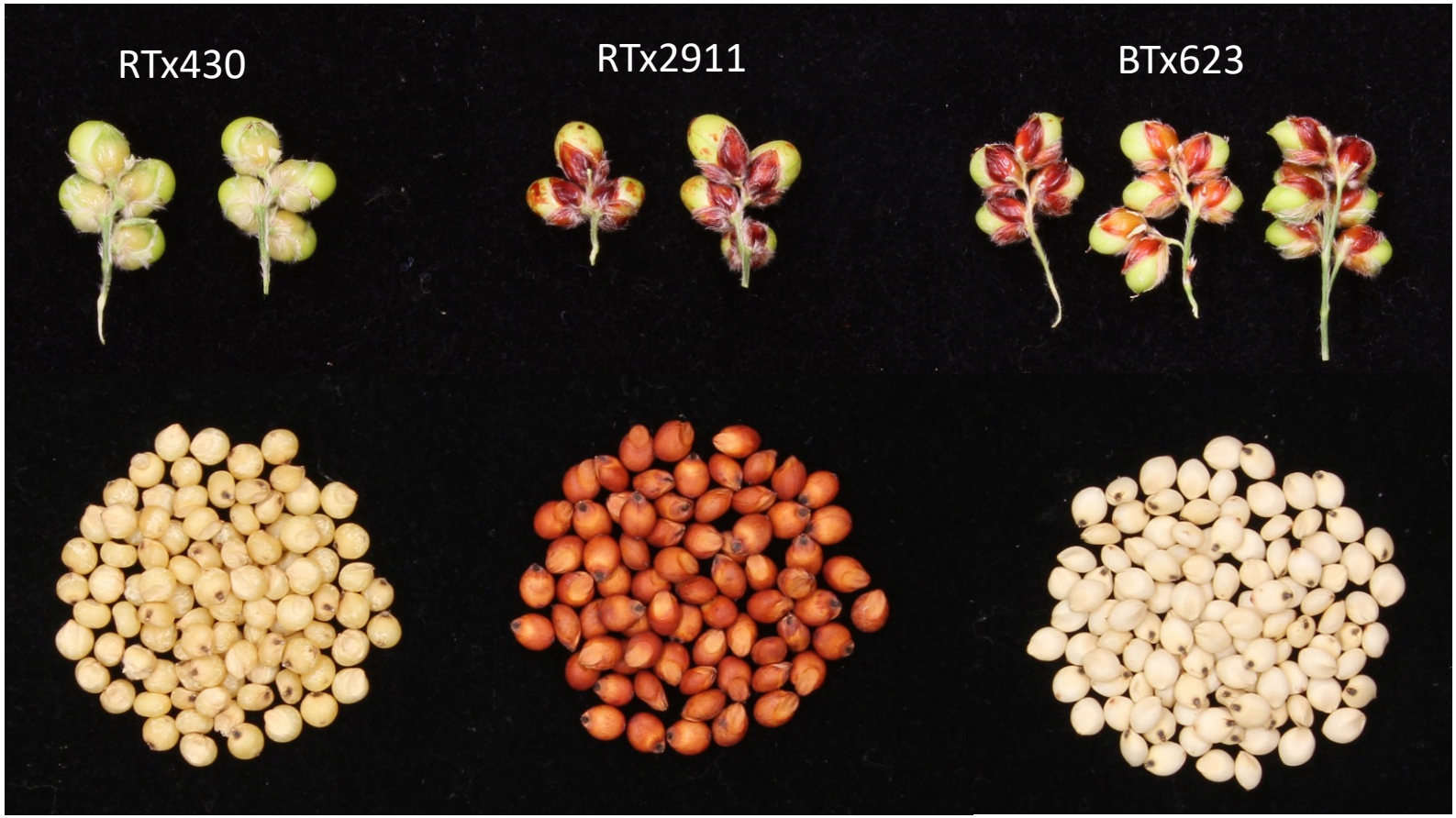


Figure 2.4. Glume and grain color of RTx430, RTx2911 and BTx623

The relative grain mold resistance and sequence polymorphisms at the Y1 and Y3 genes as well as the red glume in BTx623, suggest that *Y3* may have at least partially rescued the loss of *Y1* gene. Transgenic maize lines expressing the sorghum *Y1* gene were able to produce 3-deoxyanthocynidin and accumulate pigments in pericarp and cob glumes [89] suggesting *Y1* could be sufficient for the synthesis of these compounds. However, it is still unclear whether both *Y1* and *Y3* or either of the two are sufficient to induce the required phenotypes. Although, the physical distance between the SNP closest to *Y1* and the other four tightly linked significant SNPs in *Y3* gene may not be far enough to define these as two separate loci, our data suggest that *Y3* may be another functional R2R3 MYB transcription factor gene required for the biosynthesis of flavonoids. We have modeled a modified schematic representation of the biosynthetic pathway of flavonoid compounds with *Y1* or *Y3* genes involved in regulation of the pathway (Figure 2.5).

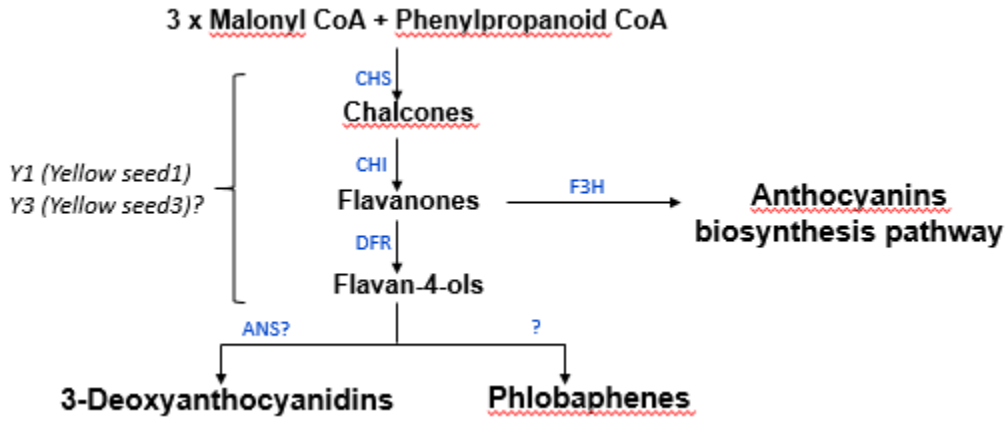


Figure 2.5. A simplified schematic representation of the sorghum biosynthetic pathway for flavonoid compounds with *Y1* or *Y3* genes involved in regulation of the pathway. Enzyme and genes names: CHS, Chalcone synthase; CHI, Chalcone isomerase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; F3H, Flavonoid 3'-hydroxylase. Pathway adapted from [89, 129, 136]

2.4.3 Expression of *Y1*, *Y3* and flavonoid biosynthesis pathway genes

We reasoned that if either of the two genes are not expressed in the resistant line, RTx2911, it will help understand which of the two genes may be required for the biosynthesis of 3-deoxyanthocynidins and pigment accumulation in grain and glumes of sorghum. However, both *Y1* and *Y3* are expressed in the grain and glume tissues of the resistant line, RTx2911, while both lacked expression in the susceptible line RTx430 (Figure 2.6A & B). The expression of both genes

in the developing grains seems to be high from soft to hard dough stages (about 20 to 30 days after flowering) compared to flowering or milk stages. Moreover, the expression of the flavonoid biosynthesis pathway genes *CHALCONE ISOMERASE* (*CHI*) and *DIHYDROFLAVONOL 4-REDUCTASE 3* (*DFR3*) were significantly lower in the susceptible, RTx430, compared to the resistant line, RTx2911 (Figure 2.6C & D).

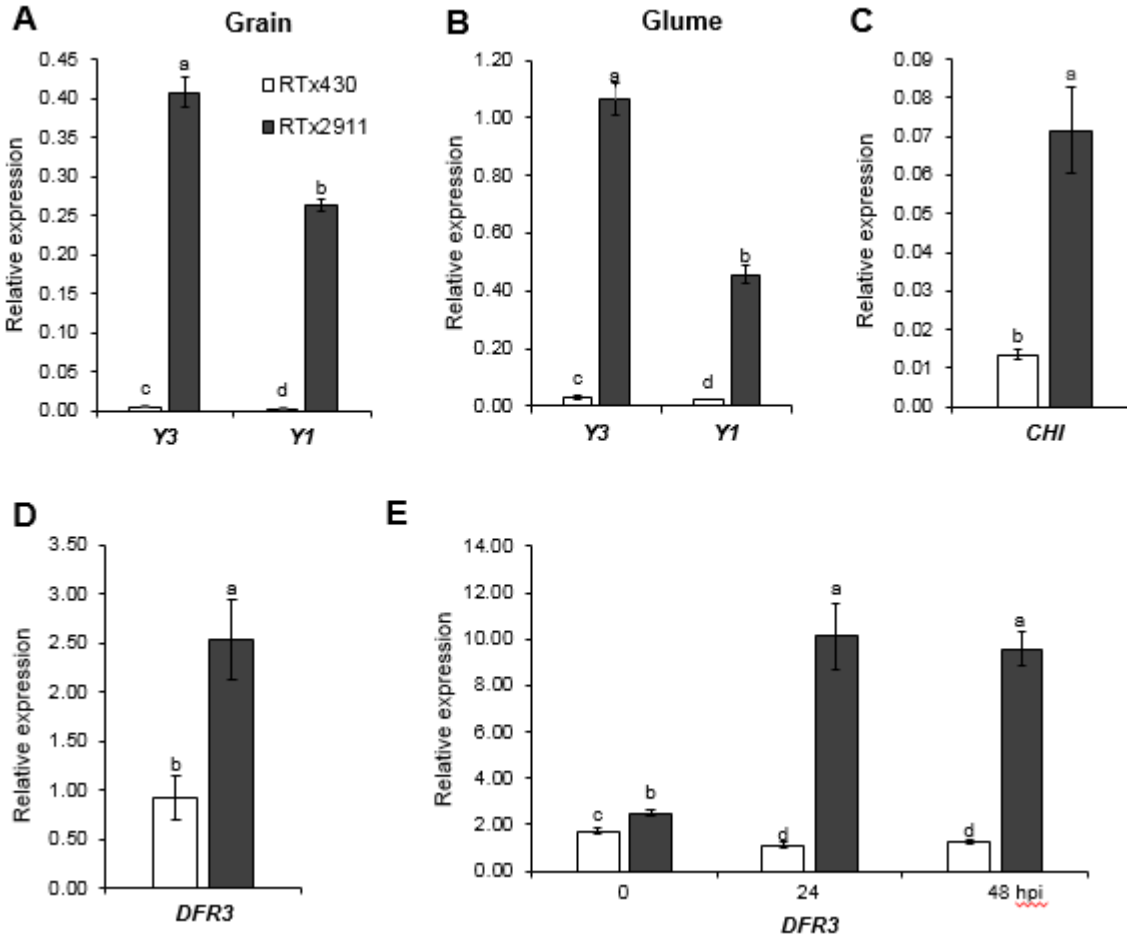


Figure 2.6. Expression of *YELLOW SEED3*, *YELLOW SEED1* and other 3-deoxyanthocyanidins biosynthesis pathway genes in grain and glume tissues of sorghum genotypes. The expression of *Y3* and *Y1* in grain (A) and glume (B) tissues. *CHI* (C) and *DFR3* (D) expressions in grain. Gene expressions were analyzed in response to grain mold fungi in grain and glume at 20 days after anthesis from susceptible RTx430 and resistant RTx2911 lines. The tissue samples in A to D were taken at 24 hours after inoculation. E. Expression of *DFR3* in developing grains after infection by a consortium of grain mold causing fungi.

In the anthocyanidin biosynthesis pathway, DFRs are enzymes involved in the reduction of flavanones to flavan-4-ols. In particular, sorghum *DFR3* is known to be pathogen inducible [129]. The expression of *DFR3* in response to inoculation of a mixture of fungal spores is significantly induced in the resistant line while its expression is reduced in the susceptible line within 24 hours after inoculation (Figure 2.6E).

2.4.4 Association between grain mold resistance and functionality traits

It is widely known that grain mold resistance is highly associated with grain pigmentation traits, although the genetics has not been well dissected, and in many instances, leads to misconception about genetic resistance to the disease and its relationship with grain functionality traits. For instance, brown pigmented sorghum lines which usually have high tannin content are generally considered more resistant to grain mold. While this is partially true, our current study indicates that red colored sorghums are comparable to brown colored lines in their grain mold resistance (Figure 2.7) suggesting factors other than tannin may play important role in resistance against the disease.

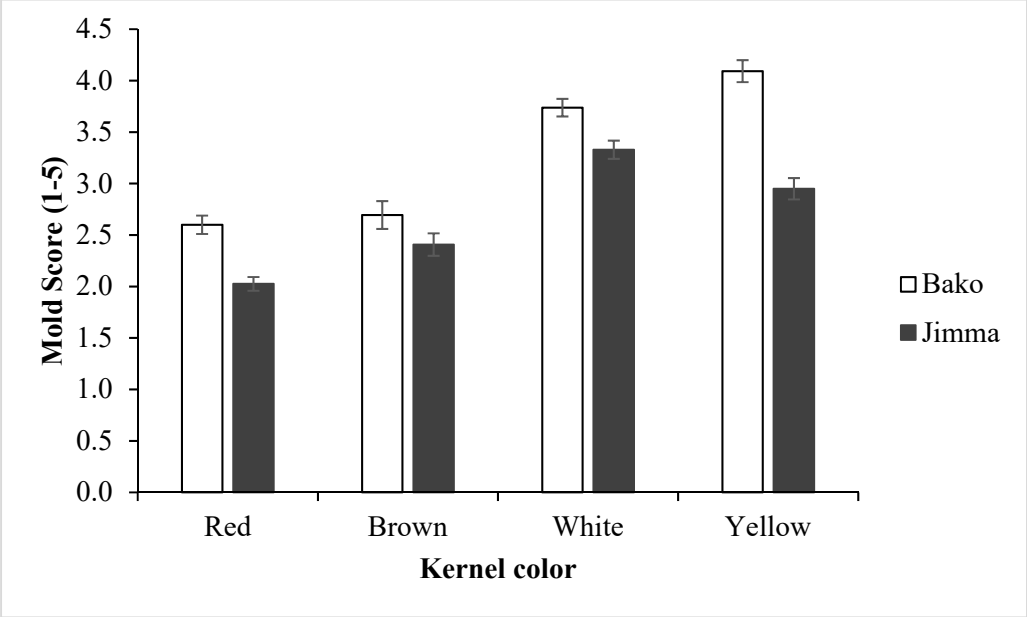


Figure 2.7. Relationship between kernel color and panicle grain mold rating at Bako and Jimma, Ethiopia. Mold scores represent means ± SE of 207 (red), 116 (brown), 178 (white) and 102 (yellow) colored landraces per location scored during 2017 season.

Sorghum lines with the other two major pigmentation groups (yellow and white) are more susceptible than red and brown kernel sorghum. A better understanding of association between genetic resistance to the disease and grain pigmentation traits as well as its impact on grain functionality requires good amount of germplasm with sufficient diversity. Most of the widely studied mapping populations such as the sorghum association panel may not fulfill these requirements. Because of limited use of modern cultivars, Ethiopian sorghum farmers maintained huge diversity for grain pigmentation traits. The current collection harbored various unique grain pigmentation phenotypes which could be useful for the genetic dissection of this trait and its relationship with grain quality and responses to biotic and abiotic stresses. It will be important to see how pigmentation trait is associated with grain functionality and resistant to grain mold. Recently, a series of functionality studies in Ethiopian sorghum collections targeting '*injera*' making quality was conducted. A comprehensive study on '*injera*' making quality of 139 key sorghum genotypes, mainly Ethiopian landraces was conducted recently [130]. Some of these landraces were scored for grain mold resistance in our study. Comparison of the pigmentation group of the top and bottom 10% of the collection ranked for '*injera*' making quality, revealed that ten of the best *injera* making landraces were found to be yellow seeded followed by three white and two red seeded (Table 2.2). On the other hand, the pigmentation of the bottom 10% of genotypes with poor '*injera*' producing qualities were red (7), brown (5), white (3) and no yellow seeded genotypes. Almost all the best '*injera*' producing yellow seeded landraces were found to be more susceptible to grain mold which could be a challenge for breeders to incorporate both traits in a single genotype. Improvement in these two traits may require a thorough understanding of the genetics of association between grain mold resistance and pigmentation traits. We also looked at the correlation between tannin content, grain mold rating and '*injera*' making quality using 93 lines in the collection. Tannin content was positively correlated with grain mold resistance both in 2015 (r=0.49) and 2016 (r=0.41) which look fairly high but the R^2 values for these correlations were weak in both 2015 (0.24) and 2016 (0.17). Moreover, '*injera*' making quality was weakly negatively correlated with grain mold resistance in both 2015 (r=-0.22) and 2016 (r=-0.25) with very low R^2 values of 0.05 and 0.06, respectively.

Table 2.2. Grain mold rating and kernel color of best and poorest injera making sorghum lines

No	Genotypes	Injera sensory score (1-9)	Kernel color	Grain mold rating (1-5)
1	ETSL 100310	8.4	Yellow	3.2
2	IS 38266	8.3	Yellow	4.0
3	IS 38263	8.3	Yellow	1.4
4	ETSL 100313	8.3	Yellow	4.0
5	IS 38358	8.2	Yellow	3.8
6	IS 38312	8.1	White	3.4
7	ETSL 101847	8.1	Yellow	3.6
8	IS 38400	8.1	Yellow	4.0
9	IS 38278	8.1	Yellow	3.0
10	IS 38281	8.0	Yellow	3.0
11	ETSL 100594	7.9	Red	1.0
12	IS 38282	7.9	Red	1.0
13	ETSL 100315	7.9	Yellow	2.2
14	ETSL 100311	7.9	White	1.0
15	ETSL 100152	7.9	White	3.4
16	IS 38392	2.4	White	1.0
17	IS 38328	2.3	Red	2.0
18	IS 38257	2.3	White	4.0
19	ETSL 100006	2.2	Brown	1.0
20	IS 38429	2.1	Red	2.4
21	ETSL 100013	1.9	Red	1.0
22	Emahoy	1.9	Brown	1.2
23	IS 38361	1.8	White	3.0
24	IS 38303	1.7	Brown	3.4
25	ETSL 100043	1.7	Brown	2.0
26	ETSL 100033	1.7	Red	1.4
27	ETSL 100111	1.3	Red	1.0
28	Abamelko	1.3	Brown	1.0
29	IS 25542	1.2	Red	1.0
30	ETSL 100004	1.0	Brown	1.0

NB. Injera sensory scores are overall acceptability scores (1 to 9) where 1 is highly undesirable and 9 is highly desirable [130]. Grain mold ratings are 1 to 5 where 1 is highly resistant and 5 is highly susceptible

2.5 Discussion

We set out to identify loci with major impact in grain mold resistance in sorghum by screening a naturally diverse population of Ethiopian landrace sorghum accessions. Genome wide association study identified the classical grain pigmentation locus *YELLOW SEED1* (*Y1*) and a linked and sequence related additional R2R3 MYB gene (*YELLOW SEED3*) to be significantly associated with grain mold resistance. Analyses of the genomic structure and sequences of the *Y3* and *Y1* and a third pseudo gene, *Y2*, suggest that the locus may have evolved from a duplication or unequal crossing over recombination event. The association of *Y1* with the production of 3-deoxyanthocyanidins and phlobaphenes accumulation in pericarp, glumes and leaves of sorghum and resistance to the hemibiotrophic pathogen *Colletotrichum sublineolum* has been described previously [131]. However, the genetic variation at this locus, the genomic structure and association with resistance to grain mold fungi was not studied previously. By studying, the sorghum genotypes, RTx430 and RTx2911, with contrasting grain color and mold resistance, we suggest that the R2R3 MYBs are key candidate genes for resistance. *Y1* and *Y3* genes are more expressed in grain and glume tissues of the grain mold resistant line, RTX2911, while they both show no expression in the susceptible line RTx430. The differences are underpinned by sequence polymorphisms at both the *Y1* and *Y3* genes. Interestingly, the two genes are co-regulated in response to inoculation by grain mold infection suggesting a functional overlap. In addition, the expression of the sorghum flavonoid biosynthesis genes is attenuated in grain mold susceptible lines in response to infection. The *DIHYDROFLAVONOL 4-REDUCTASE3* (*DFR3*) and the *CHALCONE ISOMERASE* (*CHI*) genes were significantly induced in the resistant line, RTx2911, in response to fungal infection but not in the susceptible line, RTx430. Thus, the data suggest that the *Y1* and *Y3* R2R3 MYB proteins, separately or in concert, may be involved in regulation of the production of secondary metabolites that contribute to restrict growth of molding fungi which is composed of necrotrophic species. These factors confer resistance in a race non-specific manner by modulating the accumulation of defense active secondary metabolites which makes them attractive for breeding purposes. Resistance to broad host pathogens and those causing grain mold fungi has been a major challenge due to their aggressive virulence strategies, the multiple fungal virulence factors, as well as the many fungal species involved in the pathogenesis of the disease.

The *YELLOW SEED1* gene has been implicated in the biosynthesis of the phytoalexin 3-deoxyanthocyanidins suggested to be an antimicrobial compound induced in response to pathogen

infection [88, 129]. Among the two major 3-deoxyanthocyanidins, luteolinidin and apigenindin, the former is more fungitoxic [91]. The association between grain mold infection and 3-deoxyanthocyanidins and flavon-4-ols have also been previously reported through analyses of grain phenolic compounds and secondary metabolites although genetic and molecular data were lacking [70, 71]. Flavon-4-ols is a common precursor of 3-deoxyanthocyanidins and phlobaphenes in the flavonoid biosynthesis pathway. Earlier studies considered testa pigmentation, which is attributed to tannin accumulation in pericarp of some sorghum lines, as the primary factor conferring grain mold resistance [69]. However, it will be important to differentiate between the contributions of tannins and 3-deoxyanthocyanidins especially since the two appear to have different desirability with respect to grain quality and nutritional benefits. In a separate GWAS analysis for kernel color, the *TAN1* and MYB loci as well as another locus on chromosome 3 were significant for kernel color (unpublished). The role of *TAN1* in the biosynthesis of 3-deoxyanthocyanidins and phlobaphenes is unclear, and need to be studied further. However there are speculations that *Y1*/*Y3*, *TAN1* and a third transcription factor (bHLH) form a MBW (MYB-bHLH-WDR) protein complex which may regulate transcription of flavonoid biosynthesis [136]. The fact that, in our study, both *Y1*/*Y3*, *TAN1* as well as a locus on chromosome 3 were significant for kernel color supports the idea of transcriptional regulation by the MBW protein complexes. The *TAN1* locus being significant for kernel color may indicate the regulatory role of *TAN1* in the accumulation of 3-deoxyanthocyanidins and phlobaphenes in pericarp. *Y1* or *Y3* genes may interact with the sorghum *TAN1* and unknown *bHLH* transcription factor in regulating flavonoid biosynthesis pathway as there are evidence of transcriptional regulation of the pathway by WD-repeat (WDR) protein, MYB, and bHLH complex such as in *Arabidopsis* [137]. Recent data in sorghum also suggest that *Sb02g006390* gene which encodes a putative *bHLH* transcription factor involved in accumulation of tannins in testa [92], but the presence of a similar *bHLH* transcription factor required for accumulation of 3-deoxyanthocynidin and phlobaphene in pericarp is unclear. Moreover, the study by [138] suggests that *Y1* may require *TAN1* as a key co-regulatory factor to induce 3-deoxyanthocyanidins in sorghum and maize which may also be true for *Y3*. The important roles of the 3-deoxyanthocyanidins phytoalexins in disease resistance as opposed to tannins has been strongly established, however, additional genetic and biochemical studies are required to uncouple the resistance functions of these compounds. Although associations between grain mold resistance and kernel color have been known, no genetic evidence were provided. Our study

provides genetic data supporting the role of key regulatory factors in the production of kernel pigments and antimicrobial activity against grain mold fungi.

The pathogenesis of grain mold fungi and genetics of resistance to Fusarium head blight (FHB) in wheat and ear rot in maize appear to share similar mechanisms and complexities. The fungal species that cause grain mold disease in sorghum have broad host ranges and cause various grain diseases in other cereals. FHB resistance has been identified in wheat and categorized depending on initial infection (Type I), disease spread (Type II), toxin accumulation (Type III), kernel infection (Type IV) and yield reduction (Type V) [139, 140]. Plant height and anther extrusion are widely known for their negative correlation with Type I FHB susceptibility [141, 142]. Several FHB QTLs overlap with plant height QTLs in wheat [140, 143-145]. The mapping study in sorghum [49] which identified grain mold resistance QTLs co-localized with plant height and peduncle length could be a good example of the Type I resistance of wheat which may be based on avoiding initial infection rather than a direct resistance mechanism. Similarly, due to the growth stimulants choline and betaine, anthers have been shown to promote FHB infection in wheat [146]. Studying such floral traits, plant architecture and phonological traits in sorghum would help understand and differentiate the various mechanisms of resistance to grain mold in sorghum.

In addition to genetic studies, through metabolomic profiling approaches, the hydroxycinnamic acid amides (HCAAs), coumaroylagmatine and coumaroylputrescine were recently identified to contribute to FHB resistance in the rachis of wheat [147]. The characterization of the wheat gene encoding agmatine coumaroyl transferase linked to the accumulation of these compounds in response to *F. graminearum* infection, led to the identification of polymorphisms that account for difference in resistance. HCAAs reduce pathogen ingression through their antimicrobial and cell wall reinforcement properties [148-150]. These observations suggest that the accumulation of antimicrobial compounds may be an effective mechanism for such complex diseases. The observation from FHB, and the increased accumulation of 3-deoxyanthocynidins phytoalexins implied from our research further support that resistance to grain mold could be achieved by enhancing accumulation of compounds in floral tissues such as the glumes and rachis which may provide protection at the site of infection and prevent further fungal ingress. The network of genes in the flavonoid biosynthesis pathway regulated by the *Y1*

and/or *Y3* transcription factors leading to increased synthesis of the 3-deoxyanthocyanidins or other fungi-toxic compounds may provide avenues for selection.

In a recent study, [130] sorghum genotypes with high total starch but low amylose were found to have good '*injera*' making quality, while tannin had a negative correlation with both total starch in grain and '*injera*' as well as with all sensory parameters. We examined the relationship between grain mold resistance and '*injera*' making qualities of some of the lines studied by [130]. We observed a negative correlation between grain mold resistance and grain functionality traits such as '*injera*' making quality in sorghum. The yellow seeded sorghum landraces appear to be the most desirable for '*injera*' making quality. By contrast, these lines are susceptible to grain mold which poses a challenge for combining quality and resistance traits. A better understanding of the impact of grain mold resistance factors such as the 3-deoxyanthocyanidins, tannins or their derivatives on '*injera*' making quality is necessary to improve quality without compromising disease resistance. Genetic screens for sorghum variants that accumulate 3-deoxyanthocyanidins, retain the yellow kernel color, and fungal resistance may be important. It is still unknown whether 3-deoxyanthocyanidins and phlobaphenes affect '*injera*' making property of sorghum lines the same way as tannins.

2.6 Conclusions

Genome wide association mapping using Ethiopian sorghum landraces identified a major grain mold resistance QTL containing tightly linked transcription factors, *YELLOW SEED1* (*Y1*), and a second R2R3 MYB gene, *YELLOW SEED3* (*Y3*), defining a narrow genomic region that could be used for grain mold resistance selection. Sequence polymorphisms and expression profile at this putative target locus provided genetic evidence for our conclusions. The *Y1* and *Y3* genes are expressed in response to fungal infection in developing grains and glume tissues of resistant lines while there was no expression in the susceptible lines. Similarly, expression of genes in flavonoid biosynthesis pathway is enhanced in developing sorghum grains in response to pathogen infection in resistant lines with functional *Y1* and *Y3* genes. Overall, our study suggested genes in the flavonoid biosynthesis pathway regulated by either or both of the candidate transcription factor, R2R3 MYB genes is important for grain mold resistance. This regulation may lead to an increased synthesis of secondary metabolites such as 3-deoxyanthocynidin phytoalexins in response to fungal infection in developing grain and glume tissues which may provide an effective resistance

to broad host molding pathogens. Interestingly, some of these secondary metabolites in the flavonoid biosynthesis pathway may affect grain functional qualities and health benefits from the grain. A better understanding of the genetic components of the sorghum flavonoid biosynthesis pathway, its products, and interactions with other disease resistance and quality traits is likely to provide avenues for crop improvement in nutrition and disease resistance traits. The exact molecular mechanisms of how *Y1* and *Y3* regulate the downstream genes, and ultimately the production of the compounds, and the specific point of actions needs to be determined in future studies. Expression of Y1 and/or Y3 into susceptible sorghum lines through transformation may help differentiate the roles of *Y1* and *Y3*.

CHAPTER 3. GENOME-WIDE ASSOCIATION ANALYSIS REVEALS SEED PROTEIN LOCI AS DETERMINANTS OF VARIATIONS IN GRAIN MOLD RESISTANCE IN SORGHUM

A version of this chapter was previously published by *Theoretical and Applied Genetics* [151], https://doi.org/10.1007/s00122-020-03762-2

3.1 Abstract

Grain mold of sorghum which results from concurrent infection by multiple fungal species, starting at the early stages of grain development, is the most important disease of the crop. The genetic architecture of resistance to grain mold is poorly understood especially in landrace germplasm. We conducted a multi-stage disease rating for resistance to grain mold, under natural infestation in the field, using a diverse set of 635 Ethiopian sorghum accessions. A number of accessions with near complete immunity to the disease were identified. Genome-wide association analyses (GWAS) using 173,666 SNPs with multiple models, identified two novel loci consistently associated with grain mold resistance across environments. Sequence variation at new loci containing sorghum *KAFIRIN* gene encoding a seed storage protein affecting seed texture, and *LATE EMBRYOGENESIS ABUNDANT* 3 (*LEA3*) gene encoding a protein that accumulates in seeds, previously implicated in stress tolerance, were significantly associated with grain mold resistance. The *KAFIRIN* and *LEA3* loci were also significant factors in grain mold resistance in accessions with non-pigmented grains. Moreover, we consistently detected the known SNP (S4_62316425) in *TAN1* gene, a regulator of tannin accumulation in sorghum grain to be significantly associated with grain mold resistance. Identification of loci associated with new mechanisms of resistance provides fresh insight into genetic control of the trait while the highly resistant accessions can serve as sources of resistance genes for breeding. Overall, our association data suggest the critical role of loci harboring seed protein genes, and implicate grain chemical and physical properties in sorghum grain mold resistance.

3.2 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is among the world's most important cereal crops used for food, feed, and biofuels with unique adaptation to dry lands and nutrient deficient soil conditions. Sorghum grain is used as a staple food for millions of people in developing countries while the stalk and leaves are used as livestock feed. Grain mold is a widespread and most important disease of sorghum particularly in regions with high humidity during flowering, grain development and harvest. The disease is caused by multiple pathogenic fungal species belonging to different genera. Besides its impact on grain yield, grain mold is highly detrimental to grain quality due to contamination by mycotoxins [50, 152, 153]. *Aspergillus* and *Fusarium* are widespread and major genera that produce mycotoxins that contaminate sorghum pre and post-harvest [153]. Genetic resistance is considered a major avenue to control the disease because of the complexity of fungal species causing the disease and the limited feasibility of chemical control.

Resistance to grain mold in sorghum has been largely associated with physical and chemical kernel properties. Kernel traits including testa pigmentation, higher levels of phenolic compounds, pericarp color and kernel hardness were associated with resistance to the disease [70-72, 154]. Unfortunately, resistance through these mechanisms is associated with lower nutritional value particularly in regions where sorghum grain is a staple food. There have not been new mechanisms of resistance identified that is not impacting the nutritional quality of the grain. Therefore, development of cultivars resistant to the disease while maintaining high nutritional quality traits have been challenging for breeders. Identification of new mechanisms of resistance has been hampered by complexity of grain mold diseases phenotyping, the unique nature of the tissue being studied, and the greater impact of the environment on the disease. Seeds are rich sources of carbon that makes them especially susceptible to fungi but also have a declining resistance in the course of maturity. Photoperiod sensitivity of most of the sorghum germplasm available in gene banks on the other hand hindered phenotyping of valuable sorghum collections in areas where facilities permit. Therefore, a comprehensive phenotyping of natural variants of sorghum in tropical environments is vital to identify resistance sources and map genomic regions associated with resistance. In this regard, sorghum accessions maintained in national gene banks of countries that are centers of origin and diversity for the crop need to be exploited to discover novel traits of interest. Recently, we demonstrated the potential of such genetic resources through a large-scale phenotyping and genotyping of Ethiopian sorghum landrace collection [155].

Phenotyping for grain mold resistance using the landrace collection identified a large number of resistant accessions and the sorghum *MYB* locus containing *Y1* and *Y3* was strongly associated with resistance [114]. Conversely, a similar study in the US sorghum association panel (SAP) identified very few resistant accessions and two genetic loci associated with low seed deterioration and one locus implicated in emergence rate [76]. A separate GWAS for grain mold resistance in SAP involving individual or combined inoculation with selected grain mold causing species resulted in significant SNPs associated with resistance to the disease [77]. Five QTLs located on chromosomes 4 to 7 and 9 were detected for grain mold resistance in an earlier study on 125 F5 recombinant lines derived from a cross between the resistant Sureno and susceptible RTx430 sorghum lines [49]. Some of these QTLs were co-localized with plant stature traits (height and peduncle length). Similarly, in a study on sorghum mini core accessions [75] two loci (on chromosome 2 and 8) were detected that are associated with grain mold resistance. Except in few of these studies, such as the detection of the *MYB* locus [114], major effect QTLs with direct role in grain mold resistance have not been identified, partly attributed to the limited resistance source materials used in many of the studies.

Besides good source germplasm, accurate phenotyping and implementation of appropriate analytical tools are critical to identify genes underlying quantitative traits. Rating grain mold disease levels can be difficult because symptoms are not always obvious [25]. It may be easier to rate highly resistant or highly susceptible materials in the background of light-colored grains, but disease rating in grains with intermediate level of resistance and colored grains (brown, red and intermediate color groups) are difficult. Therefore, appropriate data quality control systems through use of proper controls and multiple rating approaches followed by heritability measures are important. Moreover, disease escapes in late maturing/flowering materials may affect identification of germplasms with real resistance to the disease [25] and leading to detection of false positives by genome wide association mapping. Hence, it is important to ensure that test germplasms are within the same maturity group or the duration of wet period during grain development and maturity are sufficiently long to accommodate all maturity groups. In addition, use of appropriate analyses models is important as results may vary depending on the models used. Genome wide association analysis models have been evolving over the last couple of years providing a new opportunity in terms of gain in statistical power and reduction in computational time [111]. With such computational advances, it would be interesting to see how the limitations

in identification of loci associated with complex phenotypes could be resolved. Moreover, it would also be important to look at how the new models affect detection of new or known loci that were previously identified using the earlier models.

Here, we present results of a genome wide association study in large and diverse sorghum landrace accessions from Ethiopia that was initiated to identify loci that contribute to grain mold resistance. A large number of highly resistant accessions and loci that strongly associate with resistance were identified based on multistage grain mold disease rating of accessions.

3.3 Materials and methods

3.3.1 Plant material

A subset of 655 sorghum accessions, sampled from the recently described large collection of Ethiopian sorghum germplasm were used for this study [114, 155]. The accessions represented landrace sorghums originally collected from sorghum growing areas in the country and maintained at Ethiopian Institute of Biodiversity and the national sorghum research program. Late maturing accessions (>130 days to flowering) were excluded to avoid materials that escape infection. Seeds used for this study were obtained from single heads of each accession subjected to multiple rounds of selfing, which were subsequently genotyped. This resulted in true to type accessions and avoided within accession variability that can affect both phenotyping and genotyping.

3.3.2 Phenotyping

Field evaluation for grain mold resistance was conducted at two locations, Jimma (latitude of 7^0 $40^{'}$ N and longitude of 36^0 $47^{'}$ E with elevation of 1,753 meters above sea level) and Bako (latitude of $9^0 8^{'}$ N and longitude of $37^0 3^{'}$ E having an elevation of 1650 meters above sea level) in Ethiopia during 2017 and 2018 main cropping seasons. Both locations have long rainy months and warm weather, environmental conditions that favor disease which make the sites suitable for screening germplasm for disease resistance. The sites are located in the Western part of Ethiopia, which receives the highest annual rainfall. Jimma receives an average annual rainfall of about 1500 mm. Bako receives slightly variable rainfall ranging from 1200 to 1600 mm most of which is distributed between April and October [156]. Planting time ranges from late April to mid-May in both areas. Each accession was planted in single row plot of 3 m length with 20 cm between plants

and 75 cm between rows. To quantify for any obvious trend in spatial grain mold distribution in the field, each experimental plot was referenced by two-dimensional spatial position in the field [i.e. columns (ranges) and rows] during each test year and location. Since there were no standard checks with similar maturity time as the landrace accessions, known resistant and susceptible accessions based on a prior study [114] and elite breeding lines with known reaction to grain mold were used as checks. Two resistant (Dagim and ETSL 100612) and three susceptible local checks (IS 38285, ETSL 101853 and Melkam) were each randomly planted in replicates of six. In addition, 15 non-replicated breeding lines were included in the study to make a total number of 700 plots, which was divided into 14 columns (ranges). Each column had 50 plots (rows). The experiment was set up in a completely randomized plots arranged in a modified augmented design where checks were randomly replicated across the experimental field instead of within each block (column). Grain mold resistance response was carefully determined based on visual rating of unharvested panicles in field plots (FGMR), excised mature panicles (PGMR) and threshed grains (TGMR). FGMR was rated for each accession on whole plot basis by observing panicles prior to harvest whereas PGMR was scored in the laboratory from five representative panicles sampled prior to harvest. Then the panicles were threshed, and grains were rated to obtain TGMR. Grain mold rating scale of 1-5 was used, where 1 represents highly resistant, and 5 is highly susceptible. Out of the initial 655 accessions, 20 had incomplete data. Therefore, 635 accessions with complete data were used for the current GWAS. Moreover, data on days to flowering, glume color and plant height were recorded. Days to flowering (DTF) was recorded as the number of days from planting to 50 % flowering in a plot. Glume color was recorded as the grains reached about hard dough stages. Plant height (PHT) was measured from base to the top of the panicles of three plants in each plot and their means used for analysis.

3.3.3 Mycoflora analysis

Mycoflora analysis was performed using seeds from 10 resistant and 10 susceptible accessions as previously described [25, 157]. Ten seeds were randomly selected from each accession and used for the analysis in two replicates (five seeds per replicate). Each seed sample was examined using compound microscope to identify fungal species based on morphological features following a pictorial guide for identification of grain mold fungi on sorghum [157].

3.3.4 Genotyping-by-sequencing

DNA extraction, genotyping-by-sequencing (GBS) and SNP calling were as described [155] while SNPs for the 635 accessions were subsetted from the raw data available at Purdue University Research Repository (doi:10.4231/PYQV-AT79). Following quality control by filtering with more than 20 % individual missing rate and minor allele frequency (MAF) < 0.05, a total of 173,666 SNPs were obtained using TASSEL 5.0 [119]

3.3.5 Data analysis

3.3.5.1 Phenotypic analysis

As a measure of data quality and reproducibility, spatial and correlation analysis, and analysis of variance (ANOVA) and heritability estimates were conducted using both individual year and location data as well as combined data over years for each rating method in each location. To adjust for any spatial trend in grain mold distribution across the test plots, spatial analysis was performed using the SpATS model [158] in R. Since there were no spatial trends, adjustments were not required. Pearson correlation analysis was conducted using the psych package [159]. ANOVA was performed following linear mixed-effect model using lme4 package [160]. Variance component was estimated for accessions using the accessions as random and years as fixed factor. Years were used as replicates to conduct combined analysis of data from different years. Heritability was calculated as ratio of variance due to accessions divided by total variance. Overall variability in grain mold resistance in the accessions is summarized as frequency distribution of TGMR values combined over the years in each location.

3.3.5.2 Principal components analysis-based population genetic structure

Principal components analysis (PCA) was performed by generating the components using 'prcomp' function and plotting by ggplot package in R. The first two components, which explained most of the variation, were used to plot the PCA and individual accessions were colored based on its respective population of K = 11 determined previously [155].

3.3.5.3 Genome-wide association analysis

Grain mold phenotypic data generated using 635 accessions and 173,666 SNPs were used to conduct GWAS by using multiple models including GLM, MLM [108], CMLM [109], FarmCPU [110] and BLINK [111] implemented in GAPIT package version 3 [161] in R. Since

grain mold resistance in sorghum is commonly associated with kernel color (brown and red) [114], a separate GWAS was conducted using a subset of 373 non-pigmented accessions by excluding those with brown and red grain color. Brown and red grain sorghums usually contain tannins and phenols in seeds, which have nutritional drawbacks as food or feed although some phenols may have health benefits [2]. The goal of conducting GWAS with the non-pigmented accessions, potentially producing no tannins and/or phenols is to identify resistance loci independent of grain flavonoids and with lower anti-nutritional factors. Individual year and location data as well as data combined over the years (replicates) in each location was used to conduct GWAS. To account for population structure and familial relatedness, the first three principal components were used as covariates, while kinship matrix automatically generated from the genotype data by GAPIT using the VanRaden method [162] was used depending on each of the model's specifications. For instance, FarmCPU uses a set of markers associated with a casual gene as cofactor instead of kinship to avoid overfitting and eliminate confounding between kinship and testing markers iteratively [110]. More recently, along with improvements in statistical power and reduction in computing time compared to FarmCPU, the new model called Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) is set to eliminate FarmCPU's requirement that quantitative trait nucleotides (QTNs) are evenly distributed in the genome [111]. We evaluated both FarmCPU and BLINK along with the earlier models (GLM, MLM and CMLM). Efficiency of the models was determined by identifying previously known QTLs in sorghum.

3.4 Results

3.4.1 Grain mold resistance among accessions

The study revealed large number of accessions with highly resistant to grain mold rating (Figure 3.1). A total of 43 and 46 accessions had threshed grain mold rating (TGMR) value of 1 at Jimma and Bako, respectively implying near complete immunity to the disease. Similarly, 202 accessions at Jimma and 189 accessions at Bako had scores between 1 and 2. Most of the resistant accessions consistently expressed the trait across the two locations. A total of 139 accessions had mean grain mold score of less than 2 which indicates that they are highly resistant to the disease, whereas, the overall mean score across all accessions was 2.8. This indicated that the current set of sorghum accessions represented a large proportion of genotypes that are resistant to grain mold.

We found accessions that are consistently highly resistant or susceptible to grain mold across all environments based on all rating methods. Two accessions (ETSL 100612 and ETSL 101178) were consistently the most resistant to grain mold while IS 38285 was consistently the most susceptible accession. These accessions are ideal for use as resistant and susceptible checks for future genetic studies.

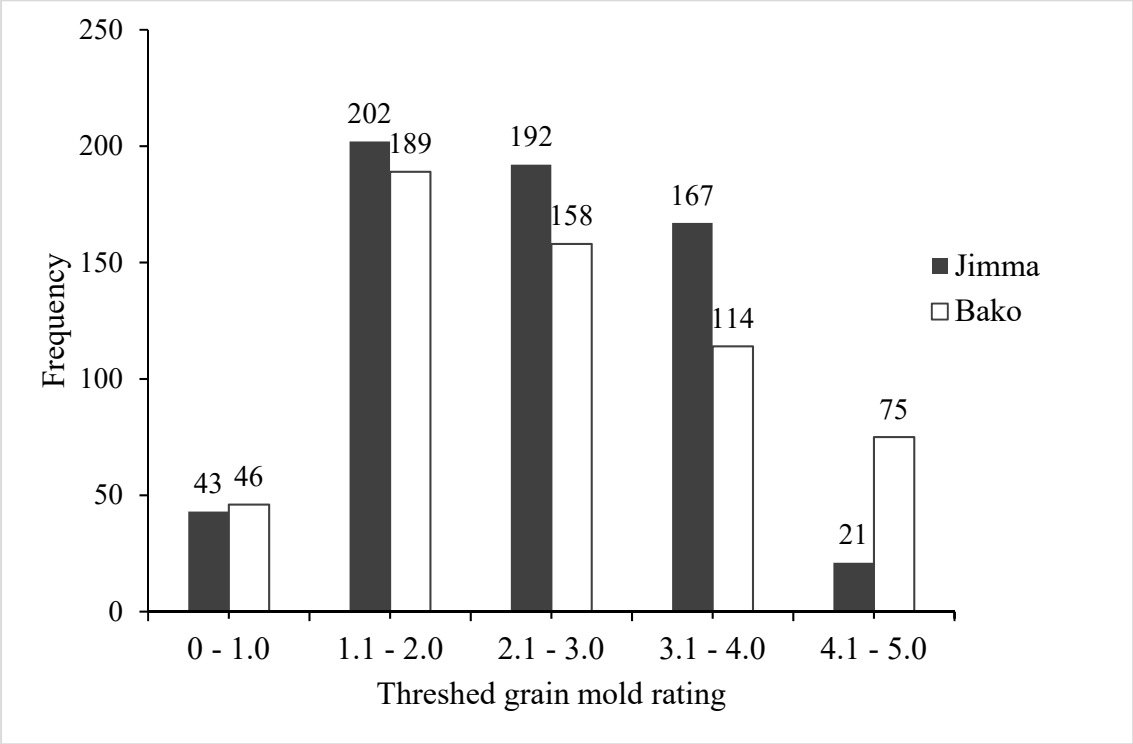


Figure 3.1. Grain mold resistance among sorghum accessions: frequency distribution of threshed grain mold rating (TGMR) across sorghum accessions. TGMR values for each accession are means of two years (2017 and 2018) where each year's scores were obtained as rating of bulk of threshed grain from five panicles per accession. Rating scale of 1-5 was used, where 1 represents highly resistant, and 5 is highly susceptible.

3.4.2 Heritability

Broad sense heritability (H^2) is an important measure of proportion of trait variance that is due to genetic factors. Heritability estimated for grain mold resistance scores obtained through the three methods indicate considerable variations at Jimma. FGMR based scores showed heritability of 32 % while PGMR and TGMR based scores had 60 and 62 % heritability, respectively (Figure 3.2). At Bako, FGMR based scores had a heritability of 47 % while PGMR scores had 53 %

heritability. These values are substantially high for a complex phenotype such as grain mold. Heritability was not estimated for TGMR at Bako, because this score was taken only during the 2018 season at this location.

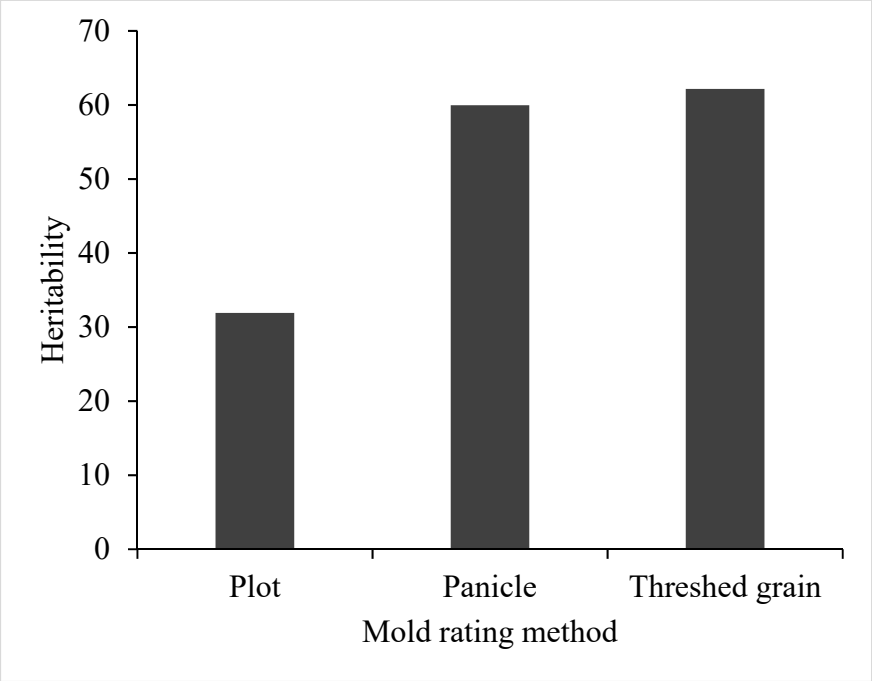


Figure 3.2. Variation in heritability of grain mold data between rating methods

3.4.3 Trait correlations

Phenotyping for grain mold resistance is complex and can be affected by scoring method, maturity time and phenology. Hence, we compared correlation among the mold scores generated through different rating methods, with flowering time and plant height. At Jimma, FGMR was positively correlated to both PGMR ($r = 0.59$, $p < 0.001$) and TGMR ($r = 0.53$, $p < 0.001$) while PGMR and TGMR also showed a very strong positive correlation ($r = 0.8$, $p < 0.001$) (Table 3.1). Similarly, at Bako, FGMR was positively correlated to both PGMR ($r = 0.76$, $p < 0.001$) and TGMR ($r = 0.65$, $p < 0.001$) while PGMR and TGMR also showed a significant positive correlation ($r = 0.69$, $p < 0.001$). FGMR was positively correlated to DTF ($r = 0.28$, $p < 0.001$) while PGMR and TGMR were not significantly correlated, respectively, to DTF at Jimma. At Bako, all three had significant positive correlation to DTF, which suggested that late flowering accessions to be more susceptible to grain mold at this location, but at Jimma such trend was not consistent between

the rating methods. PGMR and TGMR were negatively correlated to PHT at Jimma, and were weakly but positively correlated at Bako. FGMR was not correlated to PHT at both locations. Correlations between the rating methods, flowering time and plant height based on individual year and location followed similar trend with the combined data across the two years.

Table 3.1. Pearson correlation coefficients among mold rating methods, flowering time and plant height

Rating method	Jimma				Bako			
	PGMR	TGMR	DTF	PHT	PGMR	TGMR	DTF	PHT
FGMR	0.59***	0.53^{***}	0.28^{***}	-0.14^{NS}	0.76^{***}	0.65^{***}	0.24^{***}	0.03^{NS}
PGMR		0.80^{***}	0.11^{***}	-0.22^{***}		0.69^{***}	0.32^{***}	0.10^{***}
TGMR			0.04^{NS}	-0.21^{***}			0.34^{***}	0.18^{***}

Asterisk represent significance levels (*** = P ≤ 0.001, NS = Nonsignificant); FGMR = field (plot) grain mold rating, PGMR = Panicle grain mold rating, TGMR = Threshed grain mold ratingDTF = days to flowering; PHT = plant height

3.4.4 Principal components analysis

The first two principal components explained 75.6 % of the total genotypic diversity (Figure 3.3). The racial groups of accessions within each cluster group provide some insight to our understanding of the process that shaped the observed genetic diversity. Along the three extreme edges, we see the importance of the sub-populations 6, 2 and 9 for the overall diversity in this set of accessions. Sub-population 6 contains mostly the dura sorghums and intermediate race between dura and other races. Closely located to this sub-population is sub-population 1 which contains mostly caudatum and dura-caudatum mixed race. Sub-population 9 contains mostly the bicolor race while sub-population 2 contains intermediate accessions of bicolor with dura and caudatum races.

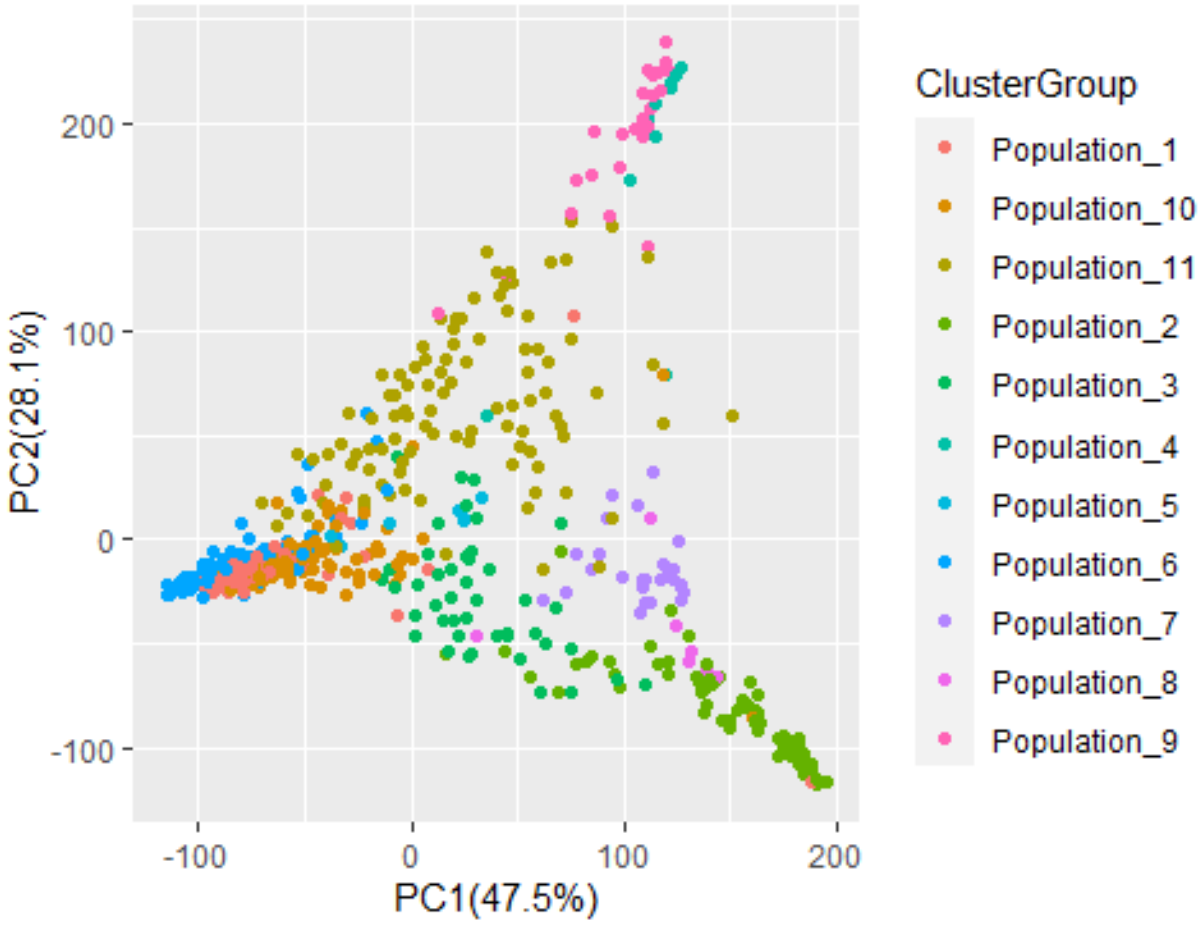


Figure 3.3. Principal Component Analysis (PCA) of accessions. Proportion of variance explained by each PC is indicated in parenthesis. Dots with different color indicate cluster groups (sub-populations)

Sub-population 2 contains mostly early flowering and short statured accessions than both sub-populations 6 and 9 whereas the latter two sub-populations are highly similar in terms of flowering time and plant height suggesting other traits also contribute to shaping the diversity. Moreover, grain yield and yield component data collected for the accessions previously [163] indicate that the three sub-populations are highly variable for grain number and grain yield per panicle as well as thousand grain weight. Sub-populations 2, 6 and 9 showed an average seed number per panicle of 1662, 2399 and 3322, respectively. Similarly, grain yield per panicle for the three sub-populations, respectively, were 45.4, 91.9 and 75.9 g. Sub-population 6 had the highest average thousand grain weight of 30.3 g followed by sub-population 2 with 26.5 g and sub-population 9 had the least of the three with 22.9 g. In terms of reaction to grain mold, sub-

population 9 showed better resistance than both sub-populations. These observations indicate that the sub-populations that shaped the PCA varies substantially in multiple traits.

3.4.5 Seed mycoflora diversity among study materials

Analyses of seed mycoflora was conducted to determine the components of the grain mold fungi at the experimental sites. The species composition of molding fungi may vary by location which may also have a bearing on the nature of host resistance. We recovered *Phoma*, *Curvularia*, *Fusarium*, *Cladosporium*, *Bipolaris*, *Alternaria*, *Colletotrichum* and *Rhizopus* species from both resistant and susceptible accessions. Except *Bipolaris* and *Rhizopus* which were only recovered from seed samples harvested from Jimma, all species were recovered from seed samples from both Jimma and Bako. *Phoma* was the most frequently recovered fungal genus at both locations followed by *Alternaria*, *Fusarium* and *Curvularia* species. Moreover, the identified *Fusarium* species were categorized into at least three morphologically distinct types.

3.4.6 Validation of GWAS results using glume color and plant height as control

Pericarp color and plant height are well studied in sorghum and commonly used to validate GWAS results. Both pericarp and glume colors are genetically associated with the *Y1* locus [131]. Plant height is one of the most studied traits in sorghum and a number of major loci regulating plant height have been identified and their location in the sorghum genome were described [164-167]. Therefore, glume color and plant height were used as controls in our study.

We consistently detected three loci for glume color by all the five models except one was not detected by FarmCPU. One of these loci is the sorghum *Y1* locus, and the other two are located at 55.5 Mbp on chromosome 4 and 57.0 Mbp on chromosome 3. The locus at 57.0 Mbp on chromosome 3 overlapped with the pericarp color locus R [165] while the locus at 55.5 Mbp is co-localized with a previously identified locus for both pericarp color and 3-deoxyanthocyanidin biosynthesis [168]. The R locus is also considered to have similar effect on both pericarp and glume colors. In our study, both BLINK and FarmCPU detected S1_68388126 as the top SNP, which is tightly linked to the R2R3 *MYB* gene paralogs at the *Y1* locus. The SNPs S4_55546047 and S3_57065511 were detected by BLINK as the second and third highly significant SNPs, respectively, which are located in the other two loci. S4_55546047, S3_57039969 and

S1_68388126 were detected as first, second and third highly significant SNPs, respectively by both MLM and CMLM. We detected S3_57065511 as a top SNP using GLM and this model also detected the SNPs located in the other two loci. S4_55546047 was detected as the least significant SNP among a total of 14 significant SNPs detected by FarmCPU while the locus at 57.0 Mbp on chromosome 3 was not detected by this model. While all the four models (GLM, MLM, CMLM and BLINK) detected only the three loci, FarmCPU has detected 12 additional significant SNPs which are all located at different loci. The identification of the three known loci, not only validates our SNP data but also confirmed that the same genomic regions regulate both kernel and glume color although there may be loci specifically regulating pigmentation of either of the two tissues.

A previous study using a large collection of Ethiopian sorghum accessions reported only two of the previously known loci on chromosome 7 regulating plant height were significant [155]. These are the QTLs at 59.6 Mbp, which correspond to the dw3 (Sobic.007G163800), and a nearby QTL at 56.4 Mbp that corresponds to qHT7.1 [167]. In this study, we detected the locus at 56.4 Mbp but not that at 59.6 Mbp. Moreover, using BLINK and GLM we detected another previously mapped locus at 62.5 Mbp on chromosome 5 [169]. The significant SNP detected by both models is S5_62503828 whereas GLM detected three more closely located SNPs (S5_62503839, S5_62503768 and S5_62503770). Only BLINK and GLM were able to detect the known locus at 56.4 Mbp while all the models detected a new locus at 40.9 Mbp on chromosome 8. All the five models detected S8_40984322 as a top SNP in this locus indicating a major locus associated with plant height.

3.4.7 GWAS for grain mold resistance

Genome-wide association study for grain mold resistance detected a total of 62 loci, distributed all over sorghum chromosomes except chromosome 9. Four of these loci were consistently detected across environments and rating methods (Table 3.2) while the remaining were specific to location or rating method. Manhattan plots presented in the following sections are based on analysis of data pooled over years in each location whereas, details of significant SNPs based on the pooled data and individual year outputs are provided as supplementary materials.

Table 3.2. Summary of significant SNPs consistently associated with grain mold resistance in sorghum accessions

Trait	Location	Model	SNP	Chr.	P.value	Maf	Effect	Locus
TGMR	Jimma	FarmCPU	S4_62316425	4	2.92E-14	0.44	0.34	*Tan1*
TGMR	Jimma	BLINK	S4_62316425	4	7.37E-13	0.44	NA	*Tan1*
PGMR	Jimma	FarmCPU	S4_62316425	4	1.39E-12	0.44	0.35	*Tan1*
TGMR	Jimma	GLM	S1_51860558	1	8.30E-12	0.48	-0.30	*SbGM1.1*
PGMR	Bako	BLINK	S3_15689447	3	3.50E-11	0.31	NA	*SbGM3.1*
FGMR	Bako	GLM	S1_51860580	1	5.96E-11	0.36	0.26	*SbGM1.1*
PGMR	Jimma	BLINK	S4_62316425	4	1.32E-10	0.44	NA	*Tan1*
PGMR	Jimma	BLINK	S3_15689578	3	1.44E-10	0.41	NA	*SbGM3.1*
PGMR	Jimma	GLM	S3_15689578	3	2.01E-10	0.41	0.31	*SbGM3.1*
FGMR	Bako	GLM	S1_51860558	1	3.20E-10	0.48	-0.22	*SbGM1.1*
PGMR	Jimma	GLM	S3_15689447	3	7.20E-10	0.31	0.32	*SbGM3.1*
PGMR	Jimma	FarmCPU	S3_15689578	3	1.28E-09	0.41	0.20	*SbGM3.1*
PGMR	Jimma	GLM	S1_51860558	1	2.91E-09	0.48	-0.31	*SbGM1.1*
FGMR	Bako	GLM	S1_68362849	1	3.36E-09	0.42	0.22	*Y1*
PGMR	Bako	FarmCPU	S3_15689447	3	3.62E-09	0.31	0.23	*SbGM3.1*
TGMR	Jimma	GLM	S4_62316425	4	6.57E-09	0.44	0.39	*Tan1*
PGMR	Bako	GLM	S3_15689447	3	9.57E-09	0.31	0.30	*SbGM3.1*
TGMR	Jimma	GLM	S1_51860580	1	2.10E-08	0.36	0.28	*SbGM1.1*
TGMR	Jimma	GLM	S3_15689578	3	3.50E-08	0.41	0.23	*SbGM3.1*
PGMR	Bako	GLM	S1_51860558	1	6.50E-08	0.48	-0.28	*SbGM1.1*
TGMR	Jimma	CMLM	S4_62316425	4	9.97E-08	0.44	0.38	*Tan1*
TGMR	Jimma	MLM	S4_62316425	4	1.11E-07	0.44	0.38	*Tan1*
TGMR	Jimma	BLINK	S1_51860558	1	1.18E-07	0.48	NA	*SbGM1.1*
PGMR	Bako	GLM	S1_51860580	1	1.18E-07	0.36	0.31	*SbGM1.1*
FGMR	Bako	CMLM	S1_68362849	1	2.25E-07	0.42	0.20	*Y1*
PGMR	Bako	GLM	S3_15689578	3	4.81E-07	0.41	0.25	*SbGM3.1*
PGMR	Bako	BLINK	S1_51860558	1	4.91E-07	0.48	NA	*SbGM1.1*
FGMR	Bako	MLM	S1_68362849	1	9.23E-07	0.42	0.19	*Y1*
FGMR	Bako	FarmCPU	S1_51860580	1	1.13E-06	0.36	0.14	*SbGM1.1*
FGMR	Bako	MLM	S1_51860580	1	1.70E-06	0.36	0.24	*SbGM1.1*

FGMR, Field (plot) grain mold rating; PGMR, Panicle grain mold rating; TGMR, Threshed grain mold rating; MAF, Minor allele frequency

At Jimma, using GLM, we detected a major peak for TGMR at 51.8 Mbp on chromosome 1 (Figure 3.4A). The top SNP (S1_51860558) at this locus had allelic effect of -0.30. Based on GLM, we also detected a known SNP (S4_62316425) located in the sorghum *TAN1* gene. Using BLINK and FarmCPU, S4_62316425 was detected as a major and highly significant SNP

associated with TGMR (Figure 3.4B &C). The SNP S1_51860558 was also detected for TGMR using BLINK. Moreover, S4_62316425 was detected by both CMLM and MLM although the significance levels obtained using these two models were not as high as that of BLINK and FarmCPU. S4_62316425 had allelic effect ranging from 0.37 to 0.39 depending on the models used. One more significant SNP, S5_67744208, is detected for TGMR using both GLM and FarmCPU while among the five models, FarmCPU seems to detect more significant SNPs than others.

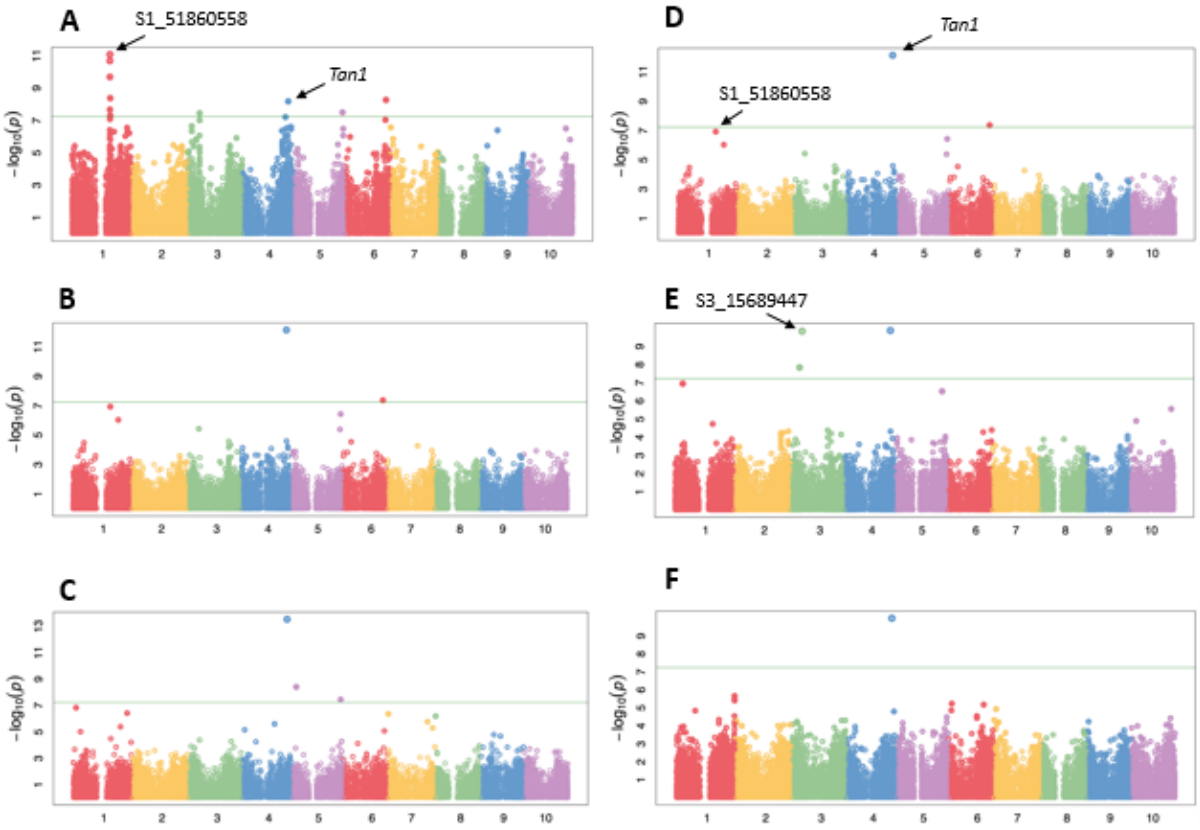


Figure 3.4. Genome-wide association analysis of grain mold resistance at Jimma comparing different models (A-C) and rating methods (D-F). Manhattan plots of association mapping for TGMR using GLM (A), BLINK (B), FarmCPU (C). Manhattan plots of association mapping using BLINK for TGMR (D), PGMR (E) and FGMR (F)

Using BLINK, we also looked at how the different mold rating methods, intact panicles in field plot (FGMR), excised panicle (PGMR) or threshed grain mold rating (TGMR) affect detection of significant loci. S4_62316425 remained highly significant for both PGMR and TGMR

but this SNP was not significant for FGMR but a closely located SNP S4_62405510 was significant for FGMR (Figure 3.4D-F). S1_51860558 was detected for TGMR but not for PGMR and FGMR. Using BLINK, another peak was detected for PGMR at 15.6 Mbp on chromosome 3. The top single SNP at this locus is S3_15689578 which was equally significant to that of S4_62316425 (Figure 3.4E). All of the detected SNPs based on the two years pooled data were also significant when analyzed for individual years.

At Bako, using GLM, we detected a peak at 51.8 Mbp that was associated with FGMR (Figure 3.5A), the same peak was detected based on TGMR at Jimma. The top SNP at this locus was S1_51860580. Another peak was detected for FGMR at 64.2 Mbp on chromosome 4 using all the three models GLM, MLM and CMLM. Significant SNPs at this locus were S4_64224360 and S4_64224349. Another significant SNPs for FGMR were detected in the sorghum *Y1* locus using GLM, CMLM and MLM (Figure 3.5A & B). Moreover, the SNP S1_51860580 was significant for FGMR using FarmCPU while other significant SNPs were detected using BLINK that were not detected by the other models.

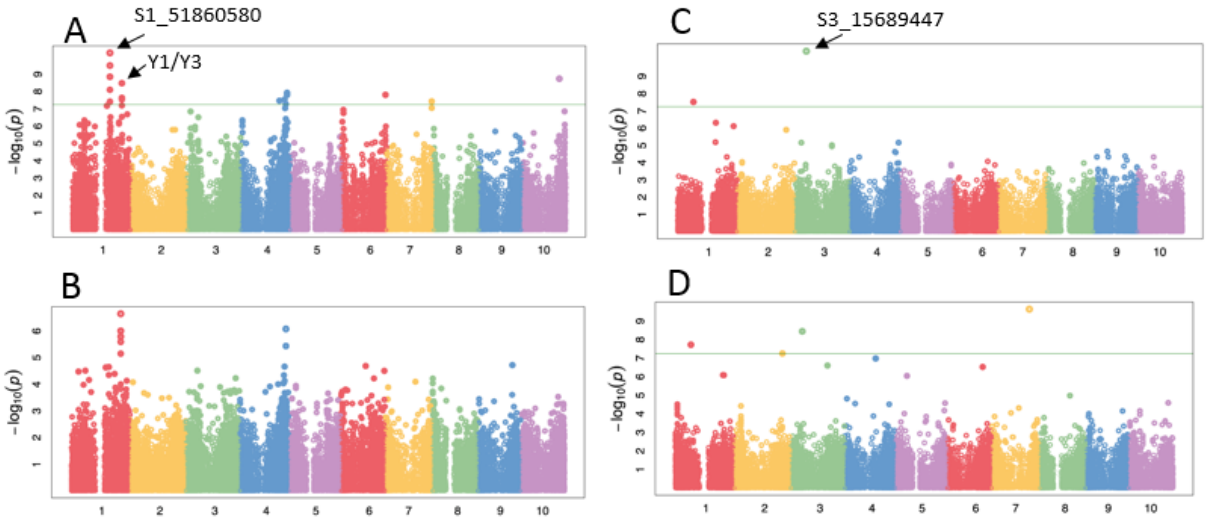


Figure 3.5. GWAS for grain mold resistance in sorghum landraces at Bako. A and B) Manhattan plot of GWAS for plot-based mold rating using GLM (A), and CMLM (B) models. C and D Manhattan plot of GWAS for panicle mold rating using BLINK (C) and FarmCPU (D) models

Two closely linked significant SNPs (S6_60023010 and S6_60022999) at 60 Mbp on chromosome 6 were detected for FGMR using the three models GLM (Figure 3.5A), FarmCPU

and BLINK. Similarly, significant SNPs associated with PGMR were detected at Bako some of which were also significant at Jimma. S3_15689447 was detected as a top SNP associated with PGMR at Bako using both BLINK and FarmCPU (Figure 3.5C & D) while the same SNP and the tightly linked SNP S3_15689578 were found significant using GLM. We also detected SNP S1_21499669 to be significantly associated with PGMR using both BLINK and FarmCPU (Figure 3.5C & D). Similar to the data from Jimma, all the significant SNPs described here, which are based on data pooled over the two years, were significant based on individual year data.

3.4.8 GWAS using non-pigmented accessions

Using 373 accessions with non-pigmented grains, we detected a total of 41 loci distributed across all the sorghum chromosomes except chromosome 8. Two of these loci located as major peaks at 51.8 on chromosome 1 and 15.6 Mbp on chromosome 3, were consistently detected across environments (Figure 3.6). It is interesting that these two loci were consistently significant across locations and models when the entire 635 accessions were used. At Jimma, all the three models BLINK, FarmCPU and GLM detected S1_51860580 as the most significant SNP to associate with TGMR (Figure 3.6A & B). Using GLM, we also detected other SNPs that associate with TGMR (S1_51860558, S1_51839599, S1_51860544, S1_51860543 and S1_51839525) tightly linked to the top SNP. Both FarmCPU and GLM detected S3_15689578 while BLINK and GLM also detected S1_74846159 for TGMR. Moreover, using FarmCPU other significant SNPs were identified that were not found significant by other models. At Bako, the two peaks at 51.8 and 15.6 Mbp were detected as the top loci for PGMR using BLINK (Figure 3.6C). The most significant SNPs in the two loci were S1_51860580 and S3_15689447. Using GLM both S1_51860580 and S3_15689447 were detected as top SNPs for PGMR (Figure 3.6D). S1_51860580 was also detected using FarmCPU while a SNP in different locus, S7_59807467, was detected for PGMR using both FarmCPU and GLM. Another SNP, S4_60400448, was detected for PGMR using both BLINK and GLM.

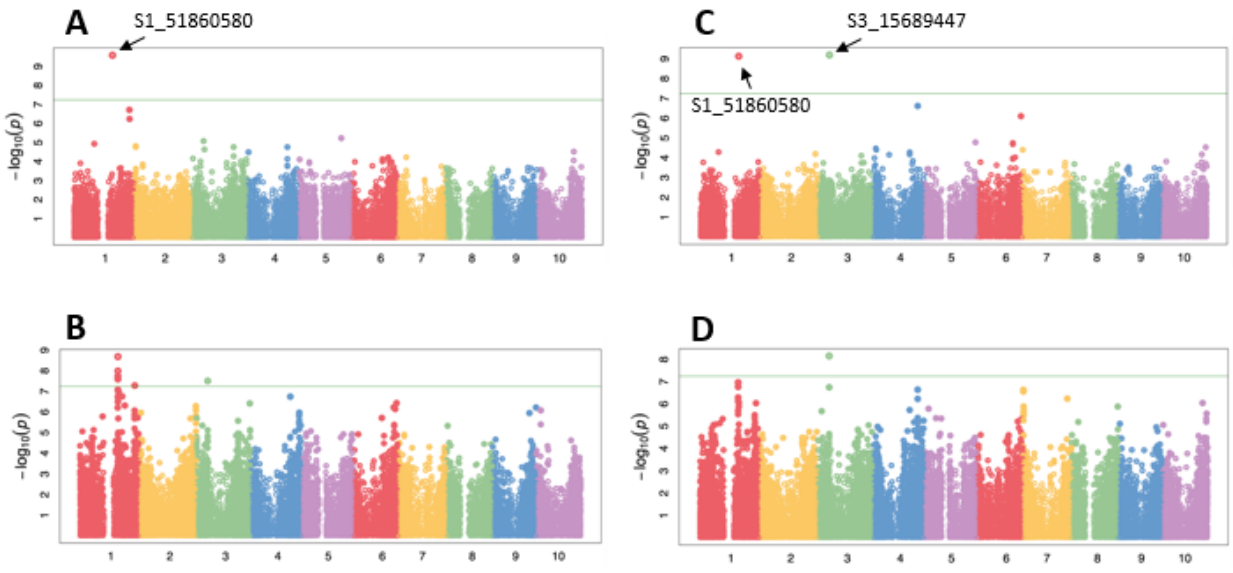


Figure 3.6. Grain mold resistance loci identified through GWAS using sorghum accessions with non-brown and red colored kernels. A and B) Manhattan plot of GWAS for threshed grain mold rating at Jimma. C and D) Manhattan plot of GWAS for panicle mold rating at Bako

3.4.9 Candidate genes in the newly detected grain mold resistance loci

Overall, we detected two novel loci consistently associated with grain mold resistance in sorghum across all environments and regardless of the GWAS models used. These two new loci are located at 51.8 Mbp on chromosome 1 and 15.6 Mbp on chromosome 3, designated here as *SbGM1.1*. and *SbGM3.1*. We also detected S4_62316425, which is located in the sorghum *TAN1* gene [170] to be significantly associated with grain mold resistance. Although the *TAN1* gene is known to regulate the biosynthesis of tannin which is presumed to play a role in grain mold resistance, tannin has not been directly demonstrated for its significant association with grain mold resistance. Moreover, the current GWAS analysis detected the sorghum *Y1* locus [114, 131] and a few other new loci. We identified candidate genes in both *SbGM1.1* and *SbGM3.1* and other loci using the reference genome of *Sorghum bicolor* v3.1.1 [120] accessible on Phytozome [171].

3.4.9.1 SbGM1.1

The significant SNPs in *SbGM1.1* locus on chromosome 1 are located in a genomic region with no annotation for predicted genes. The nearest annotated gene (Sobic.001G270200) to the most significant SNPs (S1_51860558 and S1_51860580) is 27 kb away while the other annotated genes are located 106 kb (Sobic.001G269900) and 166 kb (Sobic.001G270301) away from the

most significant SNP. We conducted further search for candidate genes including sequence tags closely linked to the significant SNPs. An expressed sequence tag (EST) of 8646 bp sequence assembled by PASA (Program to Assemble Spliced Alignments) [172] located on Chr01:51862418..51871063 (- strand) in Phytozome (Figure 3.7A-C) was identified. Interestingly, BLAST search using the EST sequence revealed significant alignment (94.8 % identity) to the 22 kDa *KAFIRIN* cluster of *Sorghum bicolor* (accession AF061282.1).

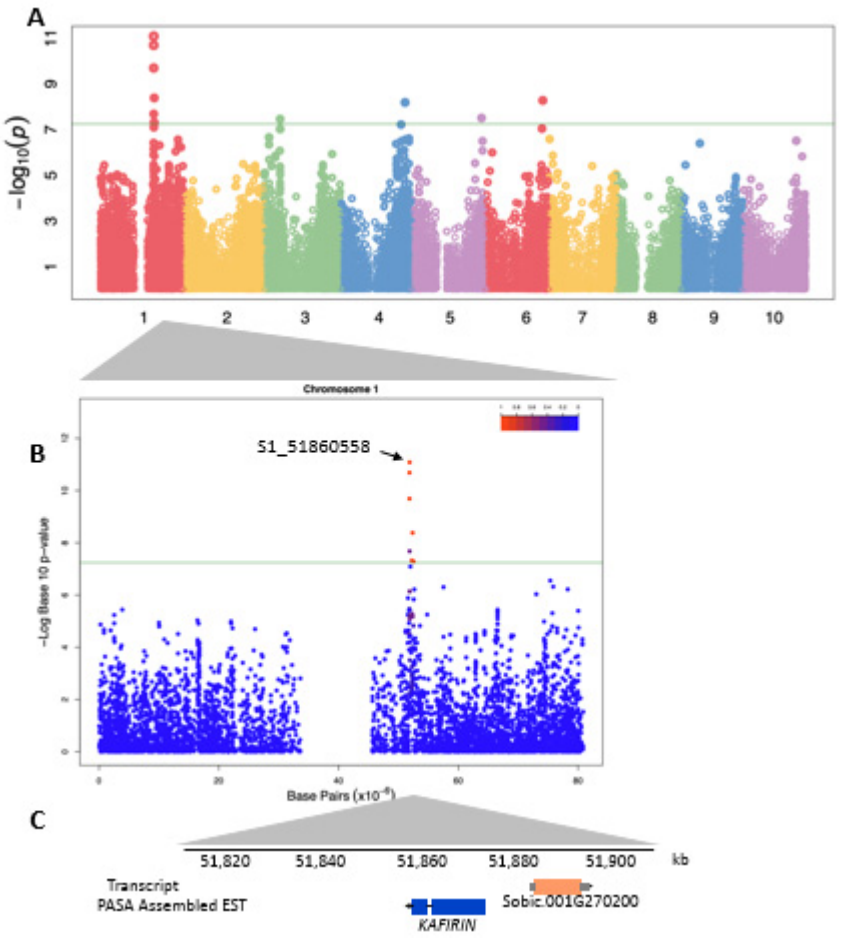


Figure 3.7. Genomic organization of a putative *KAFIRIN* gene containing locus associated with grain mold resistance. A) Genome wise manhattan plot of GWAS for TGMR using GLM. B) Chromosome wise manhattan plot indicating the genomic position of the *KAFIRIN* gene on chromosome 1. C) Details of the significant locus at 51.8 Mbp indicating the position of the *KAFIRIN* gene and a flanking Sobic.001G270200 (*cytochrome P450*) gene in relation to the significant SNPs

Kafirins are the predominant seed storage proteins in sorghum [173] associated with seed texture. The fact that this locus was significant for grain mold resistance in the non- pigmented subset of accessions, strengthens our observation that implicate seed texture-based resistance, a resistance mechanism independent of grain flavonoids. The next nearest annotated gene, Sobic.001G270200, encodes *cytochrome P450 CYP2* subfamily protein. Since, this gene was the only annotated gene in the locus close to the significant SNP, initially we considered it as our first candidate and did extensive literature search for any function in plant disease resistance. However, no evidence of defense function was found for this gene. Similarly, besides being located far from the significant SNPs (at least 106 kb), the other annotated genes were not supported by direct evidences and therefore not considered as candidates. While both *P450 CYP2* and *KAFIRIN* are potential candidate genes, in the absence of a direct evidence, the *KAFIRIN* gene is the more likely candidate for grain mold resistance. The putative storage protein kafirin supported by the EST data may have a better association, with grain mold resistance in sorghum due to close proximity to the significant SNP. Kafirins are associated with grain texture, a trait traditionally associated with resistance to grain mold.

3.4.9.2 SbGM3.1

The most significant SNP in *SbGM3.1* locus, S3_15689447, is located on the promotor region of the Sobic.003G149100 gene that encodes a putative LATE EMBRYOGENESIS ABUNDANT 3 (LEA3) family protein (Figure 3.8A-C). The other significant SNP, S3_15689578, in this locus is also located on the promoter region of the same gene. *LEA3* is a likely candidate gene based on published reports in maize and other plant species that implicate LEA3 proteins in abiotic and biotic stress responses [174, 175]. The other annotated genes close to the significant SNPs are Sobic.003G149000 (*BIDIRECTIONAL SUGAR TRANSPORTER SWEET16-RELATED*), Sobic.003G148950 (*PROTEINASE INHIBITOR I46, LEECH METALLOCARBOXYPEPTIDASE INHIBITOR*) and Sobic.003G149200, a likely duplicate of the first candidate (Sobic.003G149100). Sobic.003G149000 gene encodes a putative sugar transporter (SWEET) and is only 7 kb away from *LEA3* (Sobic.003G149100) gene which was the best candidate. *SWEET* genes have been reported for their defense functions [176-179] although their mode of action is not well understood.

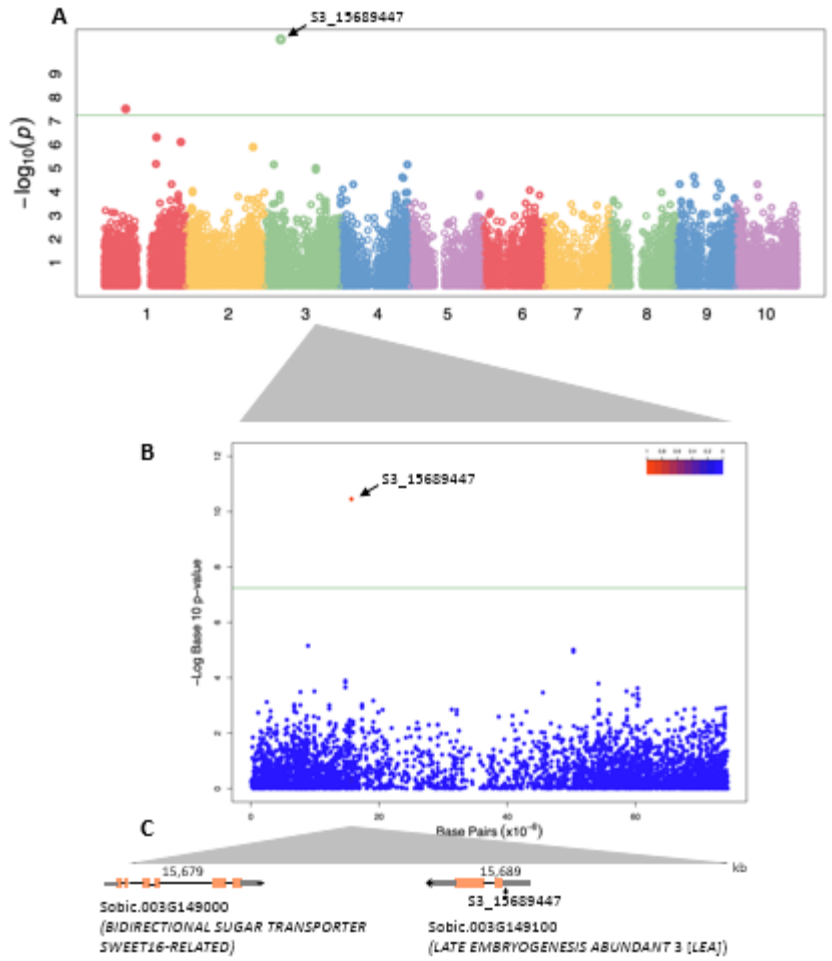


Figure 3.8. Genomic organization of a candidate *LATE EMBRYOGENESIS ABUNDANT* 3 (*LEA3*) gene associated with grain mold resistance. A) Genome wise manhattan plot of GWAS for PGMR using BLINK. B) Chromosome wise manhattan plot indicating the genomic position of *LEA3* gene on chromosome 3. C) Details of the significant locus at 15.689 kb Mbp indicating the position of candidate genes in relation to the significant SNP (S3_15689447). The exact position of the significant SNP is pointed by an arrow on promoter region of *LEA3*

3.5 Discussion

This study presents observations from a genome wide association study of a large collection of Ethiopian sorghum landraces which enabled the identification of novel loci associated with grain mold resistance, a genetically complex trait. Our study employed different phenotyping approaches, diverse germplasm and genome wide association analyses methods, which uncovered loci, associated with grain mold resistance. Sequence polymorphisms in the *TAN1*, *KAFIRIN*, and *LEA3* genes were associated with differences in grain mold resistance. Although tannins have been cited for resistance to grain diseases, we found no previous direct genetic data for the link between

tannin and grain mold resistance. Tannins were recently implicated in tolerance to bird attack [18]. The link between sorghum seed protein, kafirin and grain mold resistance is likely attributed to their contributions to endosperm texture. Kafirins are prolamin storage proteins found in sorghum endosperm and account for 70% of total proteins in the grain [173]. LEA proteins, also a family of proteins that preferentially accumulate in seeds, are associated with tolerance to desiccation in plant and animal cells [180]. The genetic association between grain mold resistance and seed proteins presents challenges for the simultaneous improvement of grain nutritional traits and grain mold resistance. Tannins and kafirins are considered nutritionally undesirable due to interference with digestibility [181] while their importance is also likely linked to pathogen and pest tolerance. Moreover, tannins are also known to have benefits to human health because of their antioxidant properties and reducing obesity [2]. Sorghum foods with high kafirin and tannin could be used as means to lower calorie intake and reduce obesity [181]. A deeper understanding of the functional link between the various grain traits including mold resistance may aid in designing better strategies for trait improvement. Besides defining new genetic loci, this study provided additional data and observations for why such loci remained undetected in previous studies. Our study reveals that the current set of landrace accessions harbors excellent source of genes for grain mold resistance and the identified loci provide new insights into understanding the genetic basis of resistance to the disease. Each of the above points are discussed in the following sections.

3.5.1 Impact of phenotyping approaches and model selection on detection of loci

Grain mold in sorghum is very complex and resistance to the disease is quantitative. Disease rating methods and accuracy of the scores can affect detection of loci associated with resistance to the disease. We followed a three-stage disease rating (see methods) for host grain mold responses that was based on field plot scoring (FGMR), laboratory scoring of excised panicles (PGMR) and threshed grain (TGMR). Heritability was improved with excised panicle and threshed grain ratings compared to plot-based ratings. This may be because excised panicle and threshed grains could easily and more accurately be rated in laboratory. Detection of loci associated with the disease varied depending on the rating methods although some loci were consistently significant across all methods. This suggests that a combination of rating methods or those which show better heritability may improve detection of loci associated with grain mold resistance.

Likewise, GWAS models have evolved over years improving the statistical power and also reducing computational time. We used five models (GLM, MLM, CMLM, FarmCPU and BLINK) to conduct GWAS. GLM, FarmCPU and BLINK commonly detected most of the known or new SNPs identified in our study while only a few of the detected SNPs were found significant using MLM and CMLM. Some of the significant SNPs ranked differently or even become non-significant with changing models indicating some loci with significant contribution to a trait of interest may be overlooked, remain undetected or become only marginally significant with some of the GWAS models. Therefore, the use of multiple models including those with limited fitting might be crucial while conducting GWAS.

3.5.2 Identification of novel grain mold resistance loci

Concentrations of grain flavonoids 3-deoxyanthocyanidins, flavan-4-ols, and tannins and grain physical characteristics have all been associated with grain mold resistance in sorghum [70-72, 154]. However, besides the nutritional drawbacks associated with these metabolites in the grain and the grain physical structures, the underlying genetics is not yet known, and new mechanisms have not been identified because of complexity of the trait and perhaps lack of appropriate mapping populations. Our major aim in conducing the current GWAS was to identify new mechanisms of resistance independent of the grain flavonoids although we were also interested to understand the role and regulation of 3-deoxyanthocyanidins, tannins and other known resistance mechanisms. Besides detecting genomic regions carrying key flavonoid regulatory genes such as the sorghum *TAN1* and *MYB* genes, we detected two novel loci with major effect on grain mold resistance in sorghum. Interestingly, both newly detected loci, *SbGM1.1* and *SbGM3.1,* contain candidate genes encoding major seed proteins. *SbGM1.1* locus harbors an expressed sequence tag (EST) highly similar to the seed storage protein kafirin which is associated with endosperm texture in sorghum [182]. Remarkably, our search for previously detected QTL in the locus using the sorghum QTL atlas [183] reveals that *SbGM1.1* is co-localized with a protein digestibility QTL which was identified using P850029/Sureno bi-parental lines [184]. However, it was not indicated whether the protein digestibility QTL harbors any *KAFIRIN* gene. The protein digestibility QTL is located about 200 kb from the significant SNPs we identified. Sureno is a grain mold resistant line [185] commonly used as a check in grain mold resistance studies. *KAFIRIN* gene organizations are very complex and are often found as clusters of tandem repeats; likely interrupted by transposable

elements and contain pseudogenes [186]. A 10-copy tandem repeat *α-KAFIRIN* genes is found in a single locus on chromosome 10 [186] while a 22 kDa *α-KAFIRIN* gene copies are located in cluster on chromosome 5 [187], suggesting the new EST within *SbGM1.1* could be an unknown *KAFIRIN* gene in sorghum. In addition, *KAFIRIN* genes in sorghum, foxtail millet and maize are flanked by *CYTOCHROME P450* genes [188] indicating the *CYTOCHROME P450* gene (Sobic.001G270200) located close to the identified EST supports this evolutionarily conserved genomic organization. Kafirins are grouped into four subclasses (α-, β-, γ- and δ-kafirins) [189, 190]. The *α*-kafirin subclass is the predominant (70-80%) of the kafirins [173]. Kafirins reduce sorghum protein digestibility [191] and hence affect nutritive value of sorghum grain. Mutations in *KAFIRIN* gene results in high value food-trait in sorghum [182]. In contrary, kafirins could confer resistance to grain mold through seed (endosperm) hardness and other mechanisms.

The *SbGM3.1* locus includes the Sobic.003G149100 gene that encodes a LATE EMBRYOGENESIS ABUNDANT 3 (LEA3) family protein. LEA proteins confer tolerance to drought and salt stress [192]. They display increased accumulation during late stage of seed development [174]. LEA3 proteins in maize are involved in both biotic and abiotic stress tolerance [174], resistance to aflatoxin in maize kernels [193, 194] and resistance to *Fusarium* head blight in wheat [195] and barley [196]. However, the evidences for resistance to aflatoxin contamination or resistance to *Fusarium* head blight in both wheat and barley were based on proteomic or transcriptional profiling of contrasting lines and direct evidences are not available. A direct evidence for LEA proteins in defense comes from overexpression of *ZmLEA3* in transgenic tobacco which increased hypersensitive cell death and enhanced expression of *PR1a*, *PR2* and *PR4* genes [174]. However, additional evidence for the involvement of LEA proteins in hypersensitive response and pathogen resistance are not available. The maize LEA3 proteins may contribute to biotic and abiotic stress tolerance by protecting protein structures [174]. A number of other studies also indicated that LEA proteins prevent protein aggregation resulting from desiccation or osmotic stress [197-199]. The same mechanism underlying tolerance to abiotic stress may account for grain mold resistance by protecting seed proteins and other compounds, which are associated with seed texture or chemicals. The other candidate gene (Sobic.003G149000) belongs to a group of *SWEET* genes, which have been studied for their defense functions [176-179], perhaps indirectly through regulation of sugar transport. Much of their role in defense seems mostly against the biotrophs or hemibiotrophs [177] and bacterial pathogens [179] which relay on nutrient supply from the host.

Sobic.003G149000 is related to *SWEET 16* groups which in *Arabidopsis* is reported as a vacuole-localized carrier involved in sugar transport and associated with germination, growth and stress tolerance [200], but their role in defense is unclear.

3.5.3 Association of sorghum *TAN1* and *Y1* loci to grain mold resistance

The sorghum *Y1* and *TAN1* genes regulate kernel color and 3-deoxyanthocyanids [131] and biosynthesis of tannins [170], respectively. Although both genes are supposed to contribute to grain mold resistance, the *Y1* locus was rarely detected by genome wide association analyses [76, 114] while the *TAN1* locus has never been reported as a significant locus for grain mold resistance. In the current study, we consistently detected the *TAN1* locus to be significantly associated with grain mold resistance, while *Y1* locus was detected based on grain mold data from one of the experimental sites. The known G-to-T transition SNP, S4_62316425, inside the coding sequence of *TAN1* gene, was detected by all the models used in our study. In our set of accessions, G was a minor allele with a frequency of 0.44. S4_62316425 is in perfect linkage disequilibrium with a 1 bp G deletion in the coding region which causes frameshift and premature stop codon leading to a non-functional allele [92, 170]. A recent GWAS for grain mold resistance using SAP [76] indicated that even if the functional allele is present at high frequency (0.79), most of the accessions were found susceptible to grain mold and the *TAN1* locus was also not significantly associated with grain mold resistance. The fact that many of the previous studies did not detect this locus could be associated with the complexity of the phenotype which requires a combination of appropriate population and a more systematic and accurate rating of grain mold. In the current study, besides the use of large and diverse natural variants, we ensured an accurate recording of grain mold through a three-stage rating and data quality control through replicated checks and heritability estimation. However, as important as detecting this key locus, which is believed to play role in grain mold resistance, is whether tannins or other flavonoids such as the 3-deoxyanthocyanids directly contribute to grain mold resistance. Tannins and 3-deoxyanthocyanids are synthesized through the same pathway [201, 202]. The widely studied *TAN1*'s orthologue in *Arabidopsis* (*TTG1*) has been shown to regulate a number of developmental and biochemical pathways [137, 203-205]. The sorghum *TAN1* is involved in biosynthesis of 3-deoxyanthocyanids [138], therefore, the resistance to grain mold mediated by *TAN1* could be due to either tannins or 3-deoxyanthocyanids. Any resistance in accessions producing tannin could also be due to other

associated factors including 3-deoxyanthocyanids. Those accessions that are able to synthesize tannins may also have genes that are required in flavonoid biosynthesis pathway, and thus are able to synthesize 3-deoxyanthocyanids and other flavonoids. Moreover, there was no evidence if tannins have any fungi-toxic properties while the 3-deoxyanthocyanids are phytoalexins that accumulate in response to pathogen infection and are known to be toxic to fungi [90, 134, 206, 207]. However, it is interesting that tannins could irreversibly and strongly bind with kafirin [181, 208]. Therefore, tannins may indirectly play role in grain mold resistance through association with other factors including seed proteins.

The other related locus, the sorghum *Y1 (MYB*) is known to regulate the biosynthesis of 3-deoxyanthocyanidins [89, 131]. 3-deoxyanthocyanidins enhance resistance against anthracnose leaf blight in sorghum [88] and maize [89]. Based on their fungi-toxic properties, accumulation in bran of sorghum grain [209] and evidences from studies conducted in leaf tissues, 3-deoxyanthocyanidins are good candidates for grain mold resistance. However, direct evidence indicating 3-deoxyanthocyanidins' role in grain mold resistance are still not available. The fact that, *Y1 (MYB)* locus has been rarely detected suggests regulations of such loci are more complex because the biosynthesis of 3-deoxyanthocyanidins is pathogen inducible and may involve additional regulatory components. Some of those regulatory mechanisms appears to be receptors that recognize pathogen structures, *MITOGEN-ACTIVATED PROTEIN KINASES (MAPK*) cascades, phosphorylation of downstream proteins and protein complexes and ultimately initiating expression of target genes involved in the biosynthesis of 3-deoxyanthocyanidins and other related flavonoids. Much of this is unknown in sorghum but such signaling components leading to phytoalexin biosynthesis are described in *Arabidopsis* [210, 211] and rice [212]. A recent study indicated that *MPK4* phosphorylation of the *R2R3 MYB* transcription factor, *MYB75/PAP1,* increases its stability and is essential for light-induced anthocyanin accumulation [213].

3.6 Conclusion

Through a combined use of large and diverse set of natural variants of sorghum from the center of origin and diversity of the crop and with enhanced accuracy of grain mold rating, we generated a multi-environment and comprehensive grain mold resistance data for 635 accessions. A large number of accessions with high level of grain mold resistance were identified that will be useful as source of genes for resistance breeding. We identified new and previously identified

candidate loci associated with grain mold resistance. Two new loci (*SbGM1.1* and *SbGM3.1*) associated with seed proteins and abiotic and biotic stress tolerance in related crop species including maize were strongly associated with grain mold resistance. With additional validation, these loci can be exploited for grain mold resistance or improving grain quality through marker assisted selection. The fact that the candidate genes in both of the newly identified loci encode seed proteins suggests importance of seed traits and other seed based macromolecules in grain mold resistance and their application for breeding. Moreover, despite a widely held assumption that tannins contribute to grain mold resistance, the role of tannins has not been directly demonstrated. Our results demonstrated that sequence variations at the *TAN1* gene is significantly and consistently associated with grain mold resistance. Overall, our observations suggest that better understanding of the genetics of complex traits can be achieved through a combination of efficient phenotyping to enhance sensitivity of GWAS, use of large and diverse variants and appropriate genomic analysis tools. For genetically complex traits such as grain mold, ensuring data quality is the first and critical step but may not be sufficient without implementation of appropriate analytical models. The fast-evolving genomic analysis models can improve statistical power and save time, but our results suggest that detection of some loci can be overlooked perhaps due to over fitting by advanced analysis models. Broadly, our study highlights the critical role of seed proteins that contribute to the physical and chemical properties of the grain are significant determinants of resistance to grain mold. These relationships may be targeted for improvement through genetic approaches.

CHAPTER 4. TRANSCRIPTOME ANALYSIS OF EARLY STAGES OF SORGHUM GRAIN MOLD DISEASE REVEALS DEFENSE REGULATORS AND METABOLIC PATHWAYS ASSOCIATED WITH RESISTANCE

A version of this chapter was previously published by *BMC Genomics* [214], https://doi.org/10.1186/s12864-021-07609-y

4.1 Abstract

To understand the genetic, molecular and biochemical components of grain mold resistance, transcriptome profiles of the developing grain of resistant and susceptible sorghum genotypes were studied. The developing kernels of grain mold resistant RTx2911 and susceptible RTx430 sorghum genotypes were inoculated with a mixture of fungal pathogens mimicking the species complexity of the disease under natural infestation. Global transcriptome changes corresponding to multiple molecular and cellular processes, and biological functions including defense, secondary metabolism, and flavonoid biosynthesis were observed with differential regulation in the two genotypes. Genes encoding pattern recognition receptors (PRRs), regulators of growth and defense homeostasis, antimicrobial peptides, pathogenesis-related proteins, zein seed storage proteins, and phytoalexins showed increased expression correlating with resistance. Notably, *SbLYK5* gene encoding an orthologue of chitin PRR, defensin genes *SbDFN7.1* and *SbDFN7.2* exhibited higher expression in the resistant genotype. The data suggest a pathogen inducible defense system in the developing grain of sorghum that involves the chitin PRR, MAPKs, key transcription factors, downstream components regulating immune gene expression and accumulation of defense molecules. We propose a model through which the biosynthesis of 3-deoxyanthocynidin phytoalexins, defensins, PR proteins, other antimicrobial peptides, and defense suppressing proteins are regulated by a pathogen inducible defense system in the developing grain. The transcriptome data suggested that the developing grain shares conserved immune response mechanisms but also components uniquely enriched in the grain. Resistance was associated with increased expression of genes encoding regulatory factors, novel grain specific antimicrobial peptides including defensins and storage proteins that are potential targets for crop improvement.

4.2 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is among the world's most important cereal crops used for food, feed, and bio-fuels with unique adaptation to arid and semi-arid parts of the world. Grain mold is the most important and complex disease of sorghum caused by different pathogenic fungal species mainly in the genus *Fusarium*, but also including species in the genera *Curvularia*, *Alternaria*, *Phoma*, *Bipolaris*, *Exserohilum*, *Aspergillus*, *Colletotrichum*, and *Penicillium*. Grain mold is widespread, with major impacts on grain yield and quality especially in regions with high humidity during grain development and harvest with highly detrimental effects on grain quality due to contamination by mycotoxins. The closely related diseases include the *Fusarium* ear rot of corn and *Fusarium* head blight of wheat, which are all caused by similar group of fungal pathogens with necrotrophic mode of nutrition.

Prior studies conducted on sorghum indicate that resistance to grain mold is associated with grain flavonoids such as testa pigmentation, concentration of phenolic compounds, 3-deoxyanthocynidns, tannins and grain physical characteristics such as grain hardness [69-72]. These observations are mainly based on trait correlations but the underlying genetics of grain mold resistance remained unclear. Recent advances in sequencing technologies, substantial reduction in the cost of genotyping and availability of efficient bioinformatics tools brought new opportunities to determine the genetic control of complex phenotypes at greater depth. Global transcriptome profiling enables the identification of genome wide variations in gene expression associated with traits of interest. Transcriptional control of gene expression is a widespread regulatory event in plant responses to pathogen infection. This is particularly important since many genes associated with disease resistance are known to be transcriptionally regulated, and such an approach may identify genes mediating responses to pathogens, with a subset likely having direct contribution to resistance. Despite numerous transcriptome studies conducted in response to pathogen infection in leaf tissue, the transcriptome responses of the grain to pathogen attack have not been studied. Consequently, the processes and pathways activated or repressed during infection remain poorly understood.

Here, we conducted a comparative transcriptome analysis of grain mold resistant and susceptible sorghum genotypes RTx2911 and RTx430, respectively. The transcriptome profiling was conducted on RNA samples from developing grain (20 days after flowering) inoculated with a combination of fungal species known to constitute the grain mold fungal complex in sorghum.

Subsequently, we found differential expression of regulatory genes, signaling components associated with major immune response pathways and potential defense active molecules. Key defense mechanisms activated in the resistant genotype in response to infection were identified providing new understandings about the genetic and molecular bases of resistance to grain mold. A subset of these define novel defense strategies against fungal infection that are likely to be specific to grain tissues. Genetic and molecular dissection of defense responses in grain presents unique challenges, and our study lays the foundation for further genetic studies in grain mold resistance of sorghum.

4.3 Materials and Methods

4.3.1 Plant materials

The grain mold resistant and susceptible sorghum genotypes RTx2911 and RTx430, respectively, were used for the transcriptome analysis. RTx2911 is resistance to grain mold [215] while RTx430 is highly susceptible to the disease [74]. The resistance and susceptibility reaction of the two genotypes to a number of grain mold causing fungal species has been confirmed in a serious of greenhouse (humidity chamber) based experiments that we have conducted recently [114].

4.3.2 Inoculation of the developing sorghum grain with grain mold fungi

A mixture of spore suspension from five *Fusarium* (*F. proliferatum, F. graminearum, F. thapsinum*, *F. verticillioides* and *F. oxysporum*) and one *Alternaria* species were spray inoculated on to panicles of both RTx2911 and RTx430 at 20 days after anthesis. Inoculation and disease establishment were conducted in a humidity chamber equipped with a humidifier that has adjustable humidistat to retain humidity at required level (85-90%). Details of isolation, fungal species identification through sequencing of the ribosomal internal transcribed spacer (ITS) region of fungal DNA, multiplication and inoculation of the fungal species used in this study were described previously [114].

4.3.3 Total RNA extraction

RNA was extracted from the developing grain before and after the two genotypes were challenged by a mixture of spore suspension of equal proportions of the five *Fusarium* and an *Alternaria* species. The sampling time points were 0, 24 and 48 hours after inoculation. Total RNA was extracted as described [128] with minor modifications [114].

4.3.4 Library construction and sequencing

RNA samples were cleaned and concentrated using RNA Clean and Concentrator^{TM} -25 Kit (ZYMO RESEARCH). The quality of the RNA was evaluated using NanoDrop and Agilent Bioanalyser (RNA Plant Nano, DNA High Sensitivity and RNA Eukaryote Pico Chips). RiboZero libraries were constructed from RNA samples at 0 and 24 h time after inoculation. Then, the libraries were sequenced on an Illumina HiSeq 2500 ultra-high-throughput sequencing system with 150 bp paired-end reads. Since preliminary gene expression analysis of previously described pathogen inducible genes through real time quantitative RT-PCR (qRT-PCR) revealed induction of genes within 24 h after inoculation, the 48 h samples were not included in RNA-seq but used to study expression of individual genes through qRT-PCR.

4.3.5 Sequence data filtering and QC

Raw reads were filtered by clipping adaptors, removing low quality reads and duplicated sequences. Sequence quality was assessed by FastQC both before and after the reads were filtered for adaptors, low quality reads and duplicated sequences. Moreover, sequence GC% assessment, rRNA and phiX database matches and organism inference was conducted.

4.3.6 Differential gene expression analysis with HISAT and Cufflinks

Following data filtering and QC, the resulting high quality clean reads were used to perform differential gene and transcript expression analysis as described [216] with some modifications. The modifications include use of HISAT [217] to align the reads to the sorghum (BTx623) reference genome (PhytozomeV12: *Sorghum bicolor v3.1.1*.) instead of TopHat and all job scripts were written in python which provided more efficiency. With large number of samples, instead of writing multiple scripts which can be time consuming, a single python script was applied to

automatically generate and execute multiple job scripts for all samples. List of differentially expressed genes were obtained from the Cuffdiff analysis and genes with a log fold change (LogFC) above one (2-fold change) were used for functional classification and metabolic pathway analysis.

4.3.7 Hierarchical clustering

To assess variability among samples, hierarchical clustering analysis was performed based on Euclidean distances using WebMeV (Multiple Experiment Viewer) (http://mev.tm4.org). HeatMaps of the samples based on normalized expression values also generated using WebMeV.

4.3.8 Functional annotation and metabolic pathway analysis

Using the graphical enrichment tool ShinyGo v0.61 [218], the lists of differentially expressed genes from Cuffdiff analysis were annotated for their underlying biological process, molecular function and cellular component ontology. Metabolic pathway analysis was performed using the ShinyGo tool that also produces KEGG pathway.

4.3.9 Validation of gene expression through real time quantitative PCR

Expression of selected genes were validated via qRT-PCR as described [219]. cDNA was synthesized from 2 µg total RNA using the AMV reverse transcriptase (NEB). Quantitative PCR was performed on LightCycler® 96 system (Roche) using a SYBR Green Supermix (Bio-Rad). Sorghum *Actin* gene was used as a constitutive endogenous control. A minimum of three technical and three biological replicates were used for qRT-PCR analysis for each sample. Expression levels were calculated by the comparative C_T method [220]. Primers used for qRT-PCR analysis are listed in Table 4.1.

Table 4.1. List of primers used for qRT-PCR

Gene	Forward	Reverse
SbDFN7.1	TACCTGGGGCCCTTGGTTAT	ATGGTACTCGGCCAGTTGTG
SbDFN8.1	CGGAACCCTTGGACAAACCT	GCTTACAACATCATAACTACAGGTG
SbDFN3.2	CGTGGCCCCTTTGGAAGAAT	ATACTTCTGCCCTAGGCGTG
SbJAZ1.1	GGACAGCAAGACACCTACTCC	ATTCCCCTGAAGCAACCAGT

4.4 Results

4.4.1 RNA sequence data and mapping to the BTx623 reference genome

A total of 433,396,806 raw reads and 432,025,256 adapter trimmed and quality clipped reads were generated for the 12 RNA-seq libraries (Table 4.2). Each sample was represented by an average of 36 million high quality reads. The adaptor trimmed and quality clipped reads were mapped to the *Sorghum bicolor* reference genome [120] with 75 to 82% of the reads uniquely aligned to the reference genome in each sample.

Table 4.2. Summary statistics of RNA-seq reads generated through the HiSeq 2500 ultra-high-throughput sequencing system

Genotype	Time point (hr)	Adapter trimmed & Quality clipped reads		Mapping to the reference genome (BTx623)
		Total reads	Bases	
RTx2911	0	35,029,678	5,133,585,627	82.50%
RTx2911	0	34,631,646	4,913,623,155	82.73%
RTx2911	0	32,974,808	4,785,724,826	71.82%
RTx2911	24	38,395,698	5,647,473,903	77.70%
RTx2911	24	52,565,274	7,712,587,392	78.28%
RTx2911	24	46,082,874	6,688,093,720	74.55%
RTx430	0	40,426,498	5,894,899,883	76.98%
RTx430	0	33,298,302	4,868,287,180	75.33%
RTx430	0	23,079,728	3,369,245,420	77.87%
RTx430	24	28,741,444	4,245,933,800	77.35%
RTx430	24	33,844,156	4,979,185,158	77.93%
RTx430	24	32,955,150	4,855,366,073	78.40%

4.4.2 Overview of differential gene expression in healthy and inoculated developing grain

To identify differentially expressed genes related to grain mold resistance, transcriptomes were compared between the resistant (RTx2911) and susceptible (RTx430) genotypes. Developing grains of the two genotypes (Figure 4.1A) were inoculated with conidial suspension from a consortium of *Fusarium* and *Alternaria* species and sampled at 0 and 24 hours post inoculation (hpi) for RNA extraction which was subsequently used for RNA-seq. Hierarchical clustering analysis of expression data of all samples indicated distinct clustering by genotype, RTx2911 and

RTx430, and pathogen inoculation (Figure 4.1B). CummeRbund plots of expression level distribution (Figure 4.1C) indicated a typical expression profile while the scatter plot (Figure 4.1D) highlighted the overall similarities and outliers between the two genotypes. Transcriptome comparisons were made between the two genotypes at each time point and within each of the genotypes at the two time points.

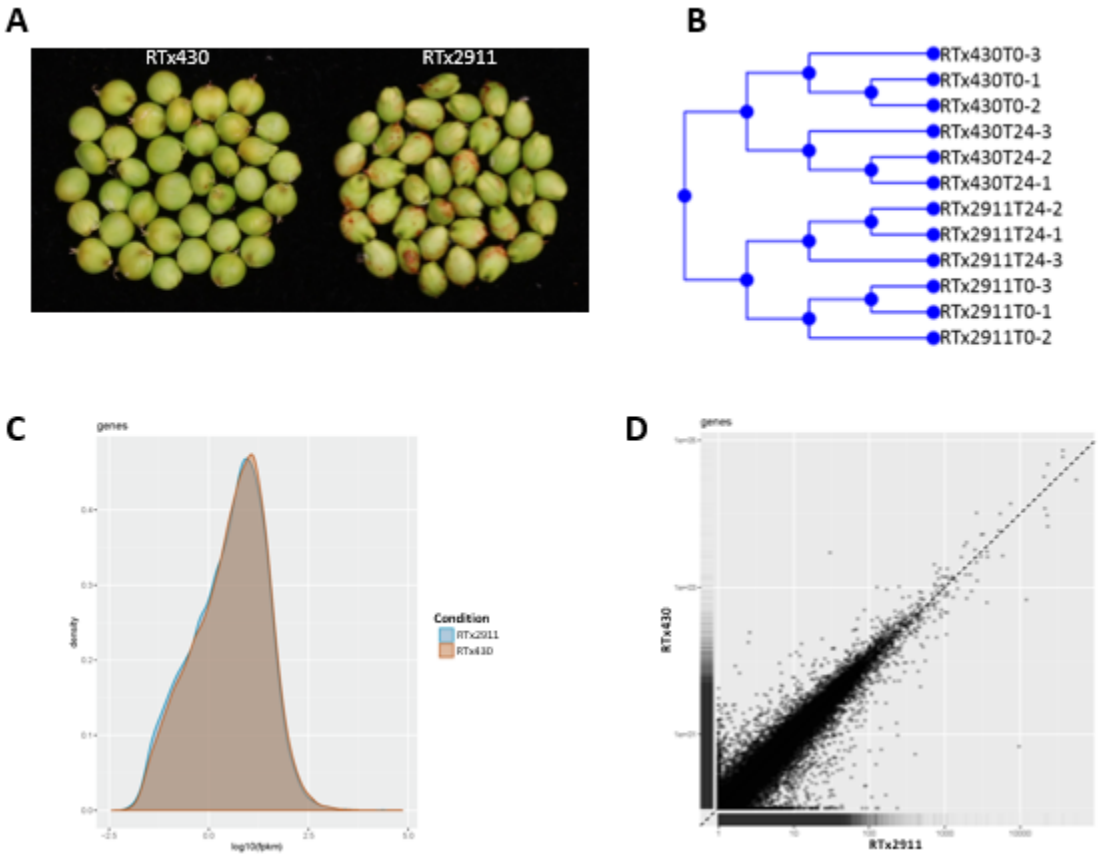


Figure 4.1. Developing grain of sorghum used for transcriptome analyses. A. Grain of sorghum RTx430 and RTx2911 at 20 days after flowering used for total RNA extraction. B. Hierarchical clustering of samples based on Euclidean distances. C. CummeRbund plots of the expression level distribution for all genes in RTx2911 and RTx430 at 24 hours after inoculation. FPKM, fragments per kilobase of transcript per million fragments mapped. D. CummeRbund scatter plots highlighting general similarities and specific outliers between RTx2911 and RTx430 at 24 hours after inoculation.

4.4.3 Genotype dependent differential gene expression

Comparisons of gene expression profiles between genotypes identified a number of genes that were differentially expressed between RTx2911 and RTx430 (Figure 4.2). A total of 1661

genes were differentially expressed at 0 hpi, of which 729 showed higher expression and 932 showed lower expression in RTx2911 compared to RTx430. At 24 hpi, 1955 genes were differentially expressed, of which 1085 were up-regulated and 870 were down-regulated in RTx2911 compared to RTx430. Some of these up and down-regulated genes were common between 0 and 24 hpi. These include 399 up-regulated and 413 down-regulated genes in RTx2911 compared to RTx430. Gene ontology (GO) enrichment analysis of genes differentially expressed between RTx2911 and RTx430 at 0 hpi revealed that there were no significantly enriched GO terms associated with this time point whereas analysis of genes differentially expressed at 24 hpi resulted in significantly enriched GO biological, molecular, kyoto encyclopedia of genes and genomes (KEGG) and cellular terms.

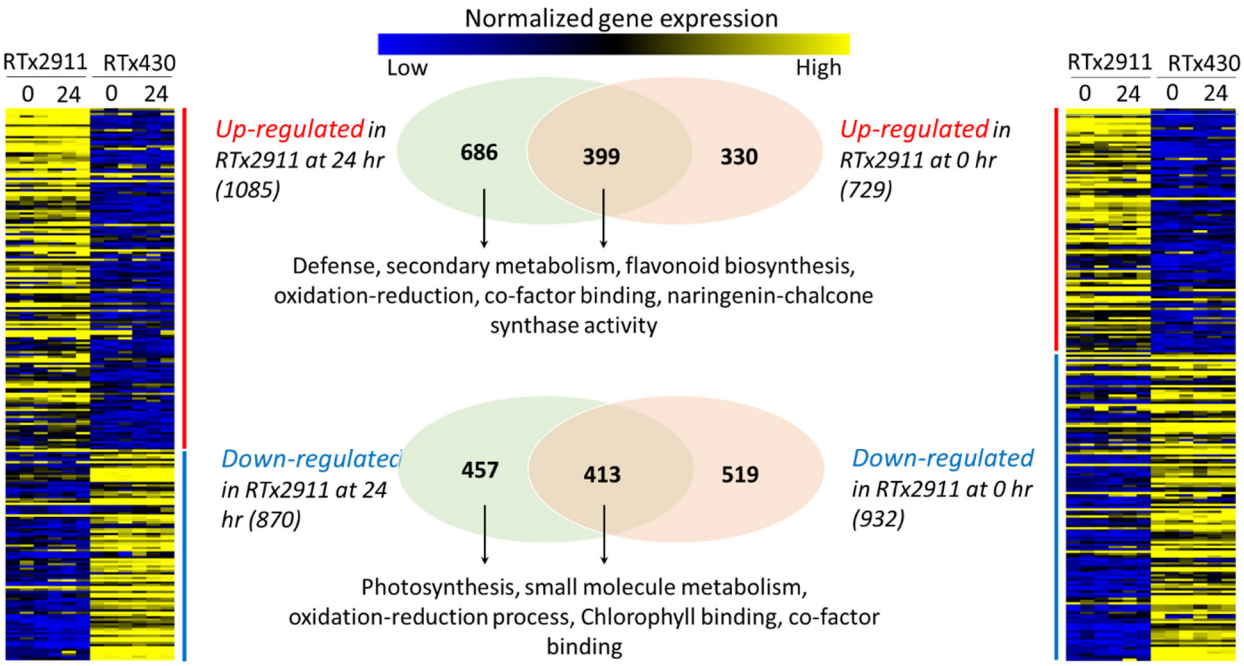


Figure 4.2. Genes differentially expressed between grain mold resistant and susceptible genotypes and significantly enriched gene ontology (GO) terms. GO terms displayed indicate those with highest significance.

4.4.4 Gene Ontology analyses of biological process regulated by fungal infection

GO enrichment analysis of genes up-regulated in the resistant genotype RTx2911 at 24 hpi identified biological processes that include defense response with 31 genes, biosynthesis of secondary metabolites with 22 genes, flavonoid biosynthesis with 6 genes, and oxidation-reduction

with 87 genes (Figure 4.3A). On the other hand, significantly enriched biological processes that were down-regulated in RTx2911 at 24 hpi included photosynthesis with 32 genes, small molecule metabolic process with 60 genes, and oxidation-reduction process with 79 genes (Figure 4.3B). The 31 defense response genes that were up-regulated in RTx2911 at 24 hpi include pathogenesis-related (PR) genes, NPR1/NIM1 (Sobic.001G143000), NPR1 interacting (Sobic.003G086200), defensins (gamma-thionin), antimicrobial peptides, receptor like kinases, WRKY transcription factors, jasmonate ZIM domain (JAZ), isoflavone 2'-hydroxylase, CHY zinc finger, and a putative nematode-resistance gene (Table 4.3). PR genes are widely known as markers of immune response activation and contribute to defense pathways including systemic acquired resistance [221] while the NPR1gene is required for both systemic acquired resistance and induced systemic resistance [222, 223]. The Sobic.004G317500 gene that encodes norcoclaurine synthase which is a member of PR10 related protein family was also up-regulated in RTx2911 [224]. Defensins (formerly called gamma-thionin) are small, highly stable, cysteine-rich antimicrobial peptides which are components of the plant immune response [225, 226]. Five defensin genes were up-regulated in response to inoculation in the resistant genotype RTx2911 compared to the susceptible RTx430. These defensin genes include Sobic.003G179300, Sobic.008G082300 and three tightly linked duplicate genes (Sobic.003G415200, Sobic.003G415300 and Sobic.003G415800). Interestingly, the up-regulated genes include the sorghum orthologue of the widely known LysM motif receptor kinase (LYK5) (Sobic.004G076100), leucine rich repeat (LRR) receptor-like serine/threonine-protein kinase (Sobic.006G217900) and somatic embryogenesis receptor-like kinase 1 (SERK1) (Sobic.006G104500) all of which were predicted to encode components of the pathogen recognitions and signaling complex. LYK5 is the major chitin receptor in Arabidopsis [227] and hence the sorghum Sobic.004G076100 gene referred here as *SbLYK5*, which encodes a LysM protein is the likely sorghum orthologue with a potential role in recognition of chitin which is a fungal microbe-associated molecular pattern.

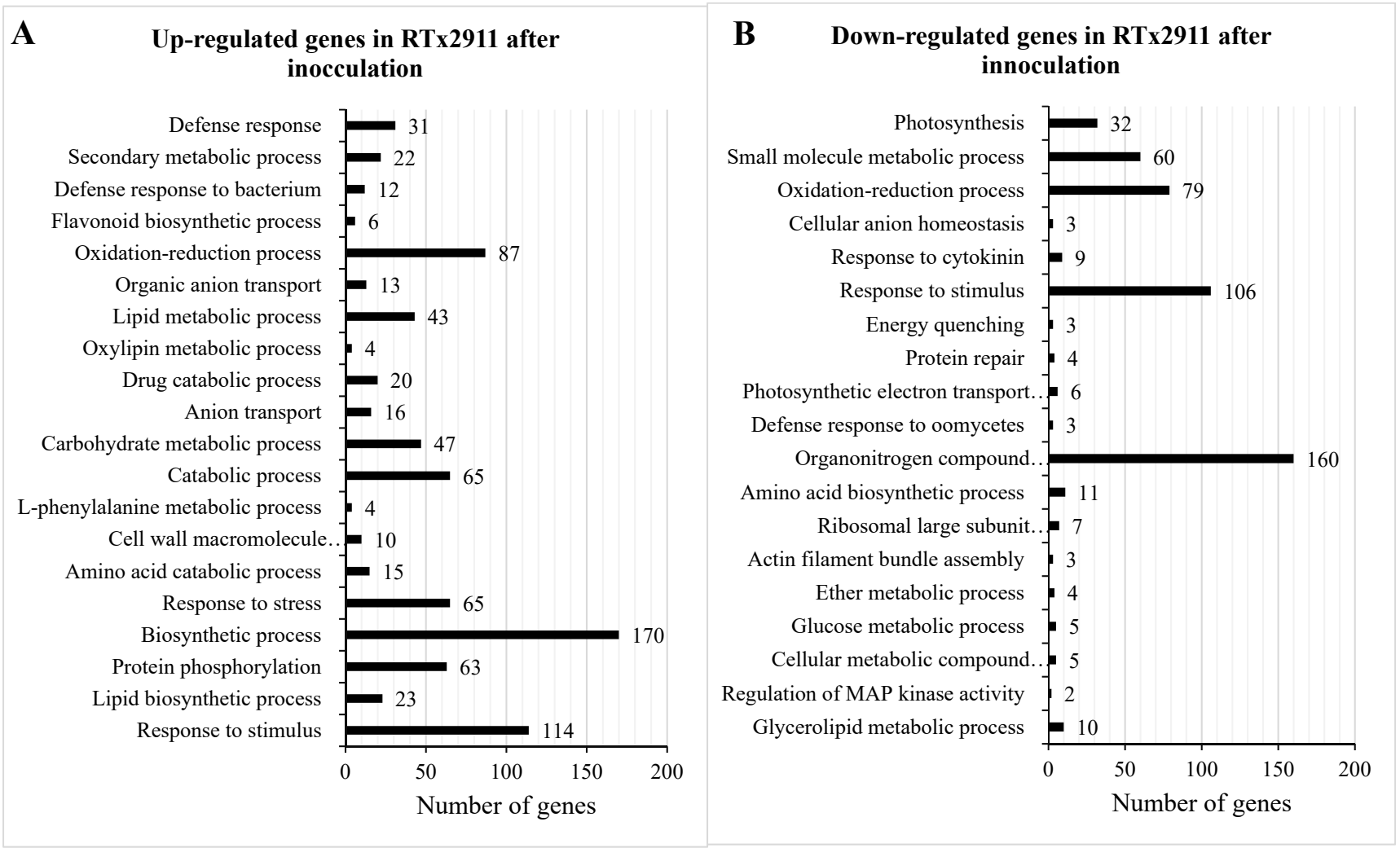


Figure 4.3. Gene Ontology enrichment analysis of DEGs between RTx2911 and RTx430 at 24hpi. Enriched GO biological process for up (A) and down (B) regulated genes at 24 hpi in RTx2911 compared to RTx430. The biological processes are sorted from top (highly significant) to bottom (least significant) enrichments.

Table 4.3. Defense genes induced at 24 hours after inoculation in RTx2911 compared to RTx430

No	Genes	Protein Family	Log_2 fold change	Test stat	p-value	q-value
1	Sobic.001G482700	Jasmonate ZIM	2.26844	3.80456	0.00005	0.000509
2	Sobic.002G214800	Jasmonate ZIM	2.14288	2.59649	0.00045	0.003348
3	Sobic.003G360900	Isoflavone 2'-hydroxylase	2.21360	3.65215	0.00005	0.000509
4	Sobic.001G143000	NPR1/NIM1	1.71135	2.65630	0.00020	0.001696
5	Sobic.001G373100	Ring finger and CHY zinc finger	1.21354	1.62130	0.00990	0.039520
6	Sobic.001G400800	Pathogenesis-related	1.12077	2.22246	0.00035	0.002713
7	Sobic.001G400900	Pathogenesis-related	1.44022	2.22800	0.00085	0.005660
8	Sobic.001G401300	Pathogenesis-related	1.65550	5.60828	0.00005	0.000509
9	Sobic.005G169200	Pathogenesis-related	1.47186	4.52536	0.00005	0.000509
10	Sobic.005G169400	Pathogenesis-related	1.84134	5.27806	0.00005	0.000509
11	Sobic.002G087500	Ricin-type beta-trefoil lectin	3.05161	5.06606	0.00005	0.000509
12	Sobic.003G086200	NPR1 interactor	2.25107	2.13331	0.00625	0.027758
13	Sobic.003G179300	Defensin	3.52422	4.23898	0.00110	0.006982
14	Sobic.003G415200	Defensin	4.32492	5.08045	0.00005	0.000509
15	Sobic.003G415300	Defensin	4.75506	7.44801	0.00005	0.000509
16	Sobic.003G415800	Defensin	2.22990	4.07364	0.00005	0.000509
17	Sobic.008G082300	Defensin	5.38255	5.58025	0.00005	0.000509
18	Sobic.003G233200	NAD dependent epimerase/dehydratase	2.58016	2.08103	0.00010	0.000942
19	Sobic.003G361100	Putative nematode-resistance	1.10019	3.01182	0.00005	0.000509
20	Sobic.004G065900	WRKY DNA binding	1.13824	1.67049	0.00785	0.033027
21	Sobic.004G076100	LYK5, LysM motif receptor kinase	3.53336	2.59092	0.00575	0.026088
22	Sobic.004G317500	(S)-norcoclaurine synthase	2.60706	3.39827	0.00005	0.000509
23	Sobic.005G165700	Plant antimicrobial peptide	2.19914	7.43702	0.00005	0.000509
24	Sobic.006G002400	Amidase family protein	1.11214	3.91534	0.00005	0.000509
25	Sobic.006G083000	Serine/threonine-protein kinase	1.04670	2.54241	0.00005	0.000509

Table 4.3 continued

No	Genes	Protein Family	Log_2 fold change	Test stat	p-value	q-value
26	Sobic.006G104500	SERK1 (somatic embryogenesis receptor-like kinase 1)	1.15820	2.03747	0.00135	0.008231
27	Sobic.006G217900	FLS2, LRR receptor-like serine/threonine-protein kinase	4.05319	4.17596	0.00005	0.000509
28	Sobic.007G030900	Copine	1.61544	5.03485	0.00005	0.000509
29	Sobic.009G113700	Peroxidase	1.24292	2.05953	0.00080	0.005378
30	Sobic.010G120800	Protein kinase	1.89065	2.33309	0.00140	0.008447
31	Sobic.010G241200	IAA-amino acid hydrolase ILR1-like 6	1.92368	3.26698	0.00005	0.000509

The sorghum RLK gene (Sobic.006G217900) encodes a putative flagellin receptor (FLS2) with 92.2% amino acid similarity to the maize gene GRMZM2G080041 [228]. FLS2 is a well characterized flagellin receptor in Arabidopsis [229] that plays a critical role in pathogen perception and signaling [228]. The WRKY transcription factor gene (Sobic.004G065900) was up-regulated in RTx2911 and shows similarity to the WRKY71 and WRKY40 genes. The Arabidopsis WRKY40 is a pathogen inducible transcription factor which along with WRKY18 and WRKY60 contributes to defense against pathogens [230]. Jasmonate ZIM domain (JAZ) proteins are transcriptional repressors in jasmonic acid (JA) responses but also play role in regulation of defense-growth balance [231]. The genes (Sobic.001G482700, Sobic.002G214800) up-regulated in RTx2911 that encode JAZ proteins may have similar roles in maintaining defense and growth balance in sorghum. Another gene up-regulated in the resistant genotype (Sobic.003G360900) encodes the isoflavone 2'-hydroxylase which catalyzes steps in phytoalexin biosynthesis pathway and modulates pathogen induced phytoalexin accumulation [232]. Moreover, Sobic.001G373100 gene, up-regulated in RTx2911 encodes ring finger and CHY zinc finger domain-containing protein. Such proteins are involved in diverse biological functions including defense against pathogens [233]. The Sobic.007G030900 gene which is up-regulated in RTx2911 encodes a copine protein that is reported as a possible suppressor of defense responses in Arabidopsis [234]. Copines are conserved calcium-dependent membrane-binding proteins [235]. The Sobic.006G002400 gene which is also up-regulated in RTx2911 encodes amidase family protein. Amidase family proteins are specific indole-3-acetamide amidohydrolase enzymes that catalyze the synthesis of indole-3-acetic acid (IAA) from indole-3-acetamide [236]. IAA is a widely known auxin that regulates plant growth and development. IAA, however, may impact disease resistance negatively [237] which could play a role in balancing immune responses and plant fitness [238]. Another up-regulated gene with a closely related function is Sobic.010G241200 that encodes an IAA-amino acid hydrolase ILR1-Like 6. IAA-amino acid hydrolases cleave IAA-amino acid conjugates releasing free IAA [239]. The sorghum homolog of putative nematode resistance gene Hs1pro-1 [240] (Sobic.003G361100) was also up-regulated in the resistant genotype.

4.4.5 Gene ontology analyses of molecular processes identify multiple differentially regulated pathways

Significantly enriched GO molecular processes associated with the DEGs in RTx2911 include cofactor binding with 69 genes, oxidoreductase activity with 88 genes, naringenin-chalcone synthase activity with 4 genes, hydrolase activity with 141 genes (Figure 4.4A). Additional DEGs fall with the molecular functions such as DNA binding, carbohydrate binding, protein kinase activity, transcriptional regulation and transmembrane activities. Various protein kinase genes up-regulated in RTx2911 include receptor like kinases, wall associated kinases, and mitogen-activated protein kinases (MAPKs). Similarly, enriched molecular processes associated with the down-regulated genes include chlorophyll binding with 9 genes, cofactor binding with 57 genes, and oxidoreductase activity with 76 genes (Figure 4.4B). Hydrolase activity, carbohydrate binding, and catalytic activity acting on protein constitute the top three categories of processes represented by the upregulated genes in the resistant genotype. Similarly, anion binding, small molecule binding, and oxidoreductase were processes that were represented by a larger proportion of down-regulated genes.

4.4.6 KEGG enrichment analysis of metabolic pathways for DEGs

KEGG enrichment analysis using genes up-regulated in RTx2911 at 24 hpi identified significantly enriched metabolic pathways associated with grain mold resistance. These include biosynthesis of flavonoid (11 genes), other secondary metabolites (53 genes), ubiquitin and other terpenoid-quinone, phenylpropanoid, phenylalanine and brassinosteriods (Figure 4.5A). The up-regulated flavonoid biosynthesis genes include 4 chalcone synthase (Sobic.005G136200, Sobic.005G136300, Sobic.005G137000, and Sobic.005G137300), chalcone-flavonone isomerase (Sobic.001G035600), cytochrome P450 (Sobic.002G126600), flavonoid 3'-hydroxylase (Sobic.004G200900), glucosyl/glucuronosyl transferase (Sobic.007G027301), shikimate O-hydroxycinnamoyl transferase (Sobic.006G136800), bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase (DFR, Sobic.004G050200) and cinnamate 4-hydroxylase (Sobic.004G141200) genes. On the other hand, down-regulated genes were in photosynthesis (11 genes), ribosome, purine, pyrimidine, and carbon metabolism functions (Figure 4.5B).

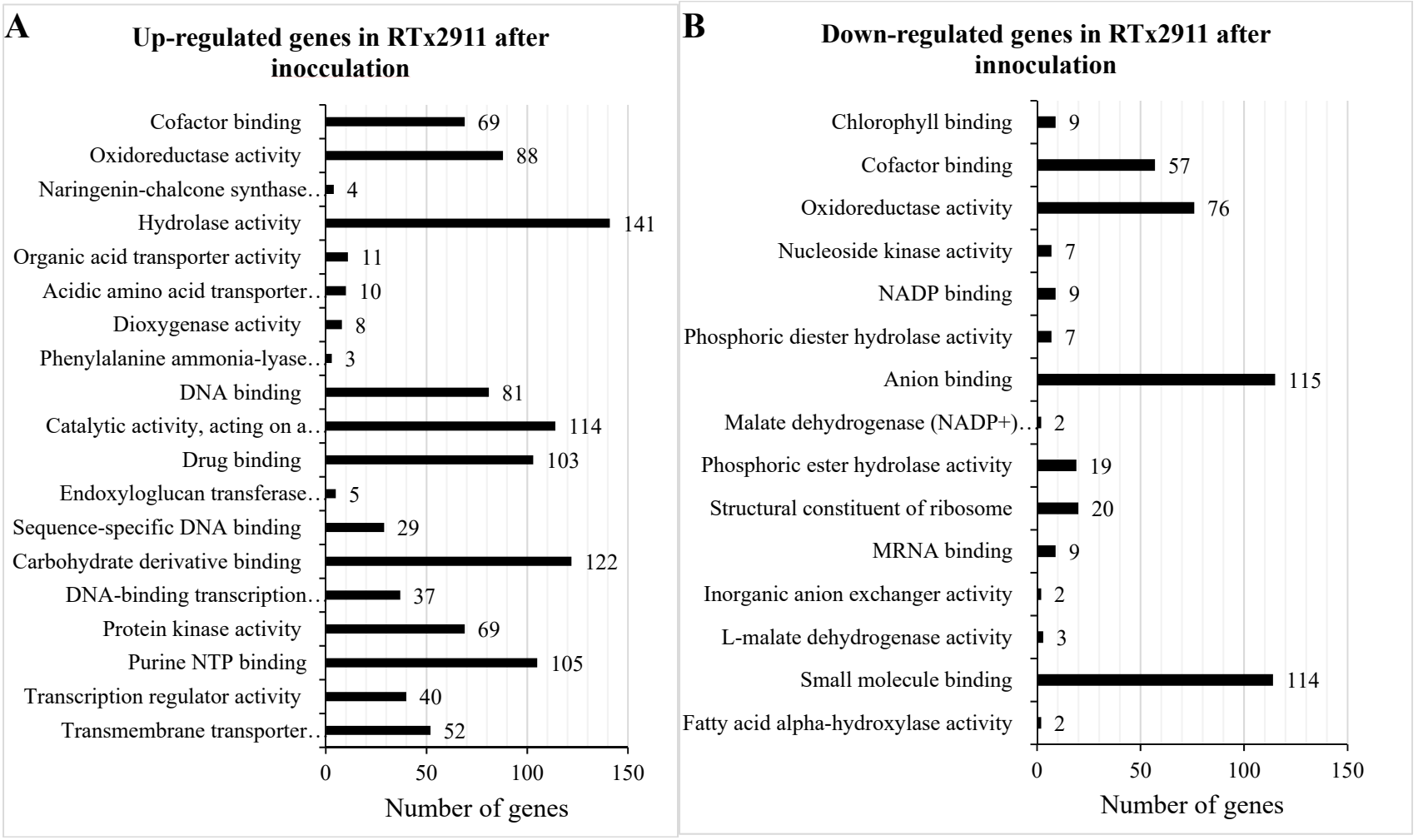


Figure 4.4. Gene Ontology enrichment analysis of DEGs between RTx2911 and RTx430 at 24hpi. A) Enriched GO molecular process of up-regulated genes at 24 hpi in RTx2911 compared to RTx430. B) Enriched GO molecular process of down-regulated genes at 24 hpi in RTx2911 compared to RTx430

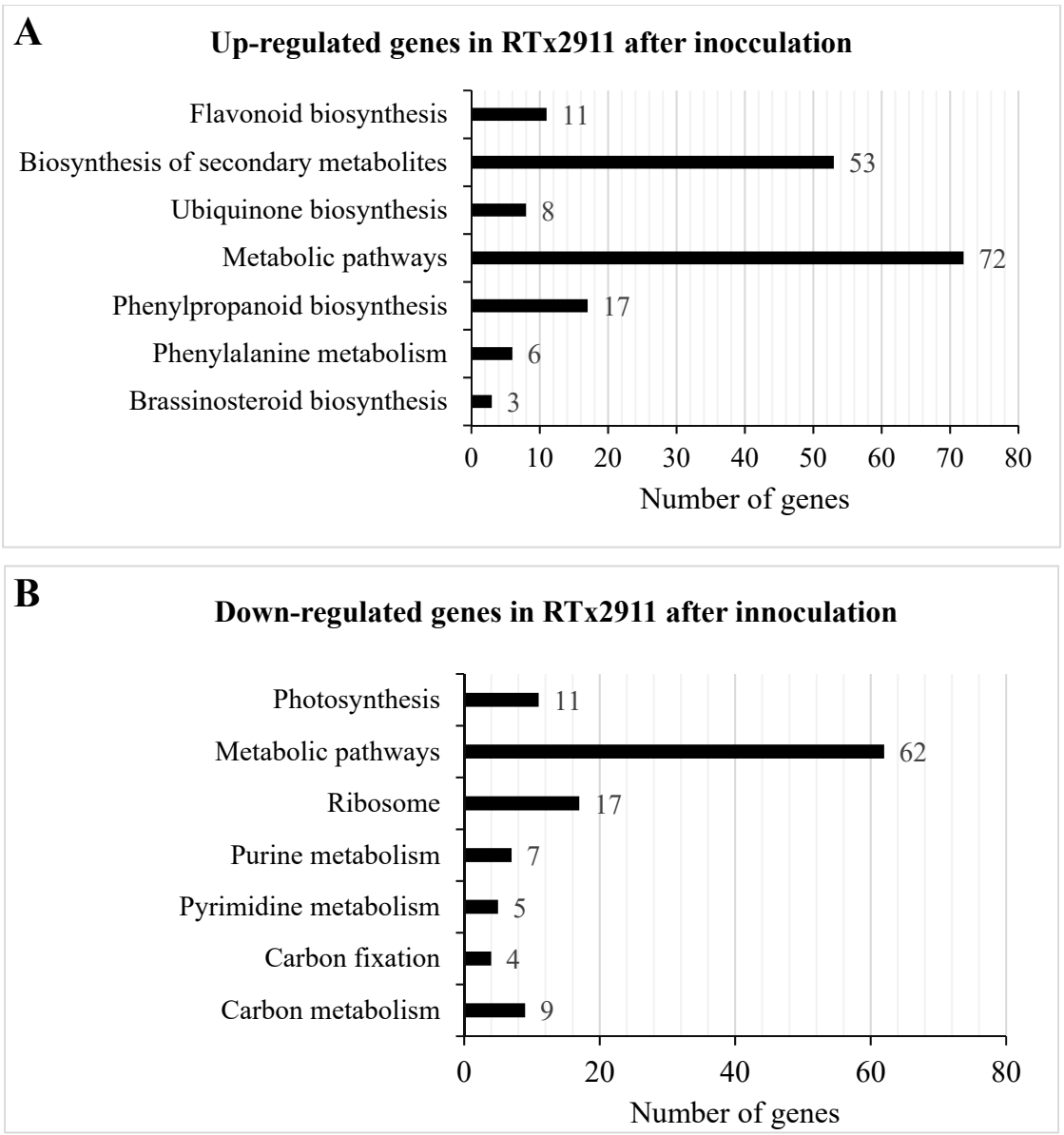


Figure 4.5. KEGG pathways for genes differentially expressed between genotypes. A) Enriched KEGG pathways for genes with higher induced expression in RTx2911 at 24 hpi compared to RTx430. B) Enriched KEGG pathways for genes down-regulated in RTx2911 at 24 hpi compared to RTx430

4.4.7 Pathogen induced differential expression of genes in developing sorghum grain

In order to decipher DEGs induced in response to fungal inoculation, transcriptome comparisons were made between samples prior to and after inoculation for each genotype. A number of genes were differentially expressed in response to fungal inoculation (Figure 4.6). Consequently, 947 DEGs with altered expression in response to inoculation were identified in

RTx2911 with 707 up-regulated and 240 down-regulated at 24 hpi compared to 0 hpi. Similarly, 706 genes were differentially expressed between the two time points in RTx430 with 359 genes up-regulated and 347 down-regulated at 24 hpi compared to 0 hpi. Among these, 59 genes were down-regulated at 24 hpi in both RTx2911 and RTx430 compared to basal expression at 0 hpi.

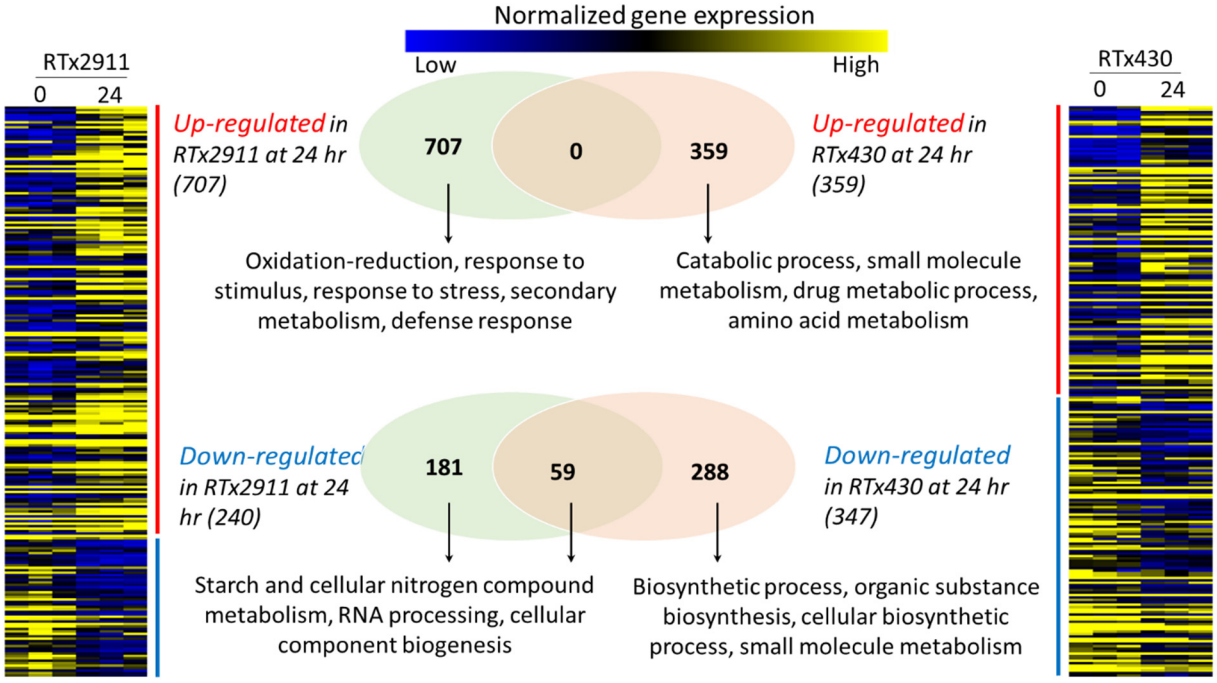
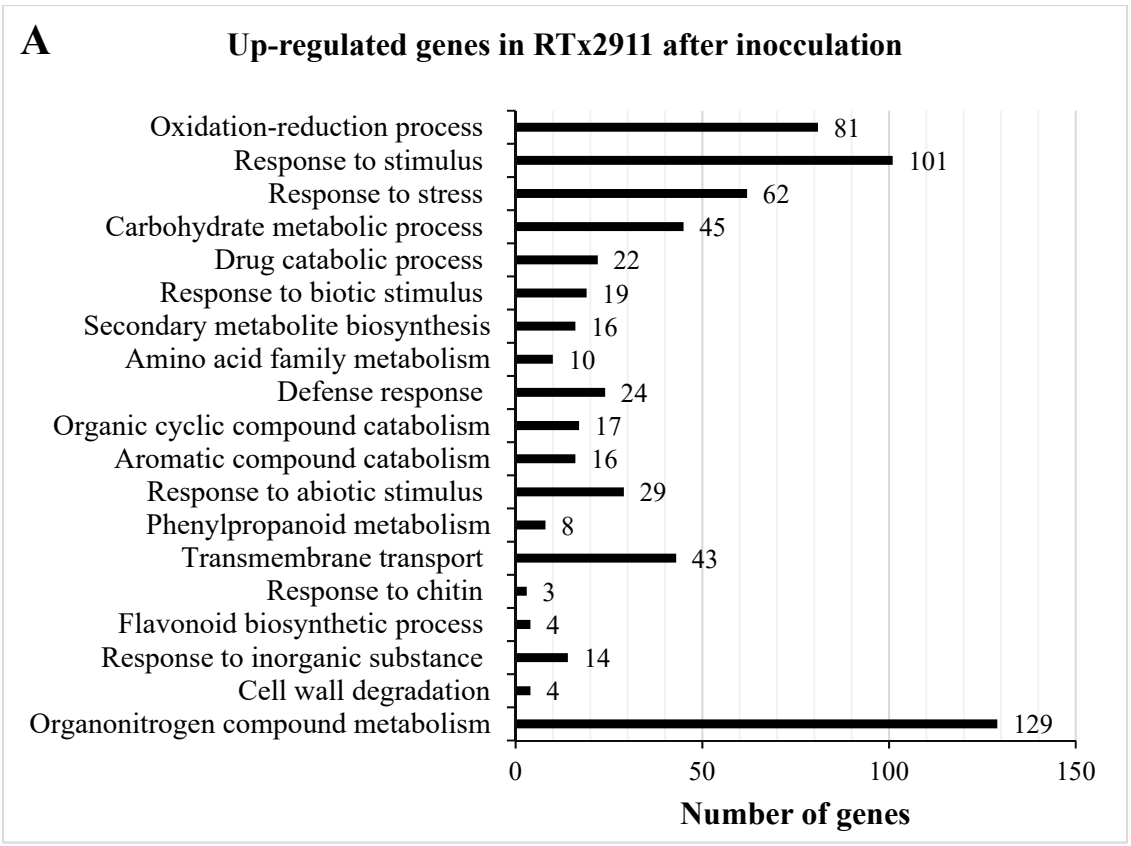


Figure 4.6. Differentially expressed genes regulated by pathogen inoculation in grain mold resistant and susceptible genotypes and significantly enriched GO terms. GO terms displayed are those with highest significance.

GO enrichment analysis of genes differentially expressed between the time points for each genotype revealed significantly enriched biological, chemical, cellular and KEGG pathways. This analysis which compares differentially expressed genes between 0 (before infection) and 24 h (after infection) indicates genes that are particularly induced upon infection in each of the genotypes. Genes up-regulated after infection in RTx2911 were assigned mainly to defense associated biological processes such as response to stimulus (101 genes), stress (62 genes), defense (24 genes), chitin response (3 genes), biotic stimulus (19 genes), oxidation-reduction, carbohydrate metabolism, secondary metabolism, flavonoid biosynthesis and others (Figure 4.7A). Genes down-regulated upon infection in RTx2911 were assigned to starch metabolism, cellular nitrogen compound metabolism, RNA processing, DNA metabolic process and others (Figure 4.7B).



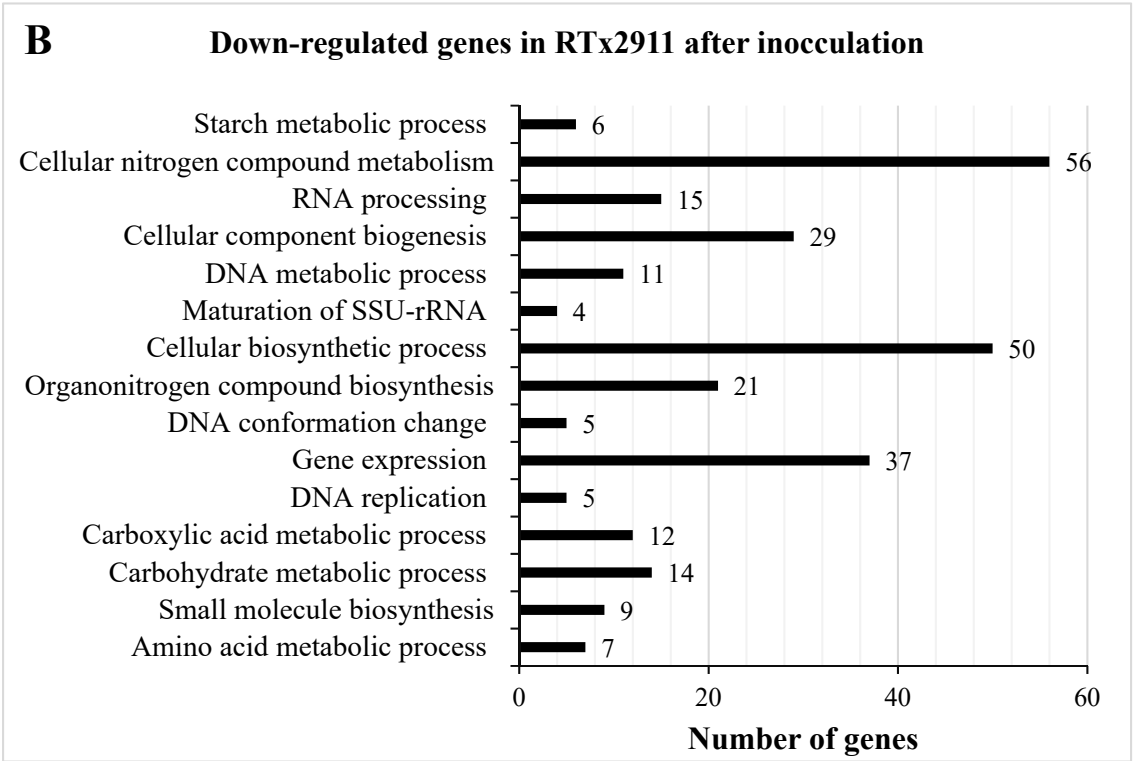


Figure 4.7. Enriched GO biological processes between 0 and 24 hpi for RTx2911. Up-regulated (A) and down regulated (B) genes at 24 hpi compared to 0 hpi.

Genes induced upon infection in RTx2911 that were also differentially expressed between the two genotypes include a JAZ repressor (Sobic.001G482700), isoflavone 2'-hydroxylase (Sobic.003G360900), ring finger and CHY zinc finger (Sobic.001G373100), nematode-resistance (Sobic.003G361100), WRKY DNA binding (Sobic.004G065900), and *SbLYK5* (Sobic.004G076100) genes (Table 4.4). Moreover, cytochrome P450 (Sobic.001G077400), ornithine aminotransferase (Sobic.001G156100), defensins (Sobic.001G165600, Sobic.005G153600, Sobic.007G075250, Sobic.007G075301), pathogenesis-related (9 genes), the NAC protein geminivirus rep a-binding 1 (GRAB1) (Sobic.003G379700), heat shock (Sobic.006G005600) and triacylglycerol (TAG) lipase (Sobic.007G194800) genes were induced upon infection in the resistant genotype.

Table 4.4. Defense genes induced upon infection in RTx2911 at 24 hpi

No	Genes	Protein Family	Log_2 fold change	Test stat	p-value	q-value
1	Sobic.001G482700	Jasmonate ZIM	1.02671	2.02240	0.00055	0.008999
2	Sobic.003G360900	Isoflavone 2'-hydroxylase	1.23157	2.26308	0.00075	0.011313
3	Sobic.001G077400	Allene oxide synthase	1.49778	2.72515	0.00010	0.002305
4	Sobic.001G156100	Ornithine aminotransferase	1.29051	4.01677	0.00005	0.001305
5	Sobic.001G165600	Defensin	1.23670	2.56264	0.00005	0.001305
6	Sobic.005G153600	Defensin	1.14584	2.57529	0.00005	0.001305
7	Sobic.007G075250	Defensin	1.68390	4.82551	0.00005	0.001305
8	Sobic.007G075301	Defensin	1.93758	5.78103	0.00005	0.001305
9	Sobic.001G373100	Ring finger and CHY zinc finger	1.71346	2.06206	0.00235	0.025444
10	Sobic.001G400700	Pathogenesis-related	1.26150	1.74839	0.00370	0.034773
11	Sobic.001G400800	Pathogenesis-related	1.42753	2.49106	0.00005	0.001305
12	Sobic.001G400900	Pathogenesis-related	2.86101	3.23454	0.00005	0.001305
13	Sobic.001G401100	Pathogenesis-related	3.26565	2.91294	0.00125	0.016513
14	Sobic.001G401200	Pathogenesis-related	4.12443	5.39278	0.00005	0.001305
15	Sobic.001G401300	Pathogenesis-related	2.19583	5.97457	0.00005	0.001305
16	Sobic.005G169200	Pathogenesis-related	4.41356	7.95931	0.00005	0.001305
17	Sobic.005G169300	Pathogenesis-related	1.70302	4.90227	0.00005	0.001305
18	Sobic.005G169400	Pathogenesis-related	2.91304	6.29651	0.00005	0.001305
19	Sobic.003G361100	Putative nematode-resistance	1.52266	3.49796	0.00005	0.001305
20	Sobic.003G379700	NAC23 (GRAB1 like protein)	1.06726	3.28441	0.00005	0.001305
21	Sobic.004G065900	WRKY DNA binding	1.47453	1.88306	0.00495	0.042569
22	Sobic.004G076100	LYK5, LysM motif receptor kinase	1.67926	1.91315	0.00325	0.031605
23	Sobic.006G005600	Heat shock protein	2.98005	5.59362	0.00005	0.001305
24	Sobic.007G194800	Triacylglycerol lipase 2	1.24681	3.73981	0.00005	0.001305

The Sobic.001G077400 gene encodes an allene oxide synthase (AOS), which is a cytochrome P450 protein. The sorghum putative AOS shares high similarity (98%) to the maize hydroperoxide dehydratase. AOS shows hydroperoxide dehydratase activity which catalyzes the first step in the biosynthesis of jasmonic acid, a major regulator of plant defense to necrotrophic fungal pathogens [241]. High level of such enzymes accumulate in pericarps and seed coats [242] which suggests their important roles in defense against grain pathogens. The Sobic.001G156100 gene encodes a highly conserved enzyme, ornithine aminotransferase which contributes to both R-gene mediated and non-host resistance through proline metabolic pathway [243, 244]. Four defensin genes (Sobic.001G165600, Sobic.005G153600, Sobic.007G075250, Sobic.007G075301) were induced upon infection in the resistant genotype. Two of these genes (Sobic.007G075250, refereed here as *SbDFN7.1*; Sobic.007G075301, refereed here as *SbDFN7.2*) are tightly linked on sorghum chromosome 7, and transcribed in opposite orientation with a likely common promotor (Figure 4.8).

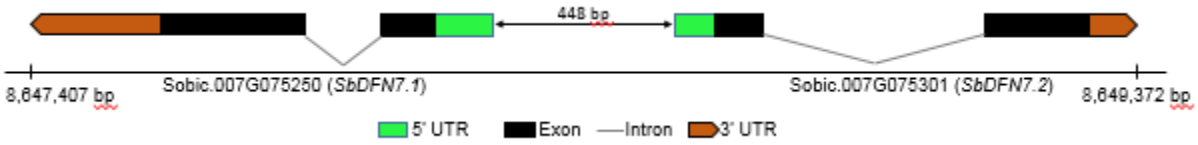


Figure 4.8. Schematic representation of the genomic organization of defensin genes *SbDFN7.1* and *SbDFN7.2* on chromosome 7

SbDFN7.1 and *SbDFN7.2* are similar to maize *ZmDEF1* (GRMZM2G368890) and *ZmDEF2* (GRMZM2G368861) in both genomic organization and sequence similarity. The intergenic region of the maize defensin genes *ZmDEF1* and *ZmDEF2* is considered as an embryo-specific asymmetric bidirectional promoter [245]. These defensin genes are specifically and highly expressed in seeds. Plant defensins are pathogen inducible [246] antimicrobial peptides [226]. The Sobic.003G379700 which encodes a NAC transcription factor was induced upon infection in RTx2911 and shares high similarity (97%) to the maize GRAB1-like protein. NAC transcription factors play role in regulation of biotic and abiotic stress responses [247] while the GRAB1 proteins which are members of the NAC domain family are known for their interaction with a geminivirus protein [248]. The other induced gene in RTx2911, Sobic.006G005600, encodes a heat shock protein (HSP90). HSP90 is the most abundant cytosolic heat shock protein family [249]

and plays important roles in immune responses [250, 251]. The Sobic.007G194800 gene induced in RTx2911 encodes an important protein TAG lipase, which is similar to the Arabidopsis *phytoalexin deficient 4* (PAD4) [252]. Plants with *pad4* mutations display defects in multiple defense responses with reduced camalexin synthesis, PR-1 gene expression and SA levels [253]. This is interesting because sorghum produces the phytoalexin 3-deoxyanthocyanidins which accumulate in response to fungal infection [90]. 3-deoxyanthocyanidins are synthesized through the flavonoid biosynthesis pathway.

Genes up-regulated in RTx430 upon infection were assigned to catabolic process, small molecule metabolism, drug metabolic process, cellular homeostasis, response to biotic stimulus (9 genes) and defense (10 genes) (Figure 4.9A). The 10 defense genes up-regulated in RTx430 include 2 genes that were specifically induced in RTx430, which are an NBS-LRR resistance gene (Sobic.005G092600) and l-type lectin-domain containing receptor kinase (Sobic.004G118800) and 8 were similar to that of RTx2911. The 8 genes commonly up-regulated in both RTx2911 and RTx430 are ornithine aminotransferase (Sobic.001G156100), 5 PR genes (Sobic.001G400800, Sobic.001G401100, Sobic.001G401200, Sobic.001G401300, Sobic.005G169400), HSP90 (Sobic.006G005600) and TAG lipase (Sobic.007G194800). On the other hand, genes down-regulated in RTx430 upon infection were assigned to biosynthetic process, organic substance biosynthesis, cellular biosynthesis process, small molecule metabolism, oxidation-reduction process, secondary metabolism biosynthesis and others (Figure 4.9B).

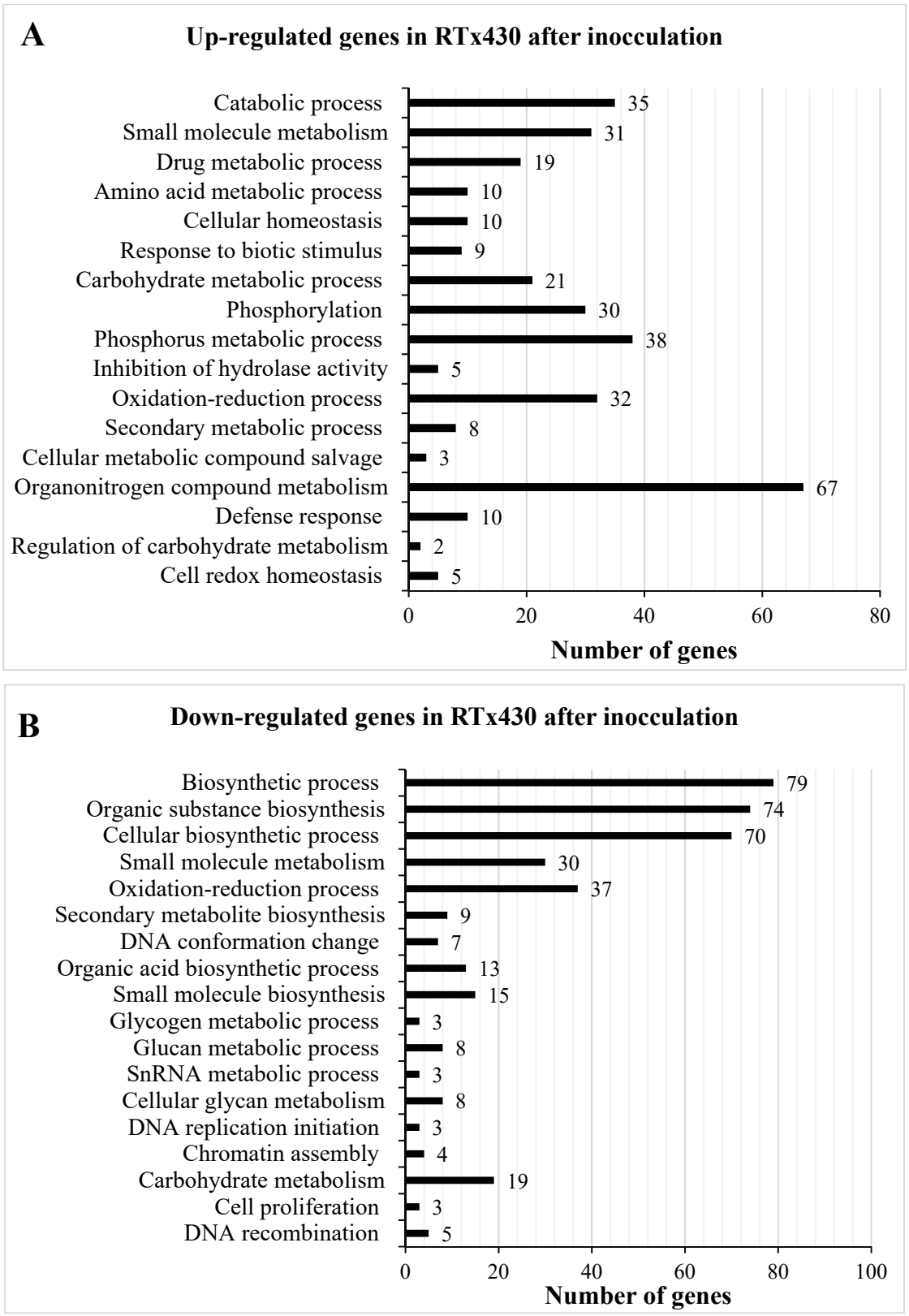


Figure 4.9. Enriched GO biological processes between 0 and 24 hpi for RTx430. Up-regulated (A) and down regulated (B) genes at 24 hpi compared to 0 hpi.

4.4.8 Increased expression of genes encoding seed storage proteins in resistant genotype

A major variation between RTx430 and RTx2911 was observed in expression of genes encoding seed storage proteins. Sobic.005G184500 annotated as zein seed storage protein was the most variable between the two genotypes in terms of expression both prior to and after inoculation. This gene showed higher expression in RTx2911 with a Log2 fold change of 9.7 in non-inoculated grain and 10.7 at 24 h after inoculation. Sobic.008G144201 was another gene with higher basal and pathogen induced expression in RTx2911 that also encodes a zein seed storage protein. Both Sobic.005G184500 and Sobic.008G144201 were highly expressed in developing grain of the resistant genotype RTx2911. Zein and kafirins are major seed storage proteins in maize and sorghum, respectively, which are associated with kernel texture [182, 254, 255]. Recently, a major kafirin locus was discovered as key determinates of grain mold resistance in sorghum [151].

4.4.9 Validation of differential expression of selected defense genes using qRT-PCR

To validate the sequence data and also determine expression pattern of some genes beyond the two time points used for RNA-seq, expression of selected genes encoding defensins and a JAZ protein genes were studied using qRT-PCR. Broadly, the resistant genotype showed a significantly higher level of expression than the susceptible genotype (Figure 4.10).

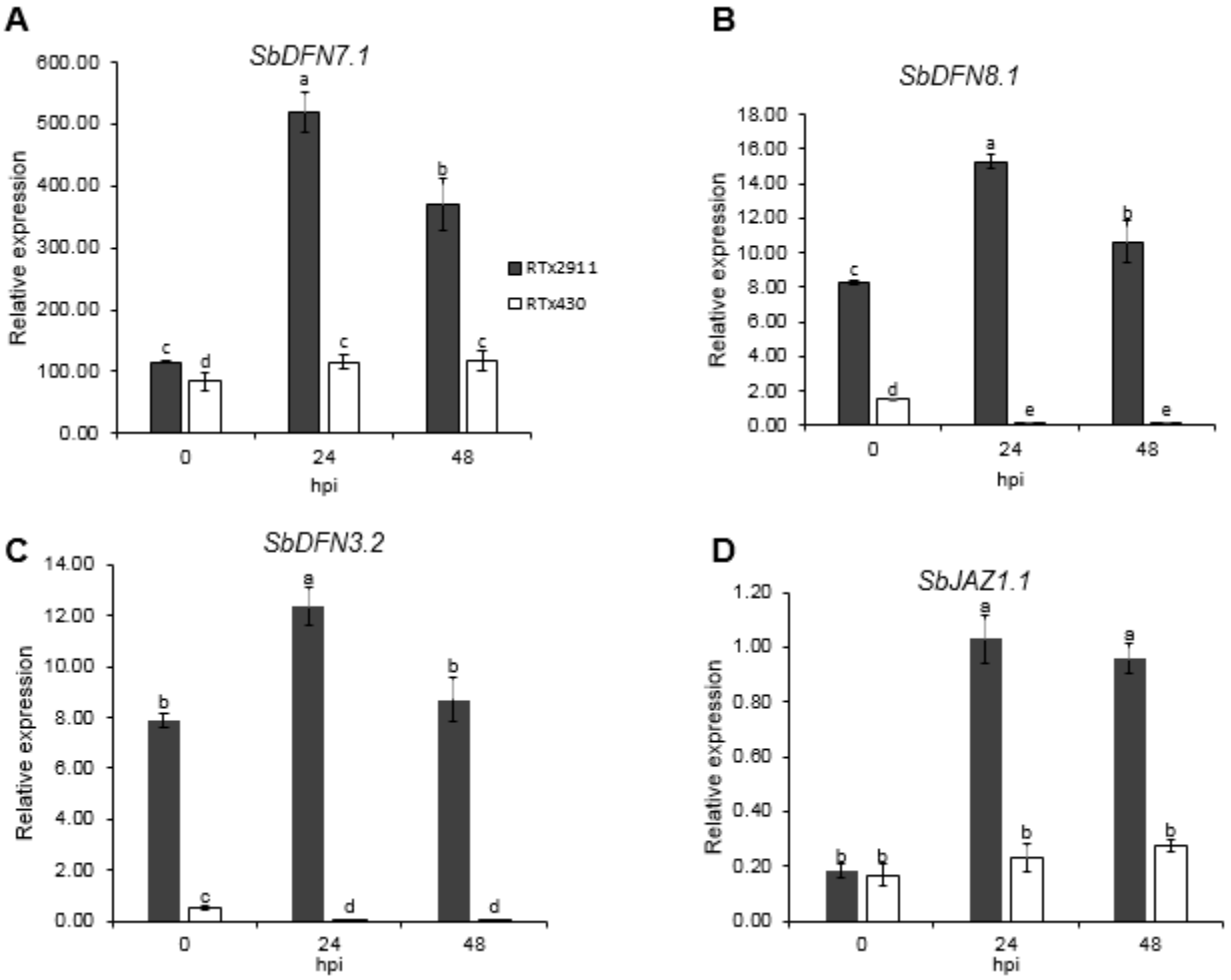


Figure 4.10. Validation of gene expression through qRT-PCR analysis of selected *Defensin (gamma-thionin)* (A-C) and Jasmonate ZIM domain (JAZ) genes (D). hpi; hours post inoculation.

The expression of the sorghum defensin genes *SbDFN7.1* (Sobic.007G075250), *SbDFN8.1* (Sobic.008G082300), *SbDFN3.2* (Sobic.003G415300) (Figure 4.10A-C) and the *SbJAZ1.1* gene (Sobic.001G482700) (Figure 4.10D) were consistent with those observed in RNA-seq. The three genes that encode defensins were highly induced at 24 hpi in the resistant genotype RTx2911. At 48 hpi, the expression of these genes varied slightly with *SbDFN7.1* (Figure 4.10A) and *SbDFN8.1* (Figure 4.10B) but remained higher than that of 0 hpi but slightly lower than 24 hpi whereas the expression of *DFN3.2* at 48 hpi leveled to the 0 hpi (Figure 4.10C). The expression of the *SbJAZ1.1* gene increased significantly at 24 hpi and remained high at 48 hpi (Figure 4.10D). The expression of these genes in the susceptible RTx430 was very low at all the time points.

4.5 Discussion

This study focused on transcriptome changes in the developing grain in response to simultaneous infection by grain molding fungal species. Defense responses in grain tissues to single or a mixture of multiple pathogenic species have not been studied previously. Responses to a mixture of fungi rather than a single species mirrors sorghum grain mold disease in the field under natural infestations. Global changes in gene expression, molecular and cellular functions, and metabolic pathways that are reprogrammed early during infection of the developing grain were delineated, which together are likely to explain variations in plant responses to the disease. Comparative transcriptome and subsequent gene ontology enrichment analysis in resistant and susceptible sorghum genotypes revealed differentially expressed genes that are associated with major plant defense pathways, seed proteins and antimicrobial protein genes that were preferentially expressed in the resistant genotype. Genes that showed higher basal and induced gene expression in the resistant genotype relative to the susceptible genotype are implicated in key plant defense pathways. Antimicrobial peptides including plant defensins and genes that encode proteins that preferentially accumulate in the seed but are also induced in response to infection were identified. This is consistent with the role of seed proteins and other compounds that regulate the physical and chemical properties of kernels, and thus provide resistance to grain mold. Interestingly, we also observed differential expression of genes encoding proteins that function in pathogen recognition, signal transduction, and other defense responses sharing similarity to immune mechanisms in leaf tissues in many plant pathogen interactions.

The major defense related genes induced in the resistant genotype RTx2911 in response to infection include PR proteins, antimicrobial peptides including defensins, receptor like kinases, regulators of systemic acquired resistance (SAR) and biosynthesis of phytoalexins as well as genes known to be involved in flavonoid biosynthesis. Analyses of enriched molecular processes identified components of pathogen recognition and response signaling such as receptor like protein kinases, wall associated kinases and mitogen-activated protein kinases (MAPKs). Thus, resistance to grain mold in developing sorghum grain involves active defense processes that involve recognition of pathogen or damage associated molecular patterns by plant receptors followed by activation of signal transduction pathways that trigger multiple immune responses consistent with the quantitative nature of grain mold resistance. Such active defense response pathways likely culminate in synthesis of antimicrobial molecules, changes in seed protein profile, and

enhancement of seed physical and biochemical defenses which may be superimposed on passive defense mechanisms.

PAMP triggered immunity (PTI) to pathogens is a form of quantitative resistance that is initiated by perception of evolutionarily conserved pathogen derived molecules, such as chitin fragments, by surface localized pattern recognition receptors (PRRs) [256]. The induced expression of the sorghum LysM motif receptor kinase (*SbLYK5*) in response to infection in the resistant genotype RTx2911 is consistent with the activation of PTI. The Arabidopsis *AtLYK5* is the receptor for chitin and is also chitin inducible [227] suggesting the sorghum orthologue identified in our study may have similar functions. Sorghum 3-deoxyanthocynidin, phytoalexins synthesized through the flavonoid pathway, and known to accumulate in response to pathogen infection may be activated by perception of fungal derived chitin fragments by *SbLYK5*. The fact that several flavonoid biosynthesis genes were induced upon infection in our study, and the co-expression of PRRs supports that the phytoalexin biosynthesis branch of the flavonoid biosynthesis pathway may be correlated with chitin perception and response signaling in the developing grain. Perception of pathogen derived elicitor by membrane localized PRRs, and their subsequent response signaling by their downstream components such as receptor like cytoplasmic kinases (RLCK) and MAPKs are known to contribute to activation of defense responses [257-259]. The enhanced expression of sorghum genes encoding putative PRRs, RLCKs and MAPKs in the resistant genotype suggest the role of PTI mechanisms in restricting the severity of grain mold in the developing grain. The data also suggest that in the developing grain that is at the physiologically active stage, the induced immune mechanism may contribute significantly, which may decline after the grain is physiologically mature when physical or passive mechanisms are likely to supersede.

Different pathogenies-related (PR) genes with higher basal and induced expression in the resistant than the susceptible genotype suggest their critical roles in resistance against grain mold in sorghum. PR-related proteins are conserved protein families involved in plant immunity [260, 261] some of which are involved in both biotic and abiotic stress responses [262]. The PR genes identified in this study occur as clusters of duplicates in two loci in sorghum which are located at 68.6 and 64.8 Mbp on chromosome 1 and 5, respectively. Those on Chromosome 1 encode proteins similar to the Bet v I family of PR-10 and those on Chromosome 5 encode chitinase-related proteins. PR-10 proteins have ribonuclease activities [263, 264]. Chitinases accumulate in

response to stress or pathogen attack [265]. Some PR genes identified in the current study were induced upon infection in both the resistant and the susceptible genotypes but some were only induced in the resistant genotype.

Our data suggest that defensins which are small (~ 5kDa) basic, cysteine-rich antimicrobial peptides [226, 266] are among the major elements of the sorghum defense system that are induced in response to grain mold fungi that are typically necrotrophic pathogens. Plant defensins are classified as PR-12 family proteins [267, 268] and are components of the plant immune response especially to necrotrophic fungi [266, 269] with high fungi toxic activities [270] and the majority of defensins reported accumulate in the seed [225]. Several genes encoding these peptides were highly induced upon infection in the resistant genotype RTx2911 but their expression was severely attenuated in the susceptible RTx430. Defensin expression is dependent on functional ethylene and jasmonic acid response pathways [246]. A cytochrome P450 gene encoding allene oxide synthase (AOS) which is involved in the biosynthesis of JA [241] was induced upon infection in the resistant genotype. JA is associated with defense against necrotrophic fungi [271] and may play critical regulatory roles in the activation of defenses against grain mold in sorghum including the expression of defensins that require JA perception and signaling. Sorghum defensins have not been studied as they are mostly grain specific and pathogen inducible whereas most previous studies focus on foliar tissues. It is notable that defensins were not prominently described in recent RNA-seq experiments conducted in leaf tissues of sorghum consistent with their grain specific expression [272, 273].

Among major and widespread plant defense responses to pathogens are the pathogen induced accumulation of phytoalexins, which are low molecular weight antimicrobial compounds [274, 275]. Sorghum produces the 3-deoxyanthocyanidin phytoalexins, apigeninidin and luteolinidin [91] through the flavonoid biosynthesis pathway. Indeed our study indicated that several flavonoid biosynthesis pathway genes were differentially expressed between the resistant RTx2911 and susceptible RTx430, a subset of which were also induced upon infection in the resistant genotype. The TAG lipase protein gene which is similar to Arabidopsis PAD4, was induced upon pathogen inoculation in both resistant and susceptible genotypes. Looking at a previous study conducted on the biosynthesis pathway of camalexin and the nature of the enzymes involved, it seems that the camalexin biosynthesis pathway has some level of similarity to that of the sorghum cyanogenic glycoside dhurrin [276]. Metabolite profiling of developing grain also

indicated that dhurrin accumulates in early stage of grain development reaching maximum amounts at 25 days after flowering but the grains were acyanogenic as demonstrated by lack of hydrogen cyanide and absence of transcripts encoding dhurrinases [277]. However, there is no evidence suggesting antimicrobial effect of dhurrin or hydrogen cyanide, which are generated during dhurrin biosynthesis.

GO enrichment analysis suggested that genes associated with photosynthesis were negatively regulated in the resistant genotype suggesting suppression of photosynthesis during enhanced defense responses. Therefore, disease resistant genotypes with good agronomic performance may harbor mechanisms that maintain the balance between defense and growth. Up-regulation of some genes that repress defense responses in the absence of pathogens is part of such mechanism. The plant hormone JA regulates inducible defenses, and plays a crucial role in growth-defense tradeoffs by regulating carbon assimilation and partitioning [278]. Interestingly, the resistant genotype shows induced expression of transcriptional repressors of JA and/ or defense responses such as JAZ proteins. Accumulation of JA in response to infection or other environmental cues promotes degradation of JAZ proteins that relieves repression on various transcription factors [231]. JAZ proteins suppress accumulation of anthocyanin by interacting with WD-Repeat/bHLH/MYB complexes while JA-induced degradation of JAZ proteins eliminates the interaction [279].

Pathogen inducible defense against major crop diseases is a vital component of resistance which may have less effect on resources that would rather be allocated to growth in the absence of pathogens. Although parts of this system is known in sorghum and other crop plants, regulation of pathogen induced defense mechanism is poorly understood. For instance, although the roles of sorghum 3-deoxyanthocyanidin phytoalexins in defense are known and that they are pathogen inducible but the upstream regulatory mechanisms that link pathogen perception to downstream target genes is unknown. In this regard, genetic evidence shows that the sorghum *Y1* and the *Tan1* genes are associated with resistance to grain mold [114, 151] and these genes regulate the biosynthesis of 3-deoxyanthocyanidin phytoalexins but the molecular link between pathogen perception and biosynthesis of the phytoalexins are not known. Based on evidences from this study and previous reports, we provide a conceptual model of pathogen inducible defense system in sorghum (Figure 4.11).

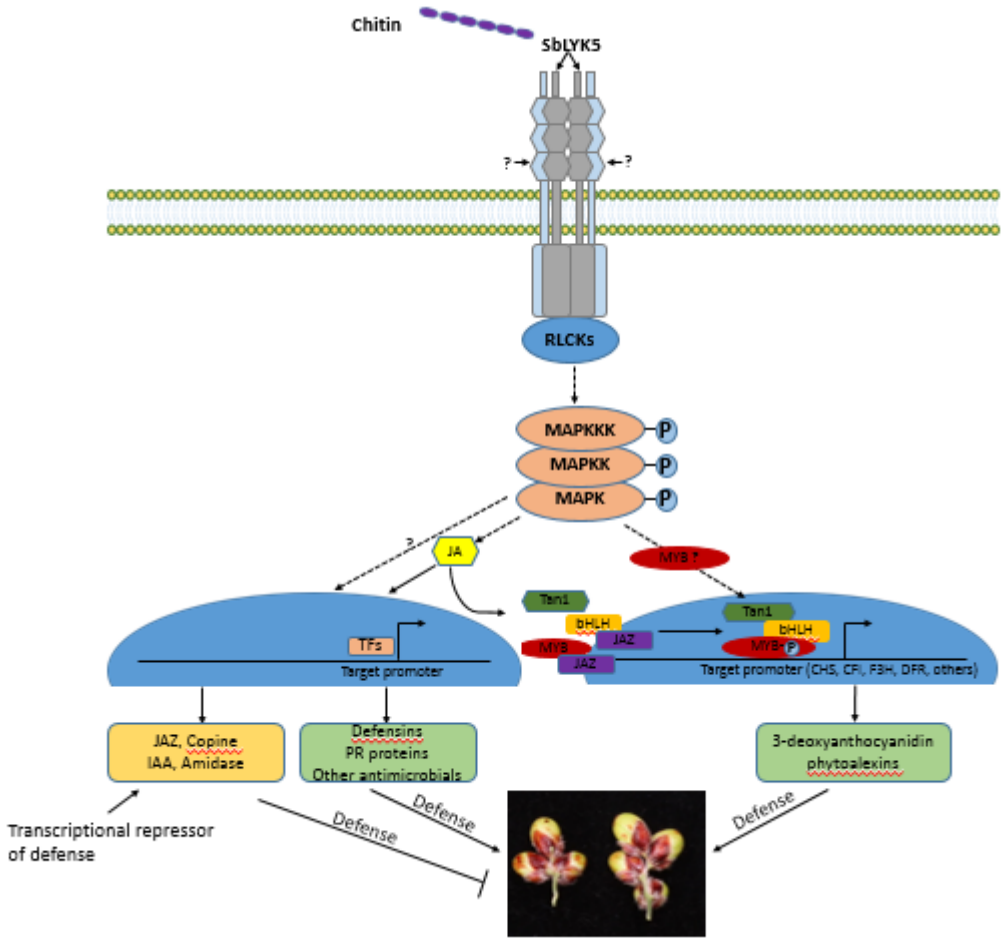


Figure 4.11. Proposed model for pathogen inducible defense system in sorghum grain and major components. The model depicts how a putative sorghum chitin receptor (*SbLYK5*) and an unknown coreceptor could be activated by perception of chitin that triggers a pathogen response signaling involving RLCKs, MAPKs, and various potential transcriptional regulators

As described in the preceding sections, we identified the *SbLYK5* gene that encodes a receptor like kinase that may function as the sorghum chitin receptor. *SbLYK5* and other RLKs likely serve as receptor complexes, and their downstream components such as RLCKs and MAPK are recruited in pathogen response signaling leading to gene expression and accumulation of defense active secondary metabolites. This is consistent with data from rice and Arabidopsis where MAPK cascades and their downstream transcription factors regulate phytoalexin biosynthesis [212]. Interestingly, a recent report suggests an R2R3 MYB transcription factor is phosphorylated by MPK4 which is required for light induced anthocyanin accumulation in Arabidopsis [213]. We therefore, speculate that the sorghum R2R3 MYB proteins encoded by *Y1* may be phosphorylated by an unidentified MPK or RLCKs in sorghum and play role in signaling of pathogen responses

and accumulation of secondary metabolites. The figure summarizes our working model of how the biosynthesis of 3-deoxyanthocynidin phytoalexins, defensins, PR proteins, and other antimicrobial peptides as well as defense suppressing proteins may be regulated through pathogen inducible defense system in sorghum grain.

4.6 Conclusion

Grain is a distinct tissue from the widely studied leaf tissue, contains rich carbon source that makes it prone to infection. Despite the importance of grain as the final and most valuable product of the crop production effort, genetic resistance and the status of defense responses in the grain have been poorly studied. Transcriptome profiling in the developing grain of sorghum genotypes revealed both conserved and unique defense mechanisms that may underlie differences in resistance to the disease. Differential expression of regulators of quantitative resistance, previously described in PTI pathways in leaf tissues, in other plant systems, were found to correlate with resistance in early stages of sorghum grain. In addition, JA response and biosynthesis pathways showed differential expression correlating with resistance extending the role of these plant hormone to grain tissue and complex diseases. These observations suggest that many responses in the grain are regulated by similar mechanisms that are active in leaf tissue despite the distinct nature of the leaf and grain tissues. By contrast, genes encoding pathogenesis-related proteins, defensins, phytoalexins and zein seed storage proteins, that are uniquely regulated in grain, and pathogen infection showed higher basal and induced expression in the resistant genotype. Interestingly, previously undescribed sorghum defensin genes that are induced upon infection or were constitutively expressed at a higher level in the resistant genotype were also identified. Further, we provide new insights into molecular, cellular, and biochemical processes underlying response to a complex disease involving a consortium of necrotrophic fungi with aggressive pathogenesis strategies, as well as a host resistance with complex genetic architecture. Potential regulators of sorghum pathogen recognition at the very early stages of attempted infection, and downstream genetic components of defense that may have antibiotic activities, or molecules that reinforce the grain structure to make it impermeable to pathogen ingress were identified. Together, the newly identified components may contribute to grain mold resistance and provide new insights into an understudied pathosystem, and will serve as targets for genetic studies and to identify resistance germplasm.

CHAPTER 5. FINE MAPPING AND IDENTIFICATION OF LINKED NBS-LRR GENES THAT CONFER BROAD SPECTRUM ANTHRACNOSE RESISTANCE IN SORGHUM

5.1 Abstract

Sorghum anthracnose caused by the hemibiotrophic fungus *Colletotrichum sublineolum* (*Cs*) is one of the most widespread diseases. Genetic resistance is the most cost effective and sustainable approach to control the disease. Despite availability of resistant germplasm and identification of multiple resistance loci, specific resistance genes and underlying resistance mechanisms are not well understood. Here we describe the identification of *ANTHRACNOSE RESISTANCE GENES* (*ARG4* and *ARG5*) encoding canonical nucleotide-binding leucine-rich repeat (NLR) receptors. *ARG4* was defined as a dominant resistance locus based on genetic studies in SAP135 that shows broad spectrum resistance to anthracnose. To identify *ARG4*, SAP135 was crossed to the susceptible line TAM428 and true-to-type anthracnose resistant or susceptible F3 families were used for BSA-seq analysis. Subsequent fine mapping narrowed the genomic region to a 1Mb region on chromosome 8. An independent but parallel study using RILs generated from a cross between the resistant sorghum line P9830 and TAM428 mapped *ARG5* to the same genomic region. Fine mapping using molecular markers, comparative genomic analyses, and gene expression studies revealed that *ARG4* and *ARG5* are resistance genes at two linked loci present in either SAP135 or P9830 lines. The corresponding loci in the reference genome BTX623 and the susceptible parent TAM428 carry susceptible alleles of *ARG4* and *ARG5* to the *Cs* strains *Csgl1* and *Csgrg*. Interestingly *ARG4* and *ARG5* are both located within clusters of duplicate NLR genes at these linked loci separated by ~1Mb genomic region. SAP135 and P9830 each carry only one of the *ARG* genes while having a susceptible allele at the second locus. RILs carrying either one of the two resistance alleles were resistant to both strains (*Csgl1*, and *Csgrg*) suggesting that *ARG4* and *ARG5* genes may have similar recognition specificities. Additional studies are required to determine specificity of these genes to other Cs strains and other pathogens. Overall, we identified two resistance genes that confer complete resistance to multiple srains of Cs with a potential for resistance breeding and molecular studies. We also demonstrate that the combined use of whole genome sequencing and biparental mapping populations are powerful tools to dissect complex loci and to mine the standing natural variation for resistance alleles.

5.2 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is among the most important cereals used for food, feed, biofuels and alcoholic beverages. It is a staple food crop in developing countries while in the developed world, it is mostly used as livestock feed, source of biofuels or alcoholic beverages. Sorghum anthracnose caused by the fungal pathogen *Colletotrichum sublineolum* is a major biotic constraint to its production. The pathogen mainly affects leaf where symptoms are more obvious but other plant parts are also affected. Symptoms include chlorotic flecks, acervuli formation, necrotic lesions and death of leaves. Host resistance is considered the most effective.

The sorghum conversion program [78] followed by identification of lines resistant to anthracnose [79] were the first set of studies that aimed at understanding inheritance of genetic resistance to the disease. These early programs lead to the identification of sorghum lines that showed resistance in diverse environments such as the sorghum line SC-748-5 [80]. The *Cg1* resistance gene carried by SC-748-5 was mapped to the distal end of chromosome 5 [81], but latter it was suggested that the resistance in SC-748-5 is not due to a single gene [82]. To understand the pathotypes of *C. sublineolum*, a set of 18 lines were developed as differentials [83] which served as sources germplasm for mapping of anthracnose resistant genes. The sorghum differential lines SC112-14, SC414-12E, SC748-5 and other sorghum lines (BK-7 and SC155-14E) were used to map anthracnose resistance loci [43, 81, 82, 84, 85]. These include three closely loci on distal end of chromosome 5 and one on chromosome 9. The four loci were consistently detected by both bi-parental mapping approaches and recent genome wide association studies [28, 86, 87]. The identified loci encompass large genomic regions that harbor genes from multiple defense response pathways [28]. This has complicated identification of candidates to a level of a few or single gene. Moreover, the identified QTLs explained only small proportion of the observed phenotypic variation [28, 87]. The source germplasms for the identified resistance alleles have mostly East African origin (Ethiopia and Sudan) while the detected QTLs are known to mediate resistance against pathotypes from USA where the studies were conducted [85]. Thus, mapping using germplasms with broad spectrum resistance to the disease from different origin may be vital to identify new loci potentially conferring resistance in diverse environments. Moreover, application of efficient next-generation genomic resources may improve mapping of anthracnose resistance loci to a single gene level.

In this study, SAP135 a sorghum line with West African origin and having a broad-spectrum resistance to anthracnose and rust is used to map anthracnose resistance genes. SAP135 (PI 576385 and SC 1070) was among the resistant accessions used in the recent GWAS [28, 87]. SAP135 has been found resistant to all the *C. sublineolum* isolates we have at Purdue (*Csgrg, Csgl1, Csgl2, Cs27 and Cs29*). Therefore, the major objectives of this study were to identify loci and candidate genes associated with anthracnose resistance in SAP135. The initial mapping experiments were done by a post-doctoral researcher Dr Gezahegn Girma (unpublished) while the fine mapping and gene identification studies were conducted as part of this thesis. Moreover, this study was to complete fine mapping and identification of candidate genes in a recently identified anthracnose resistance locus mapped using another resistant line P9830 [280]. Our study involving BSA-seq followed by fine mapping and gene identification coupled with next-generation genomic tools provides the identification of two tightly linked loci corresponding to each of the resistant lines with candidate genes discovered to a single gene level in both loci among clusters of duplicates. The two loci harbor resistant alleles of an NBS-LRR class, Sobic.008G166400 and Sobic.008G177900 carried by the two lines, SAP135 and P9830, respectively. Our study proves that efficient and rapid mapping approaches combined with advanced genomic tools may provide the way forward in understanding the genetic architecture of phenotypes of biological or agricultural importance.

5.3 Materials and methods

5.3.1 Plant materials

SAP135 is an accession collected from Nigeria, a breeding material maintained by Plant genetic Resources Conservation Unit, Griffin, GA identified by the accession number PI 576385. Other names for SAP135 include IS 17209C, SC 1070, NSL 365695 (pre-conversion). The line is described resistant to both anthracnose and rust (GRIN). Disease assays under controlled conditions in the greenhouse identified SAP135 to be resistant to all anthracnose strains available in the lab. To map anthracnose resistance loci from SAP135, a mapping population was created by crossing SAP135 to a susceptible inbred TAM428. The resulting F1 was selfed to generate F2 populations. Based on disease reaction at F2, 71 resistant and 68 susceptible individuals were selected and advanced for further evaluation at F3 stage to identify true-to-type families. A total

of 12 plants from each F3 family were planted and phenotyped to identify families that are true-to-type resistant or susceptible. This has identified 30 resistant and 50 susceptible true-to-type F3 families which were subsequently used for BSA-seq analysis and genetic mapping. Additionally, a new F2 population of 203 individuals were phenotyped and genotyped at anthracnose resistant locus identified in SAP135. These F2s are then advanced to F6 RILs for future mapping studies using different *C. sublineolum* strains because the resistant parent SAP135 has broad-spectrum resistance.

For fine mapping and identification of candidate genes associated with anthracnose resistance in P9830 [280], 80 F6 recombinant inbred lines (RILs), generated from a segregating population of a cross between P9830 and TAM428 were used. TAM428 was used as susceptible parent for both mapping populations.

5.3.2 Anthracnose disease assay

Fungal spores of *C. sublineolum* strains were cultured on an autoclaved (15 min at 121 ^{0}C) potato dextrose agar (PDA) and used for disease assays. The fungal cultures were kept under a continues florescence light for two weeks, after which spores were collected and used to inoculate 3 weeks old plants. The *C. sublineolum* strain *Csgrg* was used to map resistance in SAP135 x TAM428 population. Plants were sprayed with a spore suspension at a concentration of 1 x 10^{6} spores/ml. Inoculated plants were kept for 48 h in a humidity chamber set at a relative humidity of 70% and then transferred to greenhouse with an overhead misting system. Disease reaction of inoculated plants were scored as resistant or susceptible 4 to 6 days after the plants were transferred to the greenhouse condition. A different strain of *C. sublineolum*, *Csgl1* was used to map resistance locus in P9830. Therefore, the RILs used to narrow down this locus were inoculated with *Csgl1* while other disease assay conditions remained the same. To study the spectrum of resistance in the parental lines, the three parental lines and four other sorghum variants (RTx430, DS37, DS05 and DS25) were evaluated against a total of five strains including the two used for mapping (*Csgrg, Csgl1*, *Csgl2, Cs27* and *Cs29*).

5.3.3 DNA extraction

For whole genome re-sequencing and WideSeq of candidate genes, a high quality genomic DNA was isolated from a week old seedlings using DNeasy Plant Mini Kit (QIAGEN) following the manufacturers protocol. This includes DNA for the 30 resistant and 50 susceptible F3 families of SAP135 x TAM428, the parental lines other variant lines. Aliquots of this DNA was later also used for marker analysis and narrowing down the mapping region. For the 80 F6 RILs of P9830 x TAM428 and the new F2 population of SAP135 x TAM428, DNA was isolated using a simple and fast high-throughput DNA extraction method developed for PCR [281].

5.3.4 Whole genome re-sequencing

Equal amount of DNA from each of the 30 resistant and 50 susceptible F3 families were taken and pooled into resistant and susceptible groups and used for sequencing along with the two parental lines (SAP135 and TAM428). Whole genome re-sequencing of the resistant and susceptible pools, and SAP135 were completed at Purdue University Genomics Core facility on Illumina HiSeq 2500 while sequence for TAM428 was already available from a previous in-house sequencing effort (Lee et al., Unpublished).

5.3.5 Identification of resistance loci using BSA-seq analysis

To identify resistance loci associated with resistance in SAP135, BSA-seq analysis was conducted using whole genome resequencing of the two resistant and susceptible bulks of F3 families of the cross between SAP135 and TAM428. The sequences of the two bulks were used to conduct BSA-seq using the QTL-seq pipeline [93].

5.3.6 Visualization of target regions using Integrative Genomics Viewer (IGV)

Genome sequences of the two bulks and parental lines of both populations were aligned to the BTx623 reference and binary alignment map (BAM) files were generated for each of the bulks and parental lines. BAM files of target regions were cut out for visualization in IGV [282].

5.3.7 Marker development and fine mapping of target regions

Indel markers spanning the target regions were developed based on visualization of the genomic sequences in bam format of the two bulks and the parental lines. Primers flanking the identified indel markers were designed using the primer-blast tool at NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). PCR template size of 100 to 200 bp were targeted. The 80 F3 families and 80 F6 RILs corresponding to the two mapping populations, SAP135 x TAM428 and P9830 x TAM428, respectively were genotyped at selected indel markers to narrow down the target regions. Moreover, the new F2 populations of SAP135 x TAM428 were genotyped at selected markers in the QTL region and observed for marker-disease reaction association. PCR products were separated by agarose gel electrophoresis using 3% gel. List of the indel markers used to narrow down the QTL region and primers used to amplify the PCR based markers are in Table 5.1.

5.3.8 Use of whole genome re-sequenced sorghum lines

The advantages from the use of whole genome re-sequenced sorghum lines found in publicly available databases were utilized through phenotyping for disease reaction while their genome sequences retrieved from the public databases and visualized for patterns of polymorphism at candidate genes. This helped to get additional variants that carry alternate alleles at the candidate genes.

5.3.9 Sequencing of candidate genes

Candidate genes in the target regions were amplified from the parental lines and other variants by PCR and sequenced using the WideSeq service at Genomics Core at Purdue University, West Lafayette (Indiana). Genomic DNA for PCR amplification was extracted using DNeasy Plant Mini Kit (QIAGEN). Primers used to amplify the candidate genes and expression of selected genes are presented in Table A.1.

5.3.10 RNA extraction and gene expression analysis of candidate genes

RNA was extracted from 100 mg leaf tissues of the parental lines at 0, 48 and 72 h post inoculation (hpi). Total RNA was isolated using TRIzol (TRI Reagent). Expression of candidate

genes were studied through RT-PCR. cDNA was synthesized from 2 μg of total RNA using the AMV reverse transcriptase (NEB). Sorghum Actin gene was used as an endogenous control.

Table 5.1. Indel markers used to narrow down the QTL regions and the corresponding flanking primers used for PCR amplification

Name	Forward	Reverse	Population
ARG4-5600	GGGACTCCAACTGGATAGTGC	TGAGAGGTAGGATGCCCCTG	SAP135 x TAM428
ARG4-5890	TGATCCTTGCCGGTTTGTTTG	CTCTGGCTTTGTGGTCAAAATGAT	SAP135 x TAM428
ARG4-5945	CATCGTACAACGCCTAGCGA	AGTTTCCCACTCGTGGTGTG	SAP135 x TAM428 & P9830 x TAM428
ARG4-5998	AAGCCTCCATTGGATCCACTTAT	ACTCGGGCGAAAACTGATCT	SAP135 x TAM428 & P9830 x TAM428
ARG4-6008	ACATGATGTAATCACCGTGGAA	GTTACCAGCTGCGTTGTGTTG	SAP135 x TAM428
ARG4-6016	TTTGGCCTTCGATGCCAATATG	TGCCAGACATTCACTCCCCTA	SAP135 x TAM428
ARG4-6038	AGATACCGTCATCAAAATCCGGT	GGTTCAAGGTTCTCGCGTTG	SAP135 x TAM428
ARG4-6059	TGGGCGCACTCTATTTTCGG	ACCGGATAAACCGTCATCGG	SAP135 x TAM428 & P9830 x TAM428
ARG4-6062	CGACATCGTGTGAGTACGTGT	GTCGTCGCGACCATATGTAAC	SAP135 x TAM428
ARG5-6090	CGCCGCCGGTTAGAAAGTTAG	CCACCTGCTACTTGTACTGGG	P9830 x TAM428
ARG5-6117	GGGATATTTCCTCTTGTGCGG	GTGGCCGCCTGTACACTAC	P9830 x TAM428
ARG4-6131	GACTCCCTCGGAGACCTGTT	GTTCCTTGCCGCACACAAAA	SAP135 x TAM428 & P9830 x TAM428
ARG5-6200	TCTGCTCTGGCCATTTTGTGA	GTGGACAGCAATGTCTCCGTA	P9830 x TAM428
ARG5-6230	GTCCAGGCAAATCCCGTCTT	GGAATCTAGCTGTGCGGGAG	P9830 x TAM428

5.4 Results

5.4.1 Disease reaction of parental lines and variants at candidate genes

Spray inoculation and detached leaf disease assays showed that the sorghum line SAP135 is consistently resistant to all the strains of *C. sublineolum* while TAM428 is highly susceptible to all the strains tested (Figure 5.1).

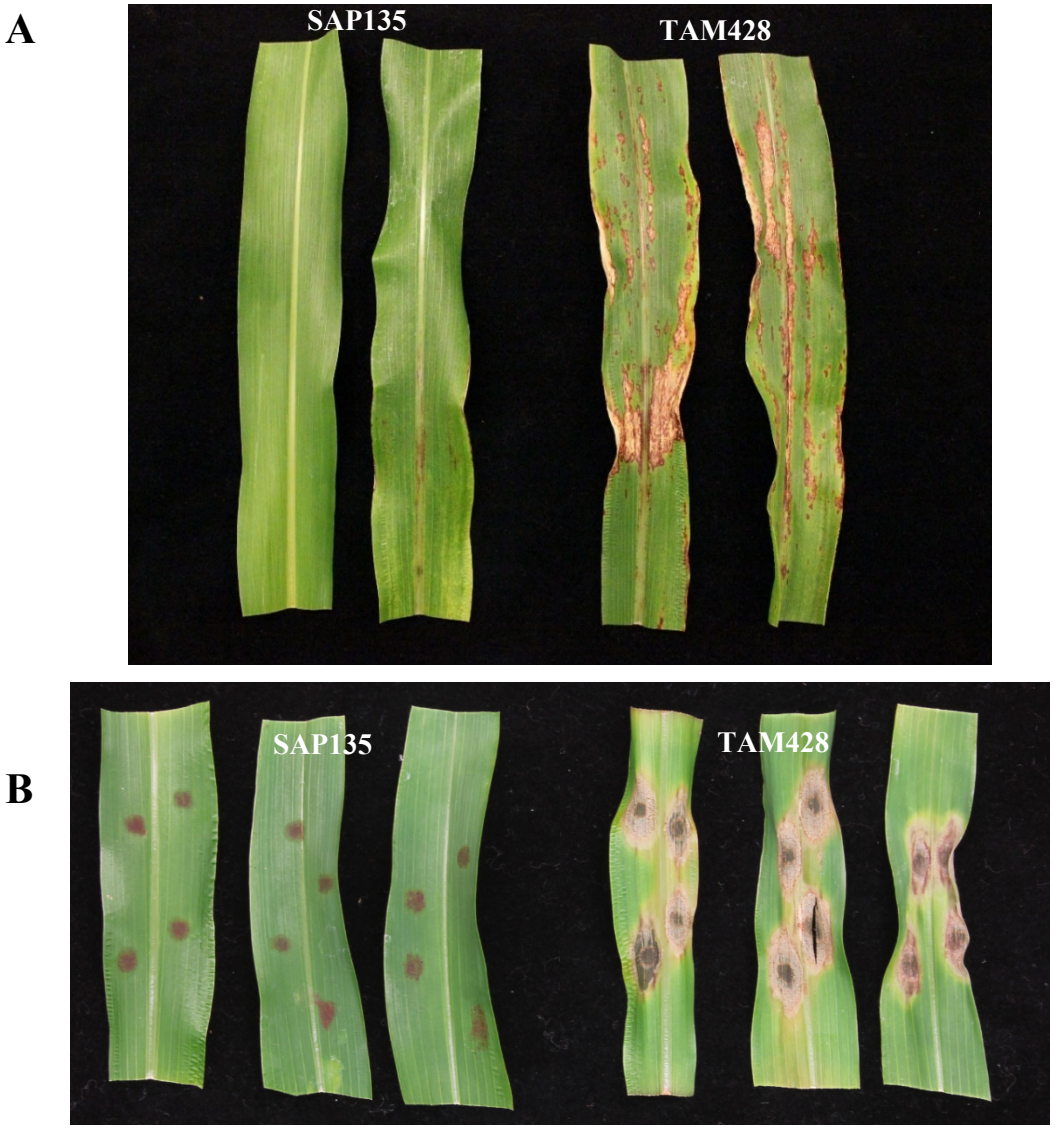


Figure 5.1. Responses of SAP135 and TAM428 to *C. sublineolum* strain *Csgrg* after spray inoculation and incubation in greenhouse (A) and detached leaf assay and drop inoculation (B)

Disease responses of the parental lines SAP135, P9830 and TAM428 as well as four other sorghum lines that are variants at candidate genes were tested using five strains of *C. sublineolum*. SAP135 was resistant to all the five strains (*Csgrg, Csgl1, Csgl2, Cs27 and Cs29*) while TAM428 was susceptible to all (Table 5.2). The other resistant parent used in the current study, P9830 is resistant to *Csgrg* and *Csgl1* but its responses to the other three were not clear. The four sorghum lines (RTx430, Ai4, BTx631 and SC23) that harbor variant alleles at candidate genes were all susceptible to the two strains used for mapping (*Csgrg* and *Csgl1*) and most of them showed susceptibility to the other three strains although some of the responses were not clear.

Table 5.2. Disease reaction of parental lines and variants to anthracnose strains used for mapping

Lines	**Anthracnose strains**				
	Csgrg	***Csgl1***	***Csgl2***	***Cs27***	***Cs29***
SAP135	R	R	R	R	R
P9830	R	R	?	S	?
TAM428	S	S	S	S	S
RTx430	S	S	S	S	S
Ai4	S	S	S	?	S
BTx631	S	S	?	?	?
SC23	S	S	S	?	S

? = disease responses were not clear

5.4.2 Mapping using BSA-seq

In the SAP135 x TAM428 population, Δ(SNP-Index) association analysis identified a single major peak for anthracnose resistance against the strain *Csgrg* on the distal end of chromosome 8 (Figure 5.2). The new locus is designated *ANTHRACNOSE RESISTANCE GENE 4* (*ARG4*), following our sequential naming of three other anthracnose resistance loci (*ARG1* to *ARG3*) identified recently using different mapping populations (unpublished). *ARG4* was mapped at ~ 60 Mbp region on chromosome 8. The segregation pattern of the SAP135 x TAM428 mapping populations varied between strains suggesting SAP135 carries multiple resistance loci conferring resistance to a number of strains. The current mapping is based on *Csgrg* which was highly virulent to TAM428, BTX623 but was avirulent to SAP135.

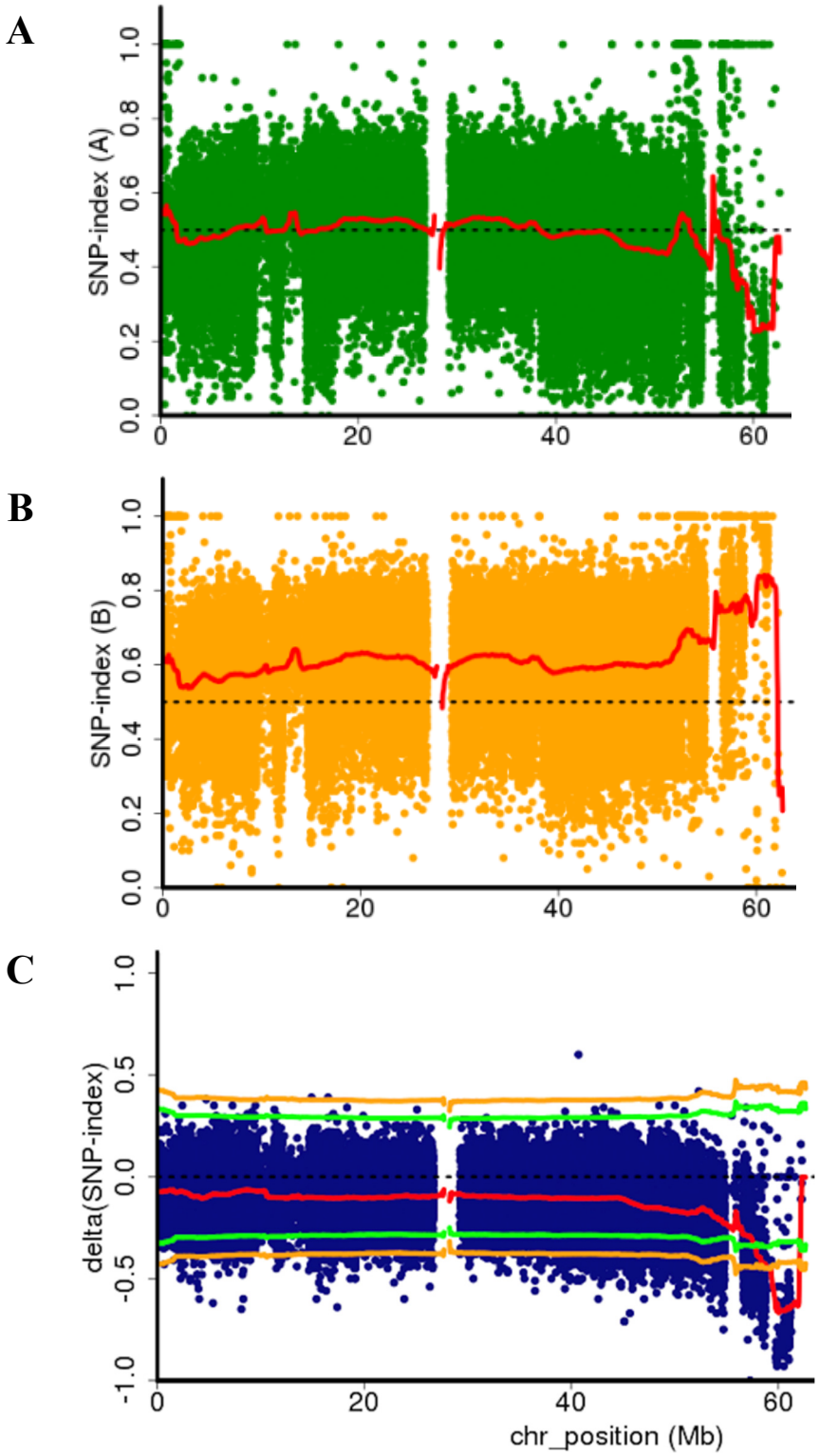


Figure 5.2. Identification of *ARG4* locus on chromosome 8 through BSA-seq analysis of resistant and susceptible pools of F3 families generated by crossing SAP135 and TAM428. A) SNP index of Rbulk. B) SNP index of Sbulk. C) Delta SNP index

5.4.3 Fine mapping of *ARG4* locus

Fine mapping of *ARG4* locus was conducted on 80 F3 families using 10 indel markers (Table 5.1) that spanned about 5 Mbp region. To delimit the location of *ARG4* locus, 23 recombinants were identified based on discordance by genotype and phenotype data (Table 5.3). Based on the number of recombinants, the locus is further narrowed down to a 928 kbp region delimited with ARG4-5945 and ARG4-6038 (Figure 5.3). The two flanking markers are physically located at 59.45 and 60.38 Mbp on the current BTx623 reference genome (Sorghum bicolor v3.1.1). The tightly linked three indel markers (ARG4-5998, ARG4-6008 and ARG4-6016) located inside the flanking markers showed no recombination and co-segregated with the phenotype. Moreover, to further confirm the resistance locus, 203 new F2 plants of SAP135 x TAM428 were genotyped at four of the indel markers including two of the inner markers and phenotyped using *Csgrg* strain that revealed the markers ARG4-5998 and ARG4-6016 showed 100% co-segregation with the resistance/susceptibility (Table 5.4).

Table 5.3. Fine mapping of *ARG4* locus: the green cells represent concordance and blue cells indicate discordance between phenotype and genotype data

No	Lines	Reaction	ARG4-5600	ARG4-5890	ARG4-5945	ARG4-5998	ARG4-6008	ARG4-6016	ARG4-6038	ARG4-6059	ARG4-6062	ARG4-6131
1	S6	S	2[a]	2	2	2	2	2	2	3	3	3
2	S7	S	3	3	2	2	2	2	2	2	2	2
3	S9	S	1	2	2	2	2	2	2	2	2	2
4	S12	S	2	2	2	2	2	2	2	2	2	3
5	S15	S	1	2	2	2	2	2	2	2	2	2
6	S17	S	3	2	2	2	2	2	2	2	2	2
7	S18	S	1	2	2	2	2	2	2	2	2	2
8	S23	S	2	2	2	2	2	2	2	2	2	3
9	S25	S	2	2	2	2	2	2	2	2	2	3
10	S26	S	3	2	2	2	2	2	2	2	2	2
11	S27	S	2	2	2	2	2	2	2	3	3	3
12	S28	S	3	3	2	2	2	2	2	2	2	2
13	S29	S	3	2	2	2	2	2	2	2	2	2
14	S30	S	1	1	3	2	2	2	2	2	2	2
15	S49	S	1	1	3	2	2	2	2	2	2	2
16	S52	S	2	2	2	2	2	2	2	2	2	3
17	S57	S	1	2	2	2	2	2	2	2	2	2
18	S58	S	3	2	2	2	2	2	2	2	2	2
19	S59	S	2	2	2	2	2	2	2	3	3	3
20	S64	S	3	3	2	2	2	2	2	2	2	2
21	R27	R	1	1	1	1	1	1	2	3	3	2
22	R67	R	2	3	3	3	3	3	3	3	3	3
23	R79	R	1	1	1	1	1	1	3	3	3	2
No. of recombinants			**14**	**5**	**2**	**0**	**0**	**0**	**1**	**3**	**3**	**9**

[a]1 = SAP135 allele, 2 = TAM428 allele, 3 = heterozygous.

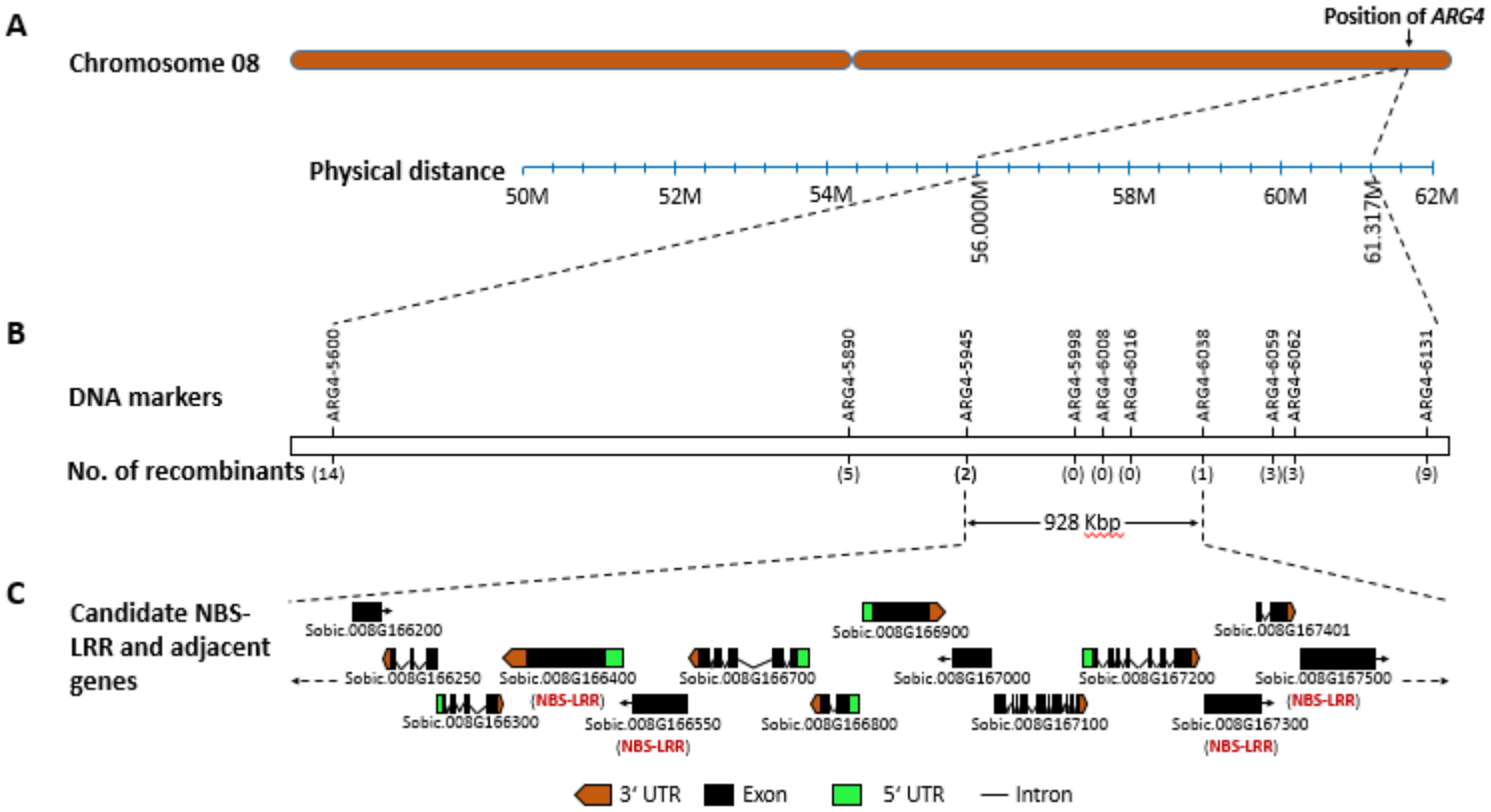


Figure 5.3. Genetic mapping of the *ARG4* locus. A) Physical position of *ARG4* locus on chromosome 8. B) Relative position of DNA markers used to narrow down the *ARG4* region and number of recombinants observed among F3 families of the cross between SAP135 and TAM428. C) Candidate NBS-LRR genes and partial illustration of adjacent genes with in the target region.

5.4.4 Resistance to *Csgrg* in SAP135 is controlled by dominant gene

Segregation ratios of the F2 populations (SAP135 x TAM428) for resistance to *Csgrg* in SAP135 showed resistance is controlled by a single dominantly inherited gene. Among the 203 F2 populations, 161 were resistant while the remaining 42 were susceptible to *Csgrg* (Table 5.4) indicating a 3:1 ratio which fits into a single dominant gene model (χ^2 value = 2.1). Moreover, resistance is qualitatively inherited although the genetic background appears to influence the level of resistance or susceptibility.

Table 5.4. Inheritance and validation of genetic resistance to *C. sublineolum* strain *Csgrg* in SAP135 based on selected indel markers in *ARG4* locus. Numbers in boxes represent the number of F2s under each phenotype and genotype groups.

Reaction	Genotype	Markers			
		ARG4-5945	**ARG4-5998**	**ARG4-6016**	**ARG4-6059**
Susceptible	1[a]	0	0	0	0
	2	38	42	42	40
	3	4	0	0	2
Sub-total		**42**	**42**	**42**	**42**
Resistant	1	62	65	65	66
	2	2	0	0	4
	3	97	96	96	91
Sub-total		**161**	**161**	**161**	**161**
Grand-total		**203**	**203**	**203**	**203**
		Single gene χ^2 value = 2.1*			

[a]1 = SAP135 allele, 2 = TAM428 allele, 3 = heterozygous; * = significant at $P < 0.05$.

5.4.5 Identification of candidate genes in *ARG4* locus

The reference sorghum genome BTx623 (v3.1.1) contained a total of 91 annotated genes within the mapped *ARG4* region, of which 67 are with known function. Among the 67 genes with predicted function, 5 genes were annotated as possibly associated with plant disease resistance that include four closely located NBS-LRR disease resistance genes and one gene encoding disease resistance-responsive (dirigent-like protein) family protein (Table 5.5).

Table 5.5. Candidate disease resistance genes in *ARG4* locus and their homologue in Arabidopsis and rice

No	Gene	Best-hit-arabi-name	arabi-defline	Best-hit-rice-name	rice-defline
1	Sobic.008G166400	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os06g49380.2	NBS-LRR disease resistance protein, putative, expressed
2	Sobic.008G166550	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os06g49380.4	NBS-LRR disease resistance protein, putative, expressed
3	Sobic.008G167300	AT3G14470.1	NB-ARC domain-containing disease resistance protein	LOC_Os04g53496.1	NBS-LRR disease resistance protein, putative, expressed
4	Sobic.008G167500	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os06g49380.6	NBS-LRR disease resistance protein, putative, expressed
5	Sobic.008G168800	AT5G42500.1	Disease resistance-responsive (dirigent-like protein) family protein	LOC_Os12g09700.1	Jacalin-like lectin domain containing protein, putative, expressed

There were four annotated NBS-LRR genes (Sobic.008G166400, Sobic.008G166550, Sobic.008G167300 and Sobic.008G167500) and a single dirigent-like gene (Sobic.008G168800). The four NBS-LRR genes are clustered in two closely located regions separated by about 150 kbp (Figure 5.3). Visualization of genomic sequences of SAP135, TAM428 and the two bulks indicated that the two parental lines are only polymorphic for three synonymous SNPs located inside the coding region of Sobic.008G168800. The three SNPs are A/C, T/G and G/A located at 60306054, 60306822 and 60306879 bp on sorghum reference genome (v3.1.1). Whereas, SAP135 and TAM428 are highly polymorphic for three of the NBS-LRR genes (Sobic.008G166400, Sobic.008G166550 and Sobic.008G167500). One of the NBS-LRR genes (Sobic.008G167300) is not polymorphic between SAP135 and TAM428 except a single synonymous SNP (T/G) at position 60141356. Therefore, the three classical disease resistance genes encoding putative NBS-LRR disease resistance protein are the most likely candidates associated with anthracnose resistance in SAP135.

5.4.6 Genomic organization, sequencing and analysis of gene expression of candidate NBS-LRR genes in *ARG4* locus

To further identify which of the three candidate NBS-LRR genes (Sobic.008G166400, Sobic.008G166550 and Sobic.008G167500) are responsible for the anthracnose resistance in SAP135, visualization of predicted gene structures, sequencing and gene expression studies were conducted. The predicted gene structure for the NBS-LRR candidates in the sorghum reference genome indicates that Sobic.008G166550 and Sobic.008G167500 have no predicted 5' and 3' UTR regions while Sobic.008G166400 has a complete gene structure including both UTR regions. Moreover, Sobic.008G166550 lacked a start codon in the sorghum reference genome, and TAM428. These observations suggest that Sobic.008G166550 and Sobic.008G167500 are pseudo genes. However, since the reference line, BTx623 which has 100% sequence similarity to TAM428 for all the candidate genes is susceptible to anthracnose, the lack of predicted UTR sequences in some of the candidates does not necessarily indicate that the genes have the same structure in SAP135. Therefore, sequencing of all the three candidates including potential 5' UTR region of Sobic.008G166550 available in SAP135 were conducted to understand the nature of polymorphisms between SAP135 and TAM428. IGV visualization of the region of Sobic.008G166550 using whole genome sequences from SAP135, TAM428 and the two bulks

(resistant and susceptible), indicated a possible deletion of large genomic region in TAM428, which was confirmed by WideSeq of genomic DNA from SAP135 and TAM428. The deleted region corresponds to the upstream sequence of the gene and 5' UTR sequences that are intact in the resistant SAP135 but deleted in TAM428. However, prediction of protein sequence from the SAP135 revealed that SAP135 contains a pre-mature stop codon in Sobic.008G166550 which excluded this gene from being the ARG4 candidate.

The remaining candidate genes, Sobic.008G166400 and Sobic.008G167500 share over 80% similarity in protein sequences although Sobic.008G167500 lacks predicted UTR region. Both IGV visualization and sequencing of both genes using genomic DNA of SAP135 and TAM428 indicated that the two lines are highly polymorphic for both genes. The fact that Sobic.008G167500 has no predicted UTRs suggests that its expression might be affected due to lack of upstream regulatory sequences. Gene expression study revealed that only Sobic.008G166400 is expressed while no transcripts were detected for Sobic.008G167500 (Figure 5.4). Therefore, Sobic.008G166400 is the most likely gene in *ARG4* locus conferring anthracnose resistance in SAP135 and hereafter designated *ANTHRACNOSE RESISTANCE GENE 4* (*ARG4*) gene.

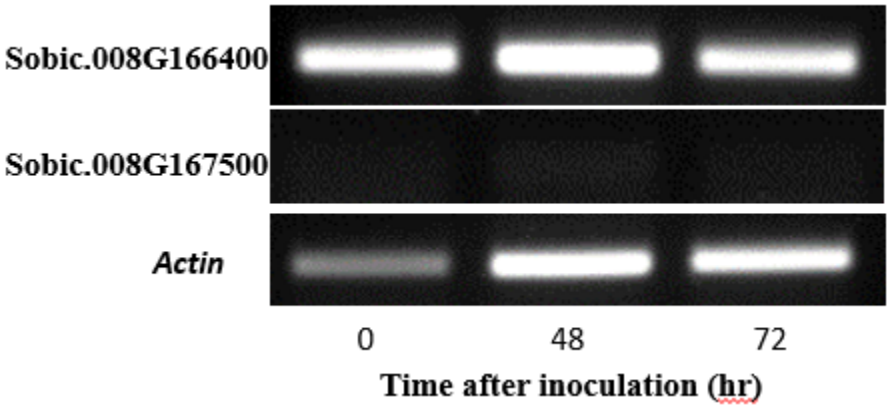


Figure 5.4. Gene expression analysis of candidate genes in *ARG4* locus. The sorghum Actin gene was used as a constitutive control

5.4.7 Polymorphisms, protein domain structure and variants of *ARG4* gene

ARG4 (Sobic.008G166400) contains a single exon that encodes 1555 amino acid protein. SAP135 and TAM428 share 95.6 and 92.6% nucleotide and amino acid identity, respectively, in

the coding region of *ARG4* gene. ARG4 is predicted to form coiled coil (CC), nucleotide-binding-site (NB-ARC), and leucine-rich-repeat (LRR) motifs (Figure 5.5) consistent with other NBS-LRR proteins. The 3D structure of *ARG4* protein was predicted using Protein Homology/analogY Recognition Engine V 2.0 (Phyre^2) web portal [283]. The 3D protein structure of *ARG4* indicates variation between that of SAP135 and TAM428. (Figure 5.5 and Figure A.1).

Additional sorghum variants were explored to identify additional alleles of the *ARG4* gene. Sequencing of genomic DNA of the sorghum line Ai4, which was highly susceptible to anthracnose strain *Csgrg* indicated that the line carries a frameshift mutation in *ARG4* gene due to an insertion of a single nucleotide at position 2767 that resulted in a pre-mature stop codon after amino acid number 930. Another sorghum line, BTx631, which is also susceptible to *Csgrg*, carries a different allele of *ARG4* gene with unique polymorphism patterns compared to all the variants studied (Figure A.2). BTx631 shares similarity with SAP135 for some of the nucleotides that are polymorphic between the variants but also displayed unique sequences at various positions of the gene. The sorghum line RTx430, which is also susceptible to *Csgrg*, was similar to that of TAM428 although there were variations for a few nucleotides.

5.4.8 Blast analysis for *ARG4*

BLASTP search of the NCBI data base (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using *ARG4* amino acid sequences from SAP135 revealed that the two genes from the BTx623 (Sobic.008G166400 [93% identity and 95% similarity]) and Rio (SbRio.08G185100 [73% identify and 81% similarity]) reference genomes as the most similar followed by a disease resistance protein RGA2-like from a grass species, green foxtail *Setaria viridis* (67% identity and 77% similarity). Sequence IDs and description of the homologues identified are presented in Table A.2. The next closely related homologues are also from grass families that include hypothetical and disease resistance proteins from foxtail millet (*Setaria italica*), white fonio (*Digitaria exilis*), and weeping lovegrass (*Eragrostis curvula*).

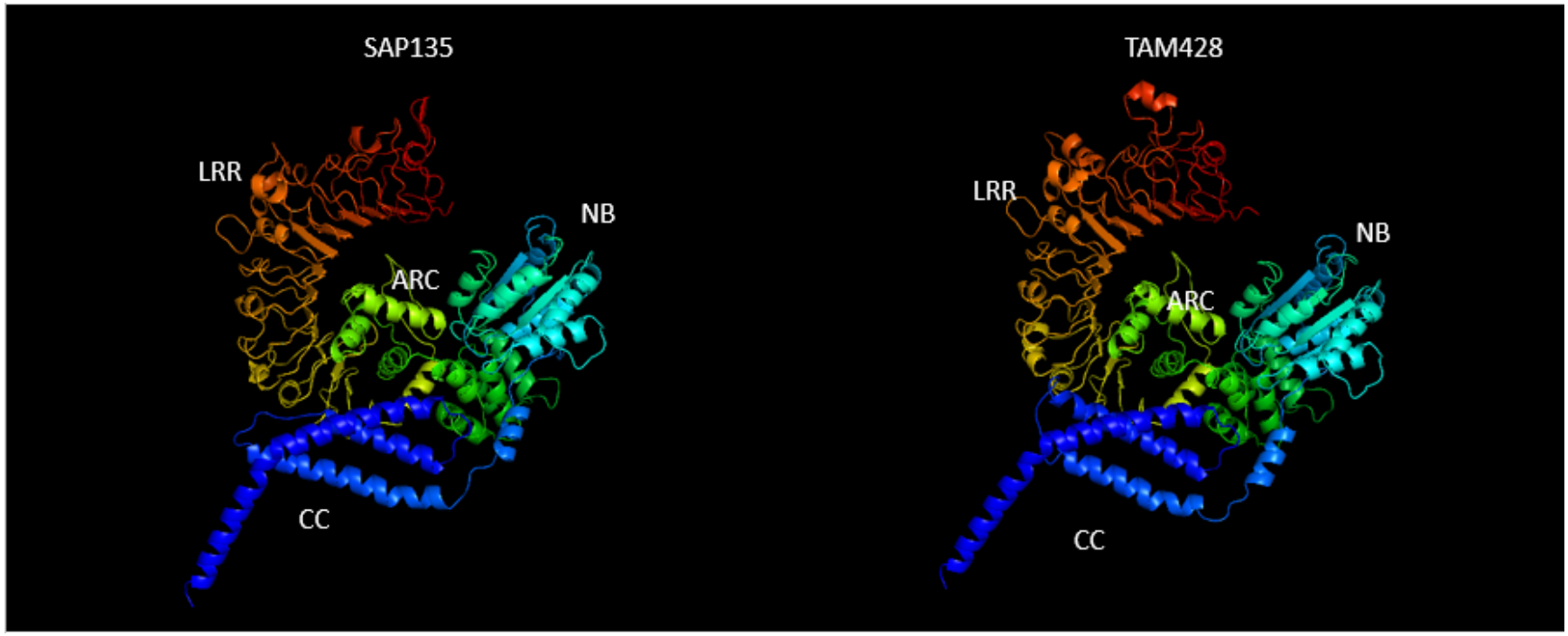


Figure 5.5. 3D protein structure of *ARG4* genes from SAP135 and TAM428

Homologues from major crops include NBS-LRR disease resistance proteins from barley (45% identity and 61% similarity), hypothetical protein from wheat (45% identity and 61% similarity) and a putative disease resistance protein RGA1 from rice (47% identity and 60% similarity). Moreover, the BLASTP search revealed another homologues protein from *Setaria italica*, described as "putative disease resistance protein At3g14460" with 45% identity and 60% similarity indicating At3g14460 is the homologue of *ARG4* from the model species *Arabidopsis thaliana*.

5.4.9 *ARG4* co-localized with a resistance locus identified in P9830

The two mapping populations, SAP135 x TAM428 and P9830 x TAM428 appear to identify the same region but it was not clear whether a single locus or tightly linked loci mediate anthracnose resistance in SAP135 and P9830. A rough mapping of anthracnose resistant locus to the *Csgl1* in P9830 x TAM428 was mapped to the same chromosomal region [280]. In parallel, based on a distinct resistant line the resistance in SAP135 to *Csgrg* was mapped to the same chromosomal region. Although the two different strains (*Csgrg* and *Csgl1*) were used to map the resistance loci, the fact that both SAP135 and P9830 are resistant to both strains suggest that the two lines may carry the same single gene conferring resistance to both strains. However, based on IGV visualization and sequencing of the candidate genes at the *ARG4* locus in the three parental lines, impactful polymorphisms were not detected between P9830 and TAM428 while SAP135 is highly polymorphic compared to both P9830 and TAM428. This is in agreement with the results of the other study [280] that there were no polymorphisms detected between P9830 and TAM428 for the candidates in *ARG4* locus. These observations suggested that the resistance locus in P9830 may be different from that in SAP135, which led to a new fine mapping of the QTL region using P9830 x TAM428 RILs including some of the markers that were used for fine mapping of *ARG4* using SAP135 x TAM428 mapping population.

5.4.10 Fine mapping of the resistance locus in P9830 using RILs

Based on fine mapping of the QTL region using 80 F6 RILs of P9830 x TAM428 and eight indel markers, four of which were shared markers with SAP135 x TAM428 population (Table 5.1), the QTL associated with resistance in P9830 was found to be a closely located new locus about 1

Mbp away from *ARG4* gene. This new resistance locus, conferring anthracnose resistance in P9830 is similarly designated as *ANTHRACNOSE RESISTANCE GENE 5* (*ARG5*). To delimit the location of *ARG5* locus, 24 recombinants were identified based on recombinants between markers and co-segregation with the phenotype data (Table 5.6). Based on the number of recombinant events, *ARG5* is mapped to a 415 Kb region delimited by indel markers ARG5-6090 and ARG4-6131. The two flanking indel markers are physically located at 60.9 and 61.3 Mbp, respectively, on the current BTx623 reference genome (Sorghum bicolor v3.1.1).

Table 5.6. Fine mapping of *ARG5* locus. The green cells represent concordance and blue cells indicate discordance between phenotype and genotype data

No	Lines	Reaction	ARG4-5945	ARG4-5998	ARG4-6059	ARG5-6090	ARG5-6117	ARG4-6131	ARG5-6200	ARG5-6230
1	8-11	S	2[a]	2	2	2	2	2	2	1
2	9-15	S	3	3	3	2	2	2	2	2
3	10-2	S	2	2	2	2	2	2	2	1
4	10-19	S	1	1	1	1	2	2	2	2
5	10-31	S	1	1	2	2	2	2	2	2
6	12-2	S	1	1	2	2	2	2	2	2
7	12-7	S	1	2	2	2	2	2	2	2
8	12-35	S	2	2	2	2	2	2	1	1
9	14-26	S	2	2	2	2	2	1	1	1
10	14-11	S	1	1	2	2	2	2	2	2
11	14-13	S	2	2	2	2	2	3	3	3
12	14-34	S	1	1	1	1	2	2	2	2
13	16-4	S	1	1	2	2	2	2	2	2
14	16-8	S	3	3	3	3	2	2	2	2
15	10-6	R	2	1	1	1	1	1	1	1
16	10-8	R	2	2	1	1	1	1	1	1
17	10-37	R	2	1	1	1	1	1	2	2
18	12-41	R	1	1	1	1	1	1	1	2
19	12-43	R	1	1	1	1	1	2	2	2
20	14-29	R	2	1	1	1	1	1	1	1
21	14-22	R	1	1	1	1	1	2	2	2
22	14-41	R	2	2	2	2	3	3	3	3
23	17-1	R	2	2	2	1	1	1	1	1
24	17-7	R	2	2	2	1	1	1	1	1
No. of recombinants			**16**	**12**	**7**	**4**	**0**	**4**	**6**	**9**

[a]1 = P9830 allele, 2 = TAM428 allele, 3 = heterozygous.

5.4.11 Identification of candidate genes in *ARG5* locus

Within the *ARG5* locus, a total of 60 annotated genes were identified in the reference sorghum genome (v3.1.1), of which 40 have predicted functions. Out of the 40 genes with predicted function, 7 genes were annotated to have plant disease resistance function which include a cluster of 5 predicted NBS-LRR disease resistance genes (Sobic.008G177900, Sobic.008G178200, Sobic.008G178300, Sobic.008G178500 and Sobic.008G178600) and 2 genes (Sobic.008G174966 and Sobic.008G175032) that encode leucine-rich repeat protein kinase family proteins (Table 5.7). The NBS-LRR genes show high sequence similarity and arranged in tandom with few genes in between (Figure 5.6). IGV visualization of the genomic sequences of resistant and susceptible bulks of P9830 x TAM428 population indicated that the 2 kinase family genes are not polymorphic between the two parental lines, therefore these two genes were not considered *ARG5* candidates. On the other hand, a major polymorphism for the NBS-LRR genes between resistant and susceptible bulks including a possible copy number variation in the NBS-LRR genes as was evident from a poor alignment to the reference of genomic sequences of some of the NBS-LRR genes from resistant bulk. To identify if there is variation in the number of the NBS-LRR genes between P9830 and TAM428, de-novo assembly of the genomic sequences of the resistant bulk was conducted. A resultant long node of 26 kbp sequence that include flanking sequences from both sides of the cluster of the 5 NBS-LRR genes available in the reference sorghum genome and alignment to the reference, revealed that there are only 2 of these genes available in the resistant parent P9830. The two NBS-LRR genes available in the resistant parent are Sobic.008G177900 and Sobic.008G178600. Therefore, these two are the only candidates for the ARG5 locus in P9830. Moreover, the Rio reference genome also showed only two of the NBS-LRR genes (SbRio.08G196000 and SbRio.08G196500) that correspond to the two present in the P9830 genome. Based on both the alignment to the reference genome of the de-novo assembled sequence and IGV view of the two candidates, the second candidate Sobic.008G178600 has a deletion of about 150 bp in its upstream region that includes part of its exon covering the start of the gene and promotor region. Hence, it was apparent that Sobic.008G177900 is the primary candidate although Sobic.008G178600 was further studied through gene expression to confirm that it is not a real candidate.

Table 5.7. Candidate disease resistance genes in *ARG5* locus and their homologue in Arabidopsis and rice

No	Gene	Best-hit-arabi-name	arabi-defline	Best-hit-rice-name	rice-defline
1	Sobic.008G174966	AT3G47570.1	Leucine-rich repeat protein kinase family protein	LOC_Os12g42520.1	receptor kinase, putative, expressed
2	Sobic.008G175032	AT3G47570.1	Leucine-rich repeat protein kinase family protein	LOC_Os12g42520.1	receptor kinase, putative, expressed
3	Sobic.008G177900	AT3G14470.1	NB-ARC domain-containing disease resistance protein	LOC_Os12g29710.1	NBS-LRR disease resistance protein, putative, expressed
4	Sobic.008G178200	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os12g29710.1	NBS-LRR disease resistance protein, putative, expressed
5	Sobic.008G178300	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os12g29710.1	NBS-LRR disease resistance protein, putative, expressed
6	Sobic.008G178500	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os12g29710.1	NBS-LRR disease resistance protein, putative, expressed
7	Sobic.008G178600	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein	LOC_Os12g29710.1	NBS-LRR disease resistance protein, putative, expressed

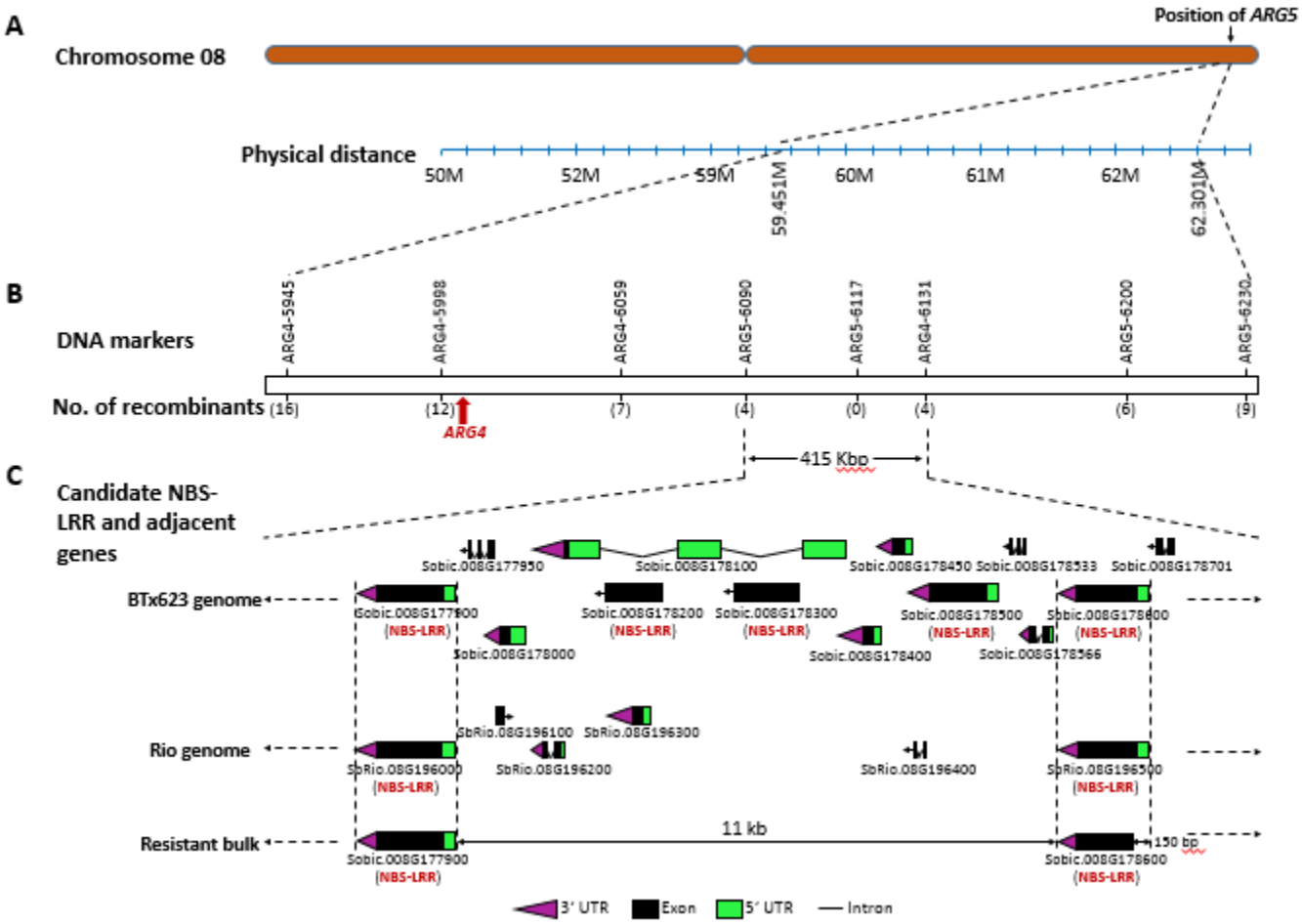


Figure 5.6. Mapping of the *ARG5* locus on chromosome 8, linked to *ARG4*. A) Physical position of *ARG5* locus on chromosome 8. B) Relative position of DNA markers used to narrow down *ARG5* locus and the number of recombinants identified from a set of recombinant inbred lines generated from a cross between P9830 and TAM428. C) Cluster of candidate NBS-LRR genes located within target region and copy number variation of the NBS-LRR genes between resistant bulk, BTx623 and Rio reference genomes.

5.4.12 Sequencing and expression of candidate NBS-LRR genes in *ARG5* locus

To further confirm that Sobic.008G177900 is the only candidate associated with anthracnose resistance in P9830, gene expression analysis was conducted for both Sobic.008G177900 and Sobic.008G178600. Because Sobic.008G178600 has a deletion in its regulatory region that might affect its expression, its expression was studied using the resistant parent P9830. This revealed that Sobic.008G178600 was indeed not expressed while Sobic.008G177900 is highly expressed in leaf tissues of the resistant parent P9830. Therefore, Sobic.008G177900 that encodes NBS-LRR disease resistance protein, hereafter referred as *ANTHRACNOSE RESISTANCE GENE 5* (*ARG5*) is the most likely gene conferring anthracnose resistance in P9830.

5.4.13 Polymorphisms, protein domain structure, and variants of *ARG5* gene

Similar to that of *ARG4*, *ARG5* has a single exon that encodes 1421 amino acid protein. P9830 and TAM428 have 94 and 90% nucleotide and amino acid identity, respectively. *ARG5* has coiled coil (CC), nucleotide-binding-site (NBS-ARC), and leucine-rich-repeat (LRR) domains (Figure 5.7). The 3D protein structure of *ARG5* from P9830 and TAM428 reveals a substantial variation between the two lines along all the motifs (Figure A.3).

Sequence of P9830 retrieved from de-novo assembly and WideSeq of genomic DNA from TAM428, SAP135 and two other variants (DS05 and DS25), which translated into amino acid sequences indicated that P9830 carries a unique variant of *ARG5* while TAM428, SAP135 and DS05 have high sequence similarity (Figure A.4). DS25 carries a different allele of *ARG5*.

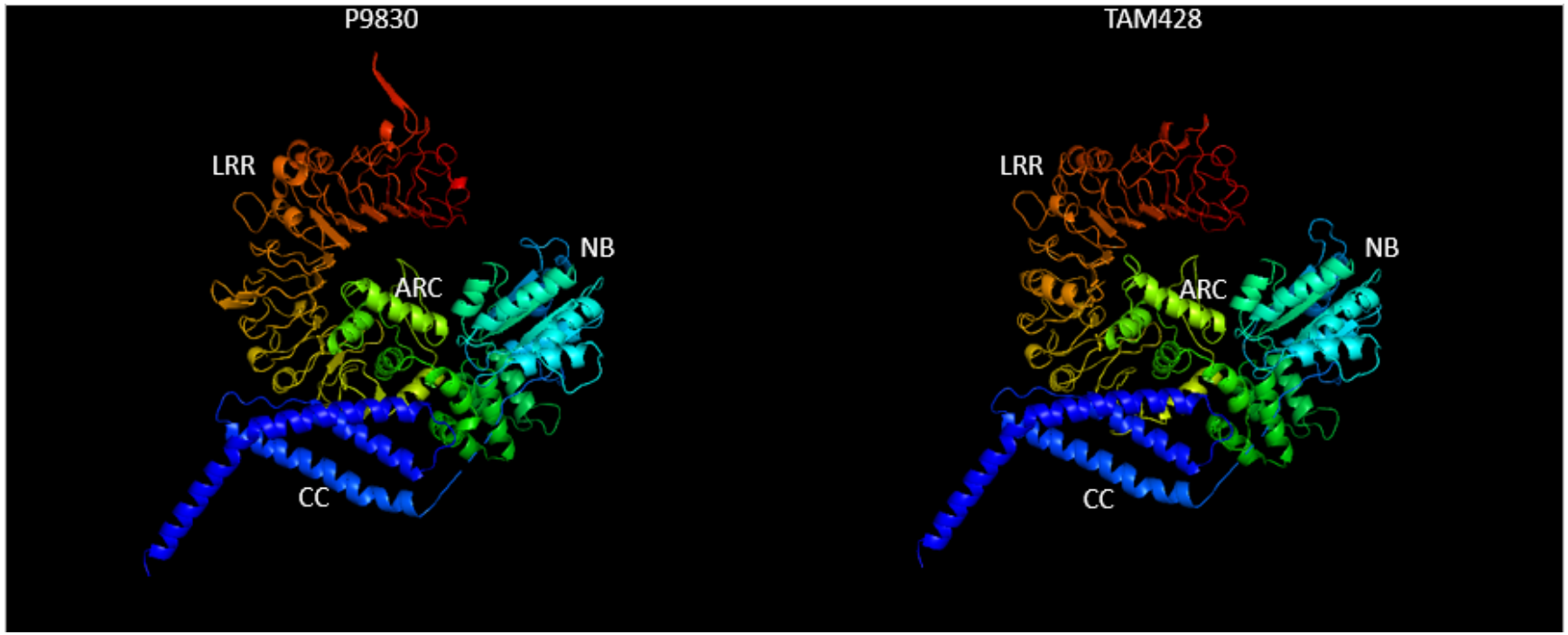


Figure 5.7. 3D protein structure of *ARG5* gene from P9830 and TAM428

5.4.14 Blast analysis for *ARG5*

BLASTP search of the NCBI database using the ARG5 amino acid sequences from P9830 revealed the two homologues of *ARG5* from Rio reference genome (SbRio.08G196000 [95% identity and 96% similarity], SbRio.08G195000 [84% identity and 89% similarity]) and the five *ARG5* duplicates from the BTx623 reference genome (Sobic.008G178300, Sobic.008G177900, Sobic.008G178600, Sobic.008G178500, and Sobic.008G178200) with identities and similarities ranging from 77-95% and 81-96%, respectively as the most related (Table A.3). Among the duplicates, the Sobic.008G178300 gene was described as "putative disease resistance protein At3g14460 isoform X1" with 91% identity and 93% similarity. Therefore, At3g14460 was found to be the homologue of both *ARG4* and *ARG5* from the model species *Arabidopsis thaliana*. A hypothetical protein from weeping lovegrass, *Eragrostis curvula* was identified as the next most similar (65% identity and 76% similarity). The additional homologues were identified from rice (50% identity and 63% similarity), various grass species (46 to 47% identity and 60 to 65% similarity), barely (47% identity and 62% similarity) and wheat (46% identity and 59% similarity).

5.4.15 Genotype by strain interaction of RILs

Reaction of randomly sampled RILs from the SAP135 and P9830 populations in response to strains of *C. sublineolum* indicated that lines resistant to either one of the two mapping strains *Csgrg* and *Csgl1* are also resistant to the other. Lines carrying either of the two resistance genes confer resistance to both *Csgrg* and *Csgl1*. This is consistent with the reaction of the two resistant parents, SAP135 and P9830 which are both resistant to the two strains. Additional experiments are required if this concurrence extends to other *C. sublineolum* strain and pathogens.

5.5 Discussion

Sorghum's immune response to the hemibiotrophic anthracnose fungus *C. sublineolum* and their genetic control are poorly understood. Despite the identification of three anthracnose resistance loci which are all located in the distal ends of chromosome 5 [81, 82, 84, 85] and one locus on chromosome 9 [43, 85], candidate genes and underlying mechanisms of resistance associated with such loci are not known. Recent studies using multi-parent mapping populations [28, 86, 87] redetected the same loci, which were initially detected by bi-parental mapping

approaches. Hence, the discovery of new anthracnose loci has paused for years. In this study, we used a combination of genomic approaches leading to the identification of two new tightly linked loci (*ARG4* and *ARG5*) which harbor resistant alleles in candidate NBS-LRR genes, Sobic.008G166400 and Sobic.008G177900 carried by two lines, SAP135 and P9830, respectively. Both candidates were identified from clusters of tightly linked and sequence related homologous that encode disease resistance proteins containing coiled coil (CC), nucleotide-binding-site (NB-ARC), and leucine-rich-repeat (LRR) domains. *ARG5* locus is co-localized with a previously reported leaf rust resistant locus in sorghum, *Rust locus 4 [75]*. Moreover, the clusters of disease resistance genes from both loci were reported as having significant homology to the wheat rust resistance protein *Lr1* [75] while blast analysis indicated that both *ARG4* and *ARG5* genes share homology to an *Arabidopsis thaliana* leucine-rich repeat (LRR) protein (At3g14460; AtLRRAC1) adenylyl cyclase [284]. Thus, *ARG4* and *ARG5* alleles from the resistant parents may serve as novel sources of resistance against anthracnose and provide opportunity to understand mechanisms of anthracnose resistance in sorghum. The following section provides an in-depth discussion of origin of *ARG4* and *ARG5* resistant alleles, approaches employed to identify the genes, possible mechanisms of anthracnose resistance conferred by *ARG4* and *ARG5* and distribution and mapping of disease resistance genes in sorghum.

SAP135, the line used to map *ARG4* locus, also known by its accession number PI 576385 in GRIN database and other names such as IS 17209C, SC 1070 and NSL 365695 (preconvertion) showed broad-spectrum resistance to anthracnose strains based on our assays and GRIN database data associated with this accession. It was also reported as resistant to anthracnose along with other anthracnose resistant accession in the sorghum association panel (SAP) [87]. It is originated from Nigeria and unrooted neighbor-joining tree of resistant accessions present in SAP and NPGS Ethiopian germplasm revealed additional related accessions, PI 534079, PI 533871, PI 534071 from the same origin [87] and PI 534037 from Chad [28]. Therefore, the anthracnose resistant allele carried by SAP135 has most probably originated from West African sorghum germplasm while the previously detected resistant alleles on chromosome 5 and 9 have origins mostly in East Africa (Ethiopia and Sudan). On the other hand, P9830, the line that carries the resistant allele of *ARG5* gene is a line in Gebisa Ejeta's lab at Purdue University, but its origin was not identified.

Although SAP135 is among the resistant accessions included in previous GWAS studies [28, 87], significant peaks associated with resistance alleles from SAP135 and related accessions

might not have been detected. It is possible that such accessions may carry resistance alleles in the loci detected by GWAS, but at least we know that loci like *ARG4* may remain undetected. *ARG4* was mapped using a virulent strain from Georgia, USA while the QTL on chromosome 9 confers resistance against pathotypes also from Georgia and Texas [85]. These pose questions over efficiency of multi-parent based mappings to detect rare resistant alleles particularly associated with phenotypes which are mostly qualitative in nature. Those detected so far by multi-parent approaches such as GWAS are more like QTLs with basal resistance as it is evident from the small explained phenotypic variation associated with resistant alleles [28]. Therefore, bi-parental approaches may still be more powerful to detect novel resistant alleles available infrequently. Moreover, segregating generations from the bi-parental crosses are vital to fine map detected regions which can be impractical with multi-parent approaches.

The disease resistant genes, Sobic.008G166400 and Sobic.008G177900 in *ARG4* and *ARG5* loci, respectively were identified through genomic approaches that involved BSA-seq, Integrative Genomics Viewer (IGV) visualization of genomic data of the candidate region from parental lines, the resistant and susceptible bulks and other variants, narrowing down of candidate region through traditional genetic mapping using molecular markers, and next generation sequencing and gene expression analysis of candidate genes. Bulk segregant analysis combined with whole genome resequencing (BSA-seq) is a rapid mapping approach [93] which is becoming a common approach to map QTLs and qualitative traits in recent years [98-103]. Integrative Genomics Viewer (IGV), which is a high performance and simple tool to interactively visualize genomic data [282] was used to explore the next generation sequence data from the parental lines and bulks as well as re-sequenced sorghum lines from public databases. This has enabled visualization of the candidate region and key polymorphisms between the parental lines for the candidate genes and helped to easily identify indel markers for marker analysis. Once the candidate regions in both loci were narrowed down and candidate genes identified, the candidates from resistant and susceptible parents and additional variants were sequenced using the WideSeq service at Purdue University (https://www.purdue.edu/hla/sites/genomics/wideseq-2/). The WideSeq service at the university is an efficient and inexpensive Next Generation Sequencing (NGS) approach that involves construction of NGS library from a target of up to 100 kb and which then sequenced and the reads are assembled back into wide sequences. Such capacity to sequence wider genomic regions was instrumental to identify the candidates among a tightly linked and sequence

related cluster of NBS-LRR genes which otherwise would have been difficult to differentiate among the candidates via the conventional sequencing approaches such as Sanger and even illumina sequencing where the short sequence reads may align to the different genes. Sequencing of homologous genes with high similarity could be challenging [28]. Moreover, the gene expression analysis of candidate genes revealed that some of the duplicates lacked expression, which may be due to alterations in regulatory regions during event of duplication leading to pseudogenization of these genes. Therefore, application of combination of genomic tools is essential to identify genes particularly found at complex loci.

The fact that both *ARG4* and *ARG5* loci contains cluster of highly similar NBS-LRR genes demonstrates the commonly observed phenomenon of disease resistance genes which often found in clusters [28, 285-288]. However, except the two genes Sobic.008G166400 and Sobic.008G177900, each found on *ARG4* and *ARG5* loci respectively, most of the candidates appeared as non-functional duplicates which are either truncated, lack regulatory sequences or not expressed. Truncated or pseudogenes of NBS-LRR class have been reported in many plant genomes [289, 290]. Moreover, *ARG5* locus contains variable number of duplicates between the resistant (P9830) and susceptible (TAM428) parents. Only two copies of the *ARG5* candidates were present in P9830 and the Rio reference genome while five copies were identified in TAM428 and the BTx623 reference genome.

The 3D protein structure of *ARG4* and *ARG5* from the corresponding resistant parents SAP135 and P9830 as well as the susceptible TAM428 revealed variation in 3D conformation. The 3D structure of *ARG4* protein from SAP135 and TAM428 showed variation mostly in their LRR region while variation in the other domains was not obvious. Whereas, the predicted 3D structure of *ARG5* from P9830 and TAM428 showed a clear variation. Interestingly, the predicted protein structures from the two resistant parents SAP135 and P9830 indicated a similar pattern toward their c-terminal region which is different in the susceptible TAM428. However, it was not clear whether such c-terminal modification is associated with resistance/susceptibility to the pathogen or other conformational changes are more important.

Plant NBS-LRR proteins, also referred, as NB-LRR or NB-ARC-LRR proteins are widely known for their key role in immune responses particularly through detection of race specific pathogen effectors [291]. The detection occurs indirectly from modifications of the host virulence target by virulence/effector proteins or some NBS-LRR proteins directly bind pathogen effector

proteins [291]. Conformational changes to the NBS-LRR proteins associated to their interaction with pathogen or altered host proteins leads to hydrolysis of ATP to ADP by the NBS domain, which in turn activates downstream process leading to resistance. Both *ARG4* and *ARG5* genes Sobic.008G166400 and Sobic.008G177900 encode NBS-LRR family proteins and blast analysis further revealed that they share homology to the *Arabidopsis thaliana* NBS-LRR protein (At3g14460; AtLRRAC1), characterized as adenylyl cyclase (ACs) that catalyzes the formation of the second messenger cAMP from ATP [284]. Knock-out mutants of Arabidopsis LRRAC1 are compromised in immune responses to fungal pathogens [284]. cAMP is thought to activate the NBS-LRR downstream signaling [292]. Interestingly, cAMP signaling is involved in elicitor-induced phytoalexin accumulation [293], a known antimicrobial molecule involved in sorghum's defense response against major diseases. cAMP is associated with induction of phenylalanine ammonia lyase (PAL) which in turn is involved in the production of phytoalexins and salicylic acid (SA) [294, 295]. The amino acid sequences of both *ARG4* and *ARG5* genes contain motifs that are highly similar to the identified AC motifs of LRRAC1 (At3g14460). Such evidences are consistent with sorghum's immune response mechanism against fungal pathogens and thus the *ARG4* and *ARG5* genes likely play resistance function through AC activity, which leads to accumulation of cAMP, an important signaling and response molecule in plants [296]. However, further studies are required to validate the AC activities of *ARG4* and *ARG5* genes.

A total of 346 NBS-encoding genes were identified in the sorghum genome distributed unevenly across the ten chromosomes [113]. Based on *Arabidopsis* and rice functional annotations for the sorghum homologues, we identified a total of 376 NB-ARC domain encoding genes in the reference genome of *Sorghum bicolor* v3.1.1, of which 117 (31%) are located on chromosome 5 followed by 55 (15%) and 50 (13%) on chromosomes 2 and 8, respectively (Figure 5.8). Many of these are annotated as NBS-LRR indicating they contain both NBS and LRR domains but all of them have at least the NB-ARC domain. Nearly a third of the disease resistance genes are located on chromosome 5 which explains why disease resistance QTLs are frequently detected on this chromosome [28, 81, 82, 84-87]. Chromosome 8, where the two new NBS-LRR genes (*ARG4* and *ARG5*) were identified also contains the third highest number of disease resistance genes after chromosome 2.

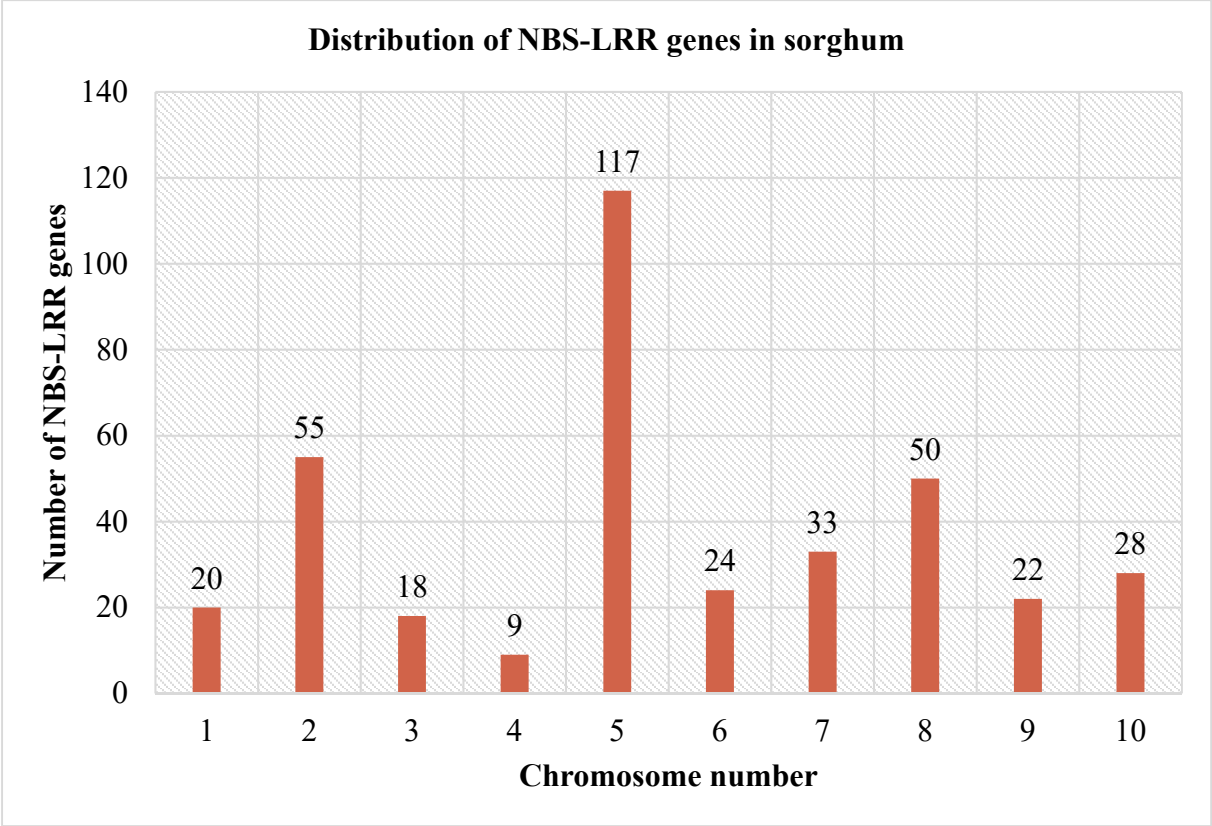


Figure 5.8. Distribution of NBS-LRR genes across sorghum chromosomes

5.6 Conclusion

Sorghum anthracnose caused by the fungus *C. sublineolum* is a major disease and host resistance mechanisms are poorly understood. Owing to the nature of the pathogen and complexity of anthracnose resistance loci, genes associated with resistance are not identified. Using BSA-seq combined with next-generation genomic resources, we identified two new tightly linked loci (*ARG4* and *ARG5*) containing candidate NBS-LRR genes, Sobic.008G166400 and Sobic.008G177900, respectively. The *ARG4* and *ARG5* resistant alleles are carried by the sorghum lines SAP135 and P9830, respectively with a likely redundant function. Although both genes were found in cluster with highly similar duplicates, most of them were truncated or were pseudogenes. Moreover, the resistant parent in *ARG5* locus, P9830 contained only two duplicates of the candidate while the susceptible parent TAM428 carried five copies. For both genes, the susceptible parent TAM428 has altered 3D protein structure compared to the resistant parents SAP135 and P9830 suggesting the susceptibility in TAM428 might be associated with failure to detect pathogen

virulence proteins from the particular isolates used to map the locus. Each of the two lines carry resistant allele only in one of the loci while having a susceptible allele in the other. Both SAP135 and P9830 showed broad spectrum resistance against anthracnose isolates. The candidate genes share homology to rust resistance genes. Thus, the *ARG4* and *ARG5* resistant alleles may confer broad-spectrum disease resistance and potentially conferring resistance to both sorghum anthracnose and rust. Moreover, blast analysis revealed that the *ARG4* and *ARG5* genes may have similar disease response function with that of an NBS-LRR gene from the model plant Arabidopsis characterized as adenylyl cyclase. Such similarities with a well studied system provide a new avenue for understanding the basis of anthracnose resistance in sorghum. Overall, we discovered a major anthracnose resistance region on sorghum chromosome 8 containing two tightly linked loci and each with a candidate NBS-LRR gene that confer broad-spectrum resistance to sorghum anthracnose and potentially to rust. The study reveal new opportunities to gene identification in complex host-pathogen systems and understanding the underlying genetic and molecular mechanisms of disease resistance associated with NBS-LRR class of genes.

APPENDIX

Table A.1. List of primers used for PCR amplification and gene expression study of target genes

Gene	Forward	Reverse	Target
Sobic.008G166550	GTGAGCCCTCAGATTGTTGA	GTGGTTCCTCTGTGCCCTTTA	Upper half of gene and upstream deletion
Sobic.008G166550	ATTTAGGGGACTCCATGGTTGC	GAACTCGGCTTTGAAGATGCG	Lower half of gene and downstream sequences
Sobic.008G166400	ACAAGCACACATCTCCTCGG	TGTACAGGGTGGAAGCAAGC	Whole ORF (ATG to stop) in TAM428 and P9830
Sobic.008G166400	ACAAGCACACATCTCCTCGG	AGAGAGAGAATCCAAAGGACAGA	Whole ORF (ATG to stop) in SAP135, DS37, DS05 and RTx430
Sobic.008G167500	GTTCCATTTCATCAGTACGCAGG	GCGGGTTGCCCAGAAGAAA	Whole ORF (ATG to stop)
Sobic.008G166400	TCGAGTGTATCAGATTTGTTGTCT	TGAACTTGGGAGACCATCCTTG	Gene expression study
Sobic.008G167500	GGTCTGGACATGTGTATCACTC	ACTACTTTCACCATCAACCTTGG	Gene expression study

Table A.2. BLASTP analysis of *ARG4* gene

Sequence ID	Description	Identity (%)	Similarity (%)
XP_021301663.1	Putative disease resistance protein RGA3 [Sorghum bicolor] (Sobic.008G166400)	93	95
KAG0521717.1	Hypothetical protein BDA96_08G185100 [Sorghum bicolor] (SbRio.008G185100)	73	81
XP_034586954.1	Disease resistance protein RGA2-like [Setaria viridis]	67	77
RCV19468.1	Hypothetical protein SETIT_3G387100v2 [Setaria italica]	67	77
KAF8655510.1	Hypothetical protein HU200_061054 [Digitaria exilis]	59	71
CAD45027.1	NBS-LRR disease resistance protein homologue [Hordeum vulgare]	45	61
TVU16126.1	Hypothetical protein EJB05_39677, partial [Eragrostis curvula]	44	59
KAF6988201.1	Hypothetical protein CFC21_005773 [Triticum aestivum]	45	61
XP_012702682.2	Putative disease resistance protein At3g14460 [Setaria italica]	45	60
VAH14370.1	Unnamed protein product [Triticum turgidum subsp. durum]	45	60
XP_015644152.1	Putative disease resistance protein RGA1 [Oryza sativa Japonica Group]	47	60
XP_037454969.1	Disease resistance protein RGA2-like [Triticum dicoccoides]	45	60
XP_025880760.1	Disease resistance protein RGA2 [Oryza sativa Japonica Group]	44	58

Table A.3. BLASTP analysis of *ARG5* gene

Sequence ID	Description	Identity (%)	Similarity (%)
KAG0521840.1	Hypothetical protein BDA96_08G196000 [Sorghum bicolor] (SbRio.08G196000)	95	96
XP_002443597.2	Putative disease resistance protein At3g14460 isoform X1 [Sorghum bicolor] (Sobic.008G178300)	91	93
XP_021301943.1	Putative disease resistance protein RGA4 [Sorghum bicolor] (Sobic.008G177900)	91	93
KAG0521845.1	Hypothetical protein BDA96_08G196500 [Sorghum bicolor] (SbRio.08G195000)	84	89
XP_002443600.1	Putative disease resistance protein RGA3 [Sorghum bicolor] (Sobic.008G178600)	83	89
XP_002443598.1	Putative disease resistance protein RGA4 [Sorghum bicolor] (Sobic.008G178500)	78	85
OQU79678.1	Hypothetical protein SORBI_3008G178200 [Sorghum bicolor] (Sobic.008G178200)	77	81
TVU50286.1	Hypothetical protein EJB05_01652, partial [Eragrostis curvula]	65	76
KAF2907818.1	Hypothetical protein DAI22_12g128800 [Oryza sativa Japonica Group]	50	63
XP_037454969.1	Disease resistance protein RGA2-like [Triticum dicoccoides]	47	60
XP_034606154.1	Putative disease resistance protein RGA3 [Setaria viridis]	47	60
KAF8756297.1	Hypothetical protein HU200_011117 [Digitaria exilis]	49	65
XP_012703299.1	Disease resistance protein RGA2 [Setaria italica]	47	61
VAH52161.1	Unnamed protein product [Triticum turgidum subsp. durum]	46	61
SPT20661.1	Unnamed protein product [Triticum aestivum]	46	59
CAD45027.1	NBS-LRR disease resistance protein homologue [Hordeum vulgare]	47	62

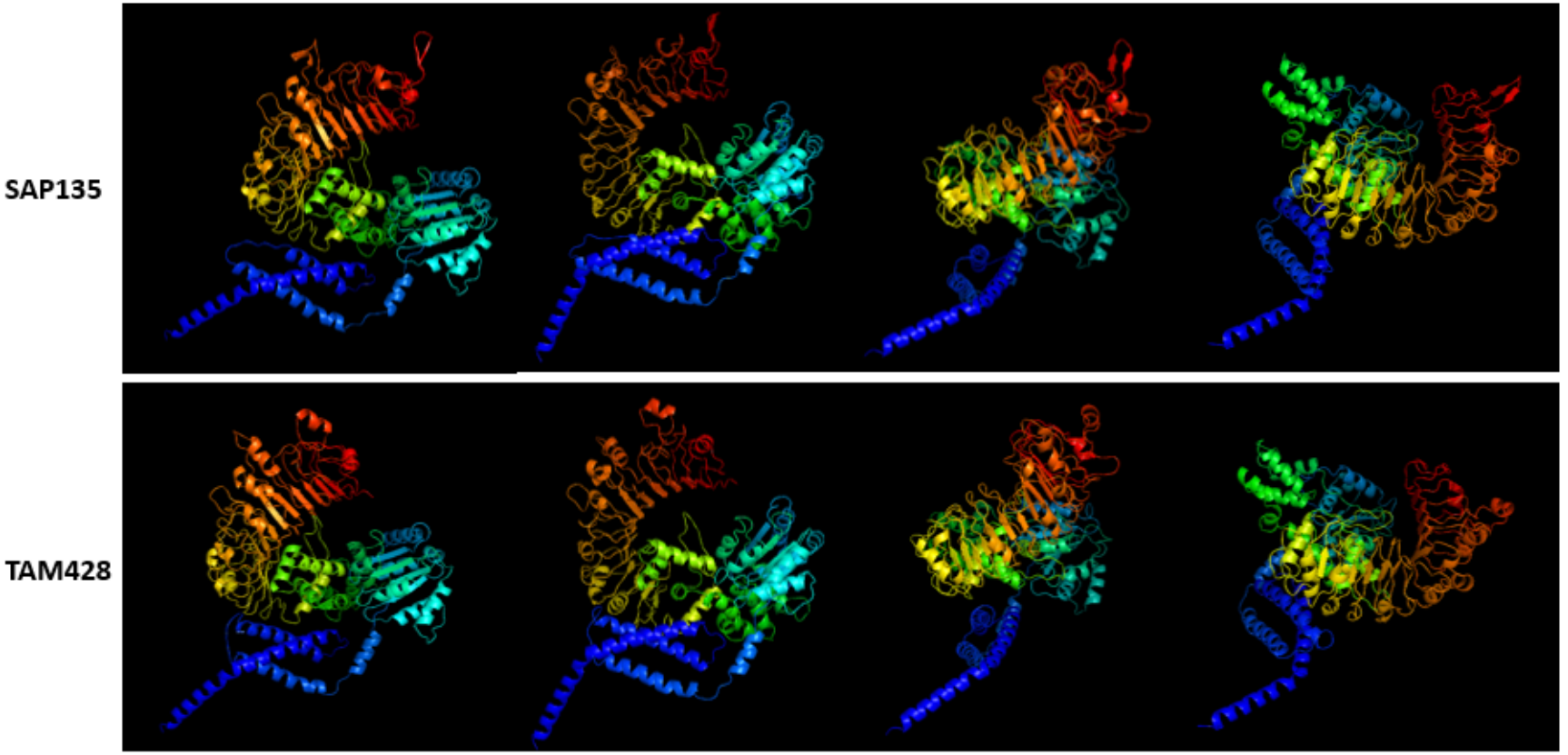


Figure A.1. Multiple view of 3D protein structure of *ARG4* gene from SAP135 (upper panel) and TAM428 (lower panel)

Figure A.2. Multiple sequence alignment of *ARG4* protein from mapping parents and variants

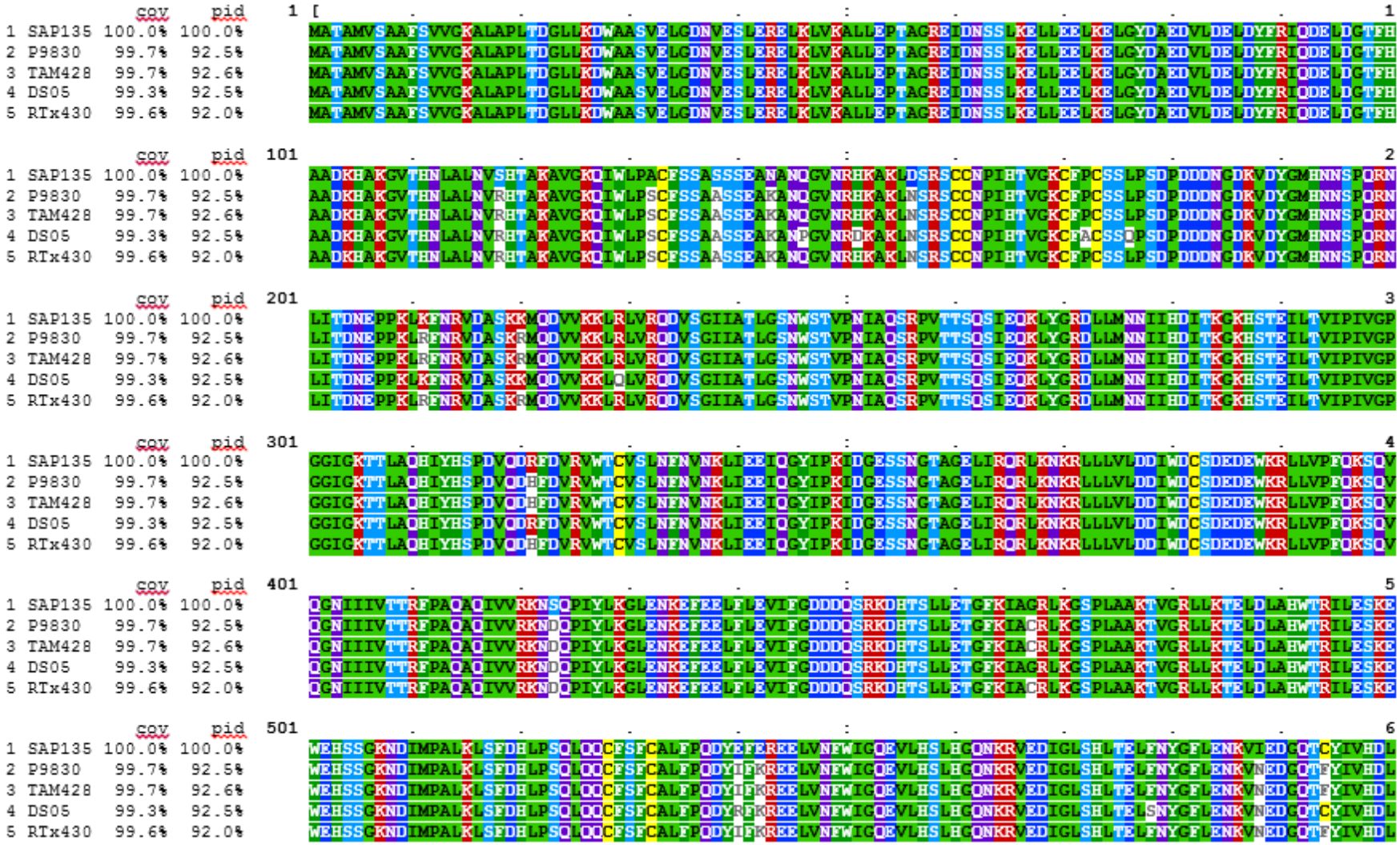


Figure A.2. continued

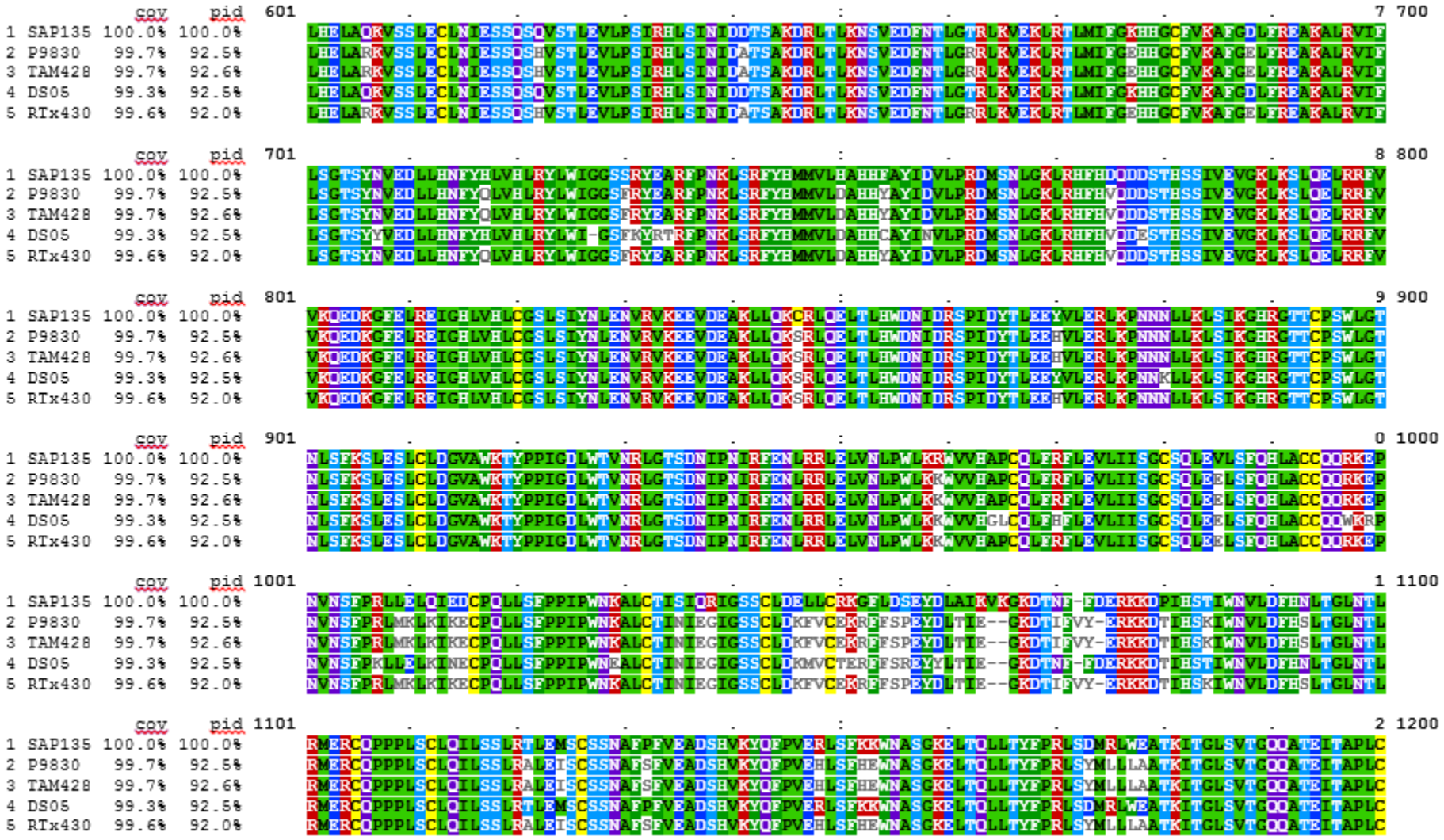
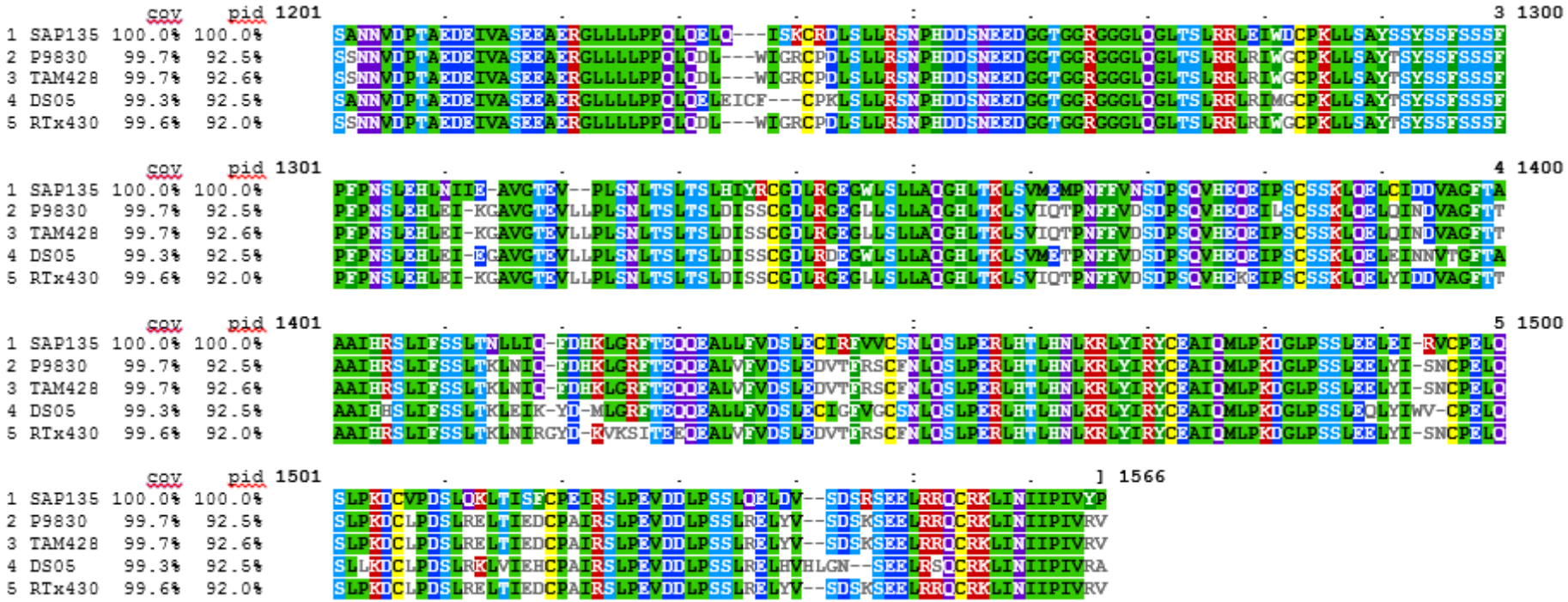


Figure A.2. continued



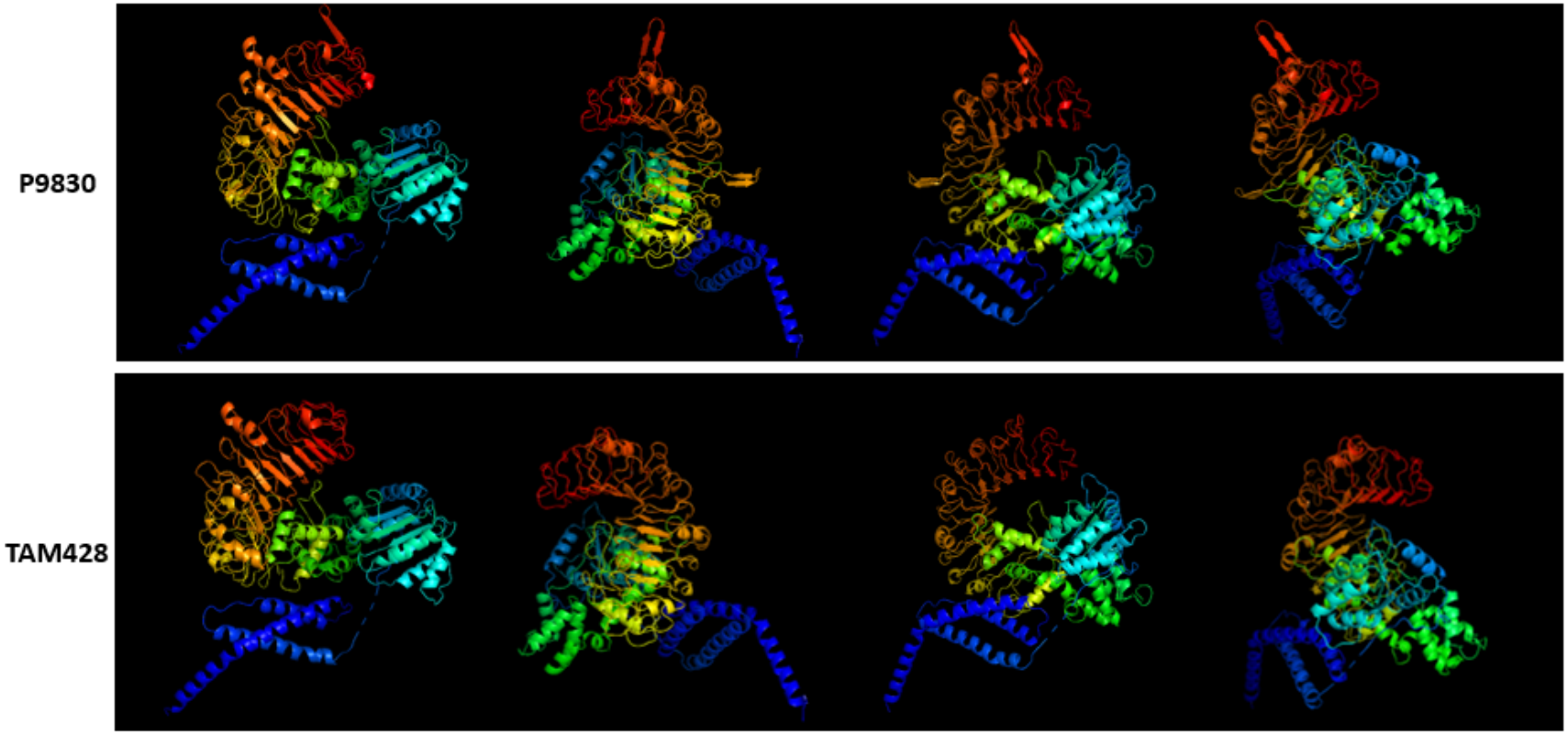


Figure A.3. Multiple view of 3D protein structure of *ARG5* gene from P9830 (upper panel) and TAM428 (lower panel)

Figure A.4. Multiple sequence alignment of *ARG5* protein from mapping parents and variants

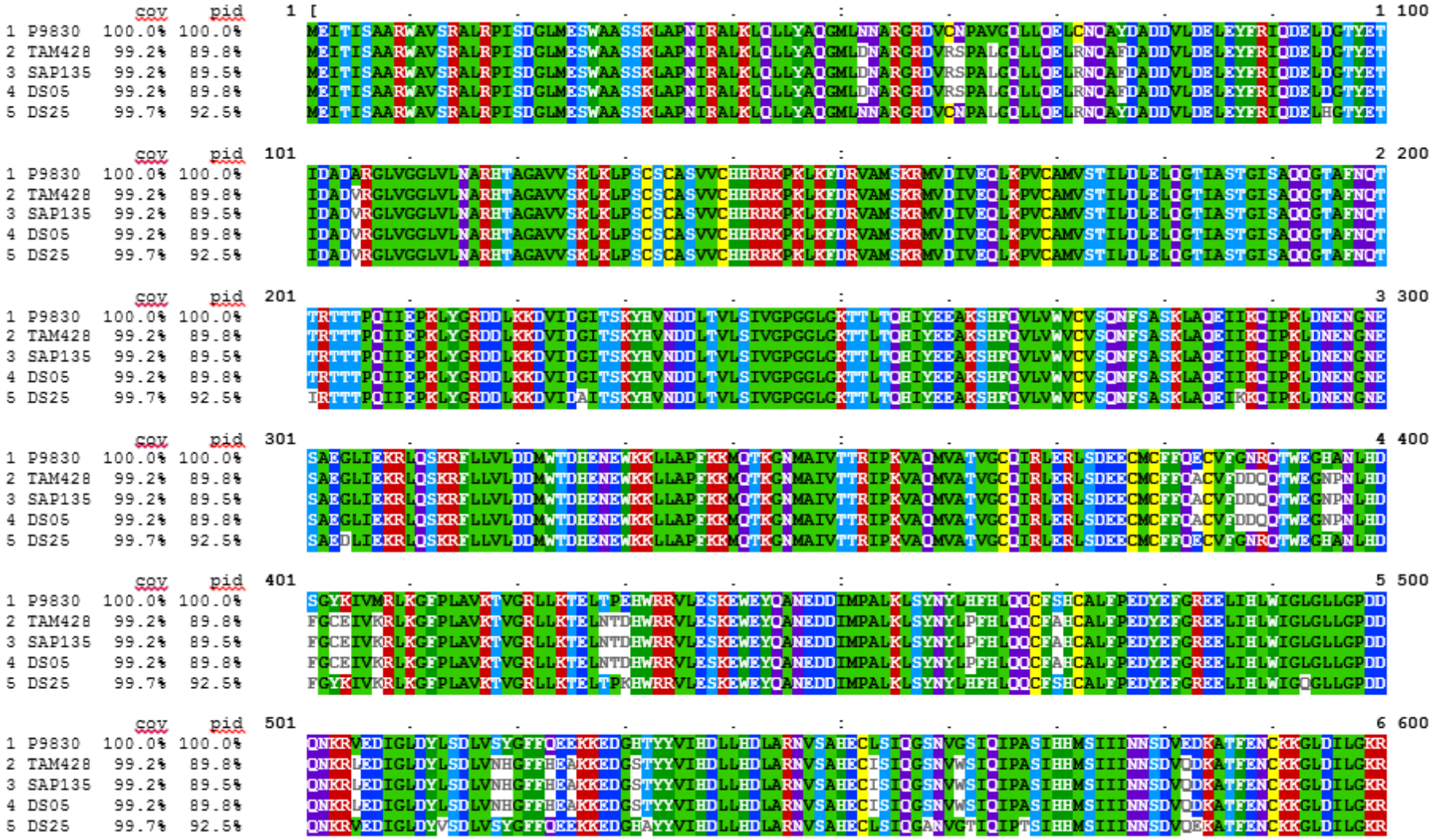


Figure A.4. continued

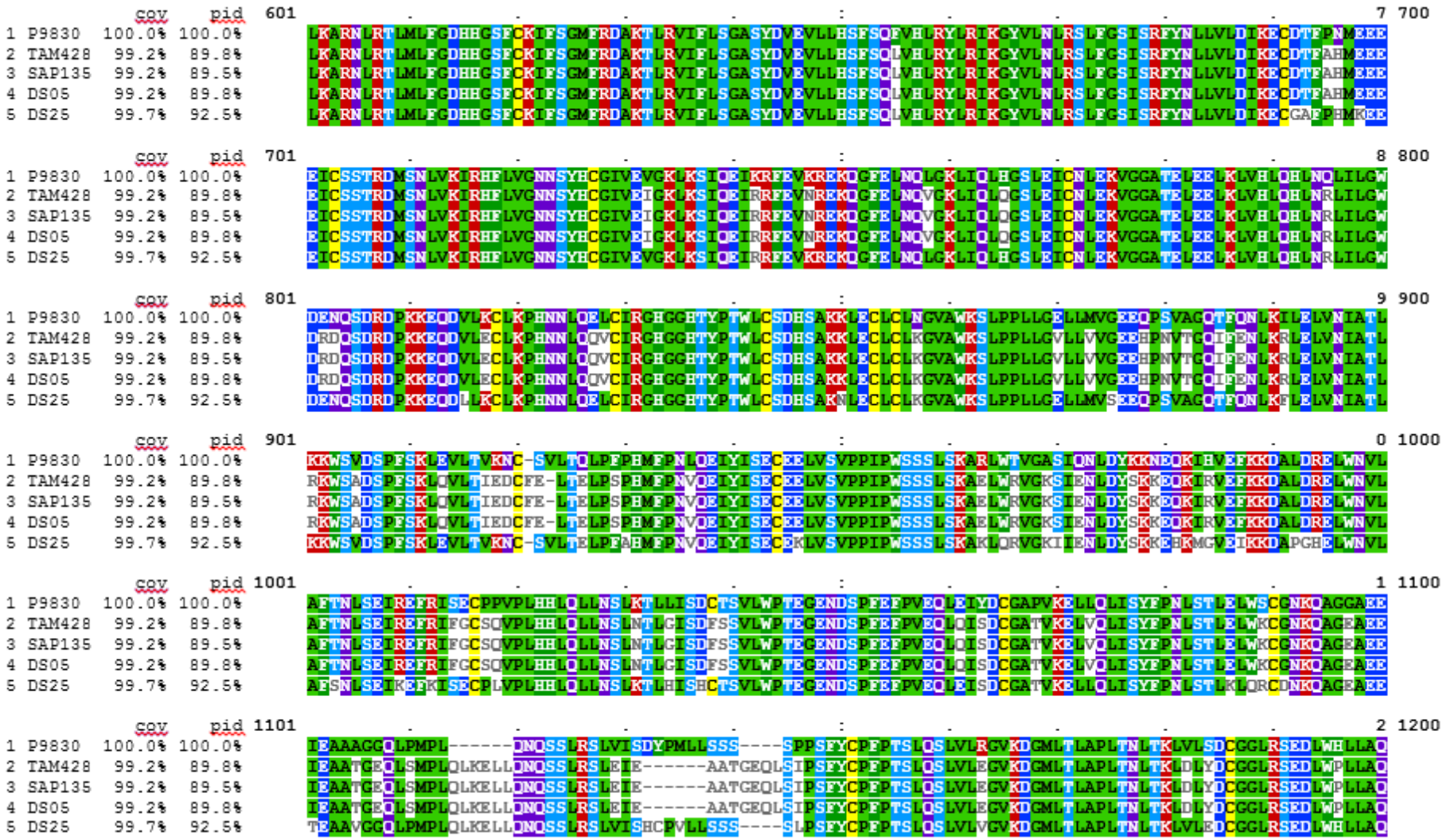
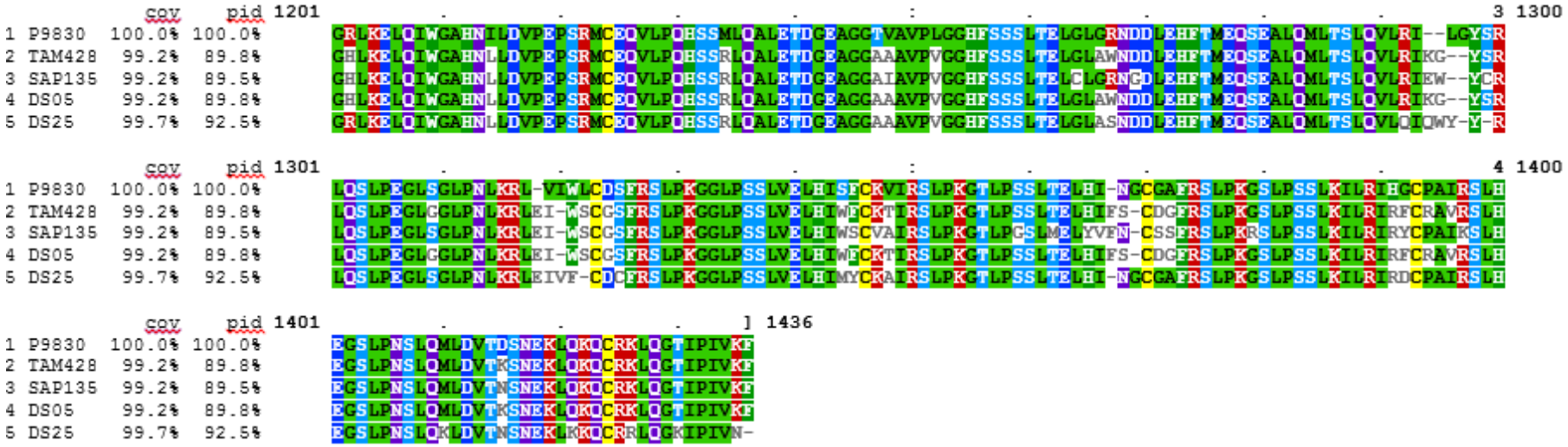


Figure A.4. continued



REFERENCES

1. FAO: **FAOSTAT. Food and Agricultural Organization of the United Nations**. In.; 2020.
2. Awika JM, Rooney LW: **Sorghum phytochemicals and their potential impact on human health**. *Phytochemistry* 2004, **65**(9):1199-1221.
3. Sharma I, Kumari N, Sharma V: **Sorghum Fungal Diseases**. In: *Sustainable Agriculture Reviews.* Edited by Lichtfouse E, Goyal A, vol. 16: Springer, Cham; 2015: 141-172.
4. Tesso T, Perumal R, Little C, Adeyanju A, Radwan G, Prom L, Magill C: **Sorghum pathology and biotechnology - a fungal disease perspective: Part II. Anthracnose, stalk rot, and downy mildew**. *European Journal of Plant Science and Biotechnology* 2011, **6**:31-34.
5. Little C, Perumal R, Tesso T, Prom L, Odvody G, Magill C: **Sorghum pathology and biotechnology - a fungal disease perspective: Part I. Grain mold, head smut, and ergot**. *European Journal of Plant Science and Biotechnology* 2012, **6**:31-44.
6. Tarr SAJ: **Diseases of sorghum, Sudan grass and broom corn**: Commw. Mycol. Inst., Kew, Surrey; 1962.
7. Bandara Y, Tesso TT, Bean SR, Dowell FE, Little CR: **Impacts of Fungal Stalk Rot Pathogens on Physicochemical Properties of Sorghum Grain**. *Plant Dis* 2017, **101**(12):2059-2065.
8. Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, Mengiste T, Ejeta G: **Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance**. *Proc Natl Acad Sci U S A* 2017, **114**(17):4471-4476.
9. Mbuvi DA, Masiga CW, Kuria E, Masanga J, Wamalwa M, Mohamed A, Odeny DA, Hamza N, Timko MP, Runo S: **Novel Sources of Witchweed (Striga) Resistance from Wild Sorghum Accessions**. *Front Plant Sci* 2017, **8**:116.
10. Mohemed N, Charnikhova T, Fradin EF, Rienstra J, Babiker AGT, Bouwmeester HJ: **Genetic variation in Sorghum bicolor strigolactones and their role in resistance against Striga hermonthica**. *J Exp Bot* 2018, **69**(9):2415-2430.

11. Guo C, Cui W, Feng X, Zhao J, Lu G: **Sorghum insect problems and management**. *J Integr Plant Biol* 2011, **53**(3):178-192.
12. Tao YZ, Hardy A, Drenth J, Henzell RG, Franzmann BA, Jordan DR, Butler DG, McIntyre CL: **Identifications of two different mechanisms for sorghum midge resistance through QTL mapping**. *Theor Appl Genet* 2003, **107**(1):116-122.
13. Satish K, Srinivas G, Madhusudhana R, Padmaja PG, Nagaraja Reddy R, Murali Mohan S, Seetharama N: **Identification of quantitative trait loci for resistance to shoot fly in sorghum [Sorghum bicolor (L.) Moench]**. *Theor Appl Genet* 2009, **119**(8):1425-1439.
14. Bowling RD, Brewer MJ, Kerns DL, Gordy J, Seiter N, Elliott NE, Buntin GD, Way MO, Royer TA, Biles S *et al*: **Sugarcane Aphid (Hemiptera: Aphididae): A New Pest on Sorghum in North America**. *J Integr Pest Manag* 2016, **7**(1):12.
15. Tetreault HM, Grover S, Scully ED, Gries T, Palmer NA, Sarath G, Louis J, Sattler SE: **Global Responses of Resistant and Susceptible Sorghum (Sorghum bicolor) to Sugarcane Aphid (Melanaphis sacchari)**. *Front Plant Sci* 2019, **10**:145.
16. Salinas-Hernandez H, Saldamando-Benjumea CI: **Haplotype identification within Spodoptera frugiperda (J.E. Smith) (Lepidoptera: Noctuidae) corn and rice strains from Colombia**. *Neotrop Entomol* 2011, **40**(4):421-430.
17. Tamiru A, Getu E, Jembere B, Bruce T: **Effect of temperature and relative humidity on the development and fecundity of Chilo partellus (Swinhoe) (Lepidoptera: Crambidae)**. *Bull Entomol Res* 2012, **102**(1):9-15.
18. Wu Y, Guo T, Mu Q, et al: **Allelochemicals targeted to balance competing selections in African agroecosystems**. *Nat Plants* 2019, **5**(12):1229-1236.
19. Williams RJ, Rao KN: **A REVIEW OF SORGHUM GRAIN MOLDS**. *Tropical Pest Management* 1981, **27**(2):200-211.
20. Forbes GA, Bandyopadhyay, R., and Garcia, G.: **A review of sorghum grain mold**. In: *Sorghum and millets diseases: a second world review.* Edited by de Milliano WAJ, Frederiksen, R.A., and Bengston, G.D. Patancheru, A.P. 502 324, India: International Crops Research Institute for the Semi-Arid Tropics.(CP 738); 1992: 265-272.
21. Veloso JS, Câmara MPS, Lima WG, Michereff SJ, Doyle VP: **Why species delimitation matters for fungal ecology: Colletotrichum diversity on wild and cultivated cashew in Brazil**. *Fungal Biol* 2018, **122**(7):677-691.

22. Baroncelli R, Sukno SA, Sarrocco S, Cafà G, Le Floch G, Thon MR: **Whole-Genome Sequence of the Orchid Anthracnose Pathogen Colletotrichum orchidophilum**. *Mol Plant Microbe Interact* 2018, **31**(10):979-981.
23. Guarnaccia V, Groenewald JZ, Polizzi G, Crous PW: **High species diversity in Colletotrichum associated with citrus diseases in Europe**. *Persoonia* 2017, **39**:32-50.
24. **Illustrated glossary of plant pathology** [https://www.apsnet.org/edcenter/resources/illglossary/Pages/A-D.aspx]
25. Thakur RP, Reddy BVS, Indira S, Rao VP, Navi SS, Yang XB, Ramesh S: **Sorghum Grain Mold. Information Bulletin No. 72**. Patancheru 502324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics; 2006.
26. Bandyopadhyay R, Butler DR, Chandrasekhar A, Reddy RK, Navi SS: **Biology, epidemiology, and management of sorghum grain mold**. In: *Technical and institutional options for sorghum grain mold management: proceedings of an international consultation, 18-19 May 2000, ICRISAT, Patancheru, India (Chandrashekar, A, Bandyopadhyay, R, and Hall, AJ, eds) Patancheru 502324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics.* 2000: 34-71.
27. Prom LK, Ahn E, Isakeit T, Magill C: **GWAS analysis of sorghum association panel lines identifies SNPs associated with disease response to Texas isolates of Colletotrichum sublineola**. *Theor Appl Genet* 2019, **132**(5):1389-1396.
28. Cuevas HE, Prom LK, Cruet-Burgos CM: **Genome-Wide Association Mapping of Anthracnose (Colletotrichum sublineolum) Resistance in NPGS Ethiopian Sorghum Germplasm**. *G3 (Bethesda)* 2019, **9**(9):2879-2885.
29. Pena GA, Cavaglieri LR, Chulze SN: **Fusarium species and moniliformin occurrence in sorghum grains used as ingredient for animal feed in Argentina**. *J Sci Food Agric* 2019, **99**(1):47-54.
30. Desjardins AE, Plattner RD: **Fumonisin B(1)-nonproducing strains of Fusarium verticillioides cause maize (Zea mays) ear infection and ear rot**. *J Agric Food Chem* 2000, **48**(11):5773-5780.
31. Munkvold GP: **Epidemiology of Fusarium diseases and their mycotoxins in maize ears**. *European Journal of Plant Pathology* 2003, **109**:705-713.

32. Borah SN, Goswami D, Sarma HK, Cameotra SS, Deka S: **Rhamnolipid Biosurfactant against Fusarium verticillioides to Control Stalk and Ear Rot Disease of Maize**. *Front Microbiol* 2016, **7**:1505.
33. Gai X, Dong H, Wang S, Liu B, Zhang Z, Li X, Gao Z: **Infection cycle of maize stalk rot and ear rot caused by Fusarium verticillioides**. *PLoS One* 2018, **13**(7):e0201588.
34. Jansen C, von Wettstein D, Schäfer W, Kogel KH, Felk A, Maier FJ: **Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted Fusarium graminearum**. *Proc Natl Acad Sci U S A* 2005, **102**(46):16892-16897.
35. Holliday P: **Fungus diseases of tropical crops**. Cambridge UK: Cambridge University Press; 1980.
36. Sutton BC: **The appressoria of C*olletotrichum graminicola a*nd C*. falcatum***. *Canadian Journal of Botany* 1968, **46**:873-876.
37. Vaillancourt LJ, Hanau RM: **Genetic and morphological comparisons of G*lomerella (*C*olletotrichum)* isolates from maize and sorghum**. *Experimental Mycology* 1992, **16**:219-229.
38. Sherriff C, Whelan MJ, Arnold GM, Bailey JA: **rDNA sequence analysis confirms the distinction between C*olletotrichum graminicola* and C. *sublineolum***. *Mycological Research* 1995, **99**:475-478.
39. Ahn E, Hu Z, Perumal R, Prom LK, Odvody G, Upadhyaya HD, Magill C: **Genome wide association analysis of sorghum mini core lines regarding anthracnose, downy mildew, and head smut**. *PLoS One* 2019, **14**(5):e0216671.
40. Schnippenkoetter W, Lo C, Liu G, Dibley K, Chan WL, White J, Milne R, Zwart A, Kwong E, Keller B *et al*: **The wheat Lr34 multipathogen resistance gene confers resistance to anthracnose and rust in sorghum**. *Plant Biotechnol J* 2017, **15**(11):1387-1396.
41. Wang L, Chen M, Zhu F, Fan T, Zhang J, Lo C: **Alternative splicing is a Sorghum bicolor defense response to fungal infection**. *Planta* 2019, **251**(1):14.
42. Acharya B, O'Quinn TN, Everman W, Mehl HL: **Effectiveness of Fungicides and Their Application Timing for the Management of Sorghum Foliar Anthracnose in the Mid-Atlantic United States**. *Plant Dis* 2019, **103**(11):2804-2811.

43. T JF, L MM, Saballos A, Vermerris W: **Using Genotyping by Sequencing to Map Two Novel Anthracnose Resistance Loci in Sorghum bicolor**. *G3 (Bethesda)* 2016, **6**(7):1935-1946.

44. Tugizimana F, Djami-Tchatchou AT, Fahrmann JF, Steenkamp PA, Piater LA, Dubery IA: **Time-resolved decoding of metabolic signatures of in vitro growth of the hemibiotrophic pathogen Colletotrichum sublineolum**. *Scientific Reports* 2019, **9**(1):3290.

45. Torres MF, Ghaffari N, Buiate EA, Moore N, Schwartz S, Johnson CD, Vaillancourt LJ: **A Colletotrichum graminicola mutant deficient in the establishment of biotrophy reveals early transcriptional events in the maize anthracnose disease interaction**. *BMC Genomics* 2016, **17**:202.

46. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N *et al*: **Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses**. *Nat Genet* 2012, **44**(9):1060-1065.

47. Ibrahim OE, Nyquist WE, Axtell JD: **QUANTITATIVE INHERITANCE AND CORRELATIONS OF AGRONOMIC AND GRAIN QUALITY TRAITS OF SORGHUM**. *Crop Science* 1985, **25**(4):649-654.

48. Little CR, Magill CW: **The Grain Mold Pathogen, Fusarium thapsinum, Reduces Caryopsis Formation in Sorghum bicolor**. *Journal of Phytopathology* 2009, **157**(7-8):518-519.

49. Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL: **Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum**. *Theoretical and Applied Genetics* 2001, **102**(2-3):307-319.

50. Chala A, Taye W, Ayalew A, Krska R, Sulyok M, Logrieco A: **Multimycotoxin analysis of sorghum (Sorghum bicolor L. Moench) and finger millet (Eleusine coracana L. Garten) from Ethiopia**. *Food Control* 2014, **45**:29-35.

51. Ssepuuya G, Van Poucke C, Ediage EN, Mulholland C, Tritscher A, Verger P, Kenny M, Bessy C, De Saeger S: **Mycotoxin contamination of sorghum and its contribution to human dietary exposure in four sub-Saharan countries**. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 2018, **35**(7):1384-1393.
52. Thomas MD, Sissoko I, Sacko M: **Development of leaf anthracnose and its effect on yield and grain weight of sorghum in West Africa**. *Plant Disease* 1996, **80**:151-153.
53. Bigeard J, Colcombet J, Hirt H: **Signaling mechanisms in pattern-triggered immunity (PTI)**. *Mol Plant* 2015, **8**(4):521-539.
54. Heath MC: **Nonhost resistance and nonspecific plant defenses**. *Curr Opin Plant Biol* 2000, **3**(4):315-319.
55. Lee HA, Lee HY, Seo E, Lee J, Kim SB, Oh S, Choi E, Lee SE, Choi D: **Current Understandings of Plant Nonhost Resistance**. *Mol Plant Microbe Interact* 2017, **30**(1):5-15.
56. Gilbert B, Bettgenhaeuser J, Upadhyaya N, Soliveres M, Singh D, Park RF, Moscou MJ, Ayliffe M: **Components of Brachypodium distachyon resistance to nonadapted wheat stripe rust pathogens are simply inherited**. *PLoS Genet* 2018, **14**(9):e1007636.
57. Tufan HA, McGrann GR, Magusin A, Morel JB, Miché L, Boyd LA: **Wheat blast: histopathology and transcriptome reprogramming in response to adapted and nonadapted Magnaporthe isolates**. *New Phytol* 2009, **184**(2):473-484.
58. Aghnoum R, Niks RE: **Specificity and levels of nonhost resistance to nonadapted Blumeria graminis forms in barley**. *New Phytol* 2010, **185**(1):275-284.
59. Thordal-Christensen H: **Fresh insights into processes of nonhost resistance**. *Curr Opin Plant Biol* 2003, **6**(4):351-357.
60. Jones JD, Dangl JL: **The plant immune system**. *Nature* 2006, **444**(7117):323-329.
61. Luo Y, Bai R, Li J, Yang W, Li R, Wang Q, Zhao G, Duan D: **The transcription factor MYB15 is essential for basal immunity (PTI) in Chinese wild grape**. *Planta* 2019, **249**(6):1889-1902.
62. Hetmann A, Kowalczyk S: **[Suppression of PAMP-triggered immunity (PTI) by effector proteins synthesized by phytopathogens and delivered into cells of infected plant]**. *Postepy Biochem* 2019, **65**(1):58-71.

63. Szatmári Á, Zvara Á, Móricz Á M, Besenyei E, Szabó E, Ott PG, Puskás LG, Bozsó Z: **Pattern triggered immunity (PTI) in tobacco: isolation of activated genes suggests role of the phenylpropanoid pathway in inhibition of bacterial pathogens**. *PLoS One* 2014, **9**(8):e102869.
64. Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE: **Compromised disease resistance in saponin-deficient plants**. *Proc Natl Acad Sci U S A* 1999, **96**(22):12923-12928.
65. Freialdenhoven A, Peterhansel C, Kurth J, Kreuzaler F, Schulze-Lefert P: **Identification of Genes Required for the Function of Non-Race-Specific mlo Resistance to Powdery Mildew in Barley**. *Plant Cell* 1996, **8**(1):5-14.
66. Miyamoto K, Matsumoto T, Okada A, Komiyama K, Chujo T, Yoshikawa H, Nojiri H, Yamane H, Okada K: **Identification of target genes of the bZIP transcription factor OsTGAP1, whose overexpression causes elicitor-induced hyperaccumulation of diterpenoid phytoalexins in rice cells**. *PLoS One* 2014, **9**(8):e105823.
67. Deice Raasch-Fernandes L, Bonaldo SM, de Jesus Rodrigues D, Magela Vieira-Junior G, Regina Freitas Schwan-Estrada K, Rocco da Silva C, Gabriela Araújo Verçosa A, Lopes de Oliveira D, Wender Debiasi B: **Induction of phytoalexins and proteins related to pathogenesis in plants treated with extracts of cutaneous secretions of southern Amazonian Bufonidae amphibians**. *PLoS One* 2019, **14**(1):e0211020.
68. Ube N, Yabuta Y, Tohnooka T, Ueno K, Taketa S, Ishihara A: **Biosynthesis of Phenylamide Phytoalexins in Pathogen-Infected Barley**. *Int J Mol Sci* 2019, **20**(22).
69. Esele JP, Frederiksen RA, Miller FR: **THE ASSOCIATION OF GENES-CONTROLLING CARYOPSIS TRAITS WITH GRAIN MOLD RESISTANCE IN SORGHUM**. *Phytopathology* 1993, **83**(5):490-495.
70. Menkir A, Ejeta G, Butler L, Melakeberhan A: **Physical and chemical kernel properties associated with resistance to grain mold in sorghum**. *Cereal Chemistry* 1996, **73**(5):613-617.
71. MelakeBerhan A, Butler LG, Ejeta G, Menkir A: **Grain mold resistance and polyphenol accumulation in sorghum**. *Journal of Agricultural and Food Chemistry* 1996, **44**(8):2428-2434.

72. Audilakshmi S, Stenhouse JW, Reddy TP, Prasad MVR: **Grain mould resistance and associated characters of sorghum genotypes**. *Euphytica* 1999, **107**(2):91-103.
73. Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P: **Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor L. Moench)**. *Theoretical and Applied Genetics* 1998, **97**(4):605-616.
74. Rodriguez-Herrera R, Rooney WL, Rosenow DT, Frederiksen RA: **Inheritance of Grain Mold Resistance in Grain Sorghum without a Pigmented Testa**. *Crop Science* 2000, **40**(6):1573-1578.
75. Upadhyaya HD, Wang YH, Sharma R, Sharma S: **SNP markers linked to leaf rust and grain mold resistance in sorghum**. *Molecular Breeding* 2013, **32**(2):451-462.
76. Cuevas HE, Fermin-Perez RA, Prom LK, Cooper EA, Bean S, Rooney WL: **Genome-Wide Association Mapping of Grain Mold Resistance in the US Sorghum Association Panel**. *Plant Genome* 2019, **12**(2).
77. Prom LKC, H. E. Ahn, E. Isakeit, T. Rooney, W. L. Magill, C.: **Genome-wide association study of grain mold resistance in sorghum association panel as affected by inoculation with *Alternaria alternata* alone and Al*ternaria alternata*, Fusa*rium thapsinum, a*nd Curvularia lunata combined**. *European Journal of Plant Pathology* 2020, **157**:783-798.
78. Stephens JC, Miller FR, Rosenow DT: **Conversion of Alien Sorghums to Early Combine Genotypes1**. *Crop Science* 1967, **7**(4):cropsci1967.0011183X000700040036x.
79. Rosenow DT, Dahlberg JA, Stephens JC, Miller FR, Barnes DK, Peterson GC, Johnson JW, Schertz KF: **Registration of 63 Converted Sorghum Germplasm Lines from the Sorghum Conversion Program**. *Crop Science* 1997, **37**(4):cropsci1997.0011183X003700040090x.
80. Mehta PJ, Wiltse CC, Rooney WL, Collins SD, Frederiksen RA, Hess DE, Chisi M, TeBeest DO: **Classification and inheritance of genetic resistance to anthracnose in sorghum**. *Field Crops Research* 2005, **93**(1):1-9.

81. Ramasamy P, Menz MA, Mehta PJ, Katilé S, Gutierrez-Rojas LA, Klein RR, Klein PE, Prom LK, Schlueter JA, Rooney WL *et al*: **Molecular mapping of Cg1, a gene for resistance to anthracnose (Colletotrichum sublineolum) in sorghum**. *Euphytica* 2008, **165**(3):597.
82. Burrell AM, Sharma A, Patil NY, Collins SD, Anderson WF, Rooney WL, Klein PE: **Sequencing of an Anthracnose-Resistant Sorghum Genotype and Mapping of a Major QTL Reveal Strong Candidate Genes for Anthracnose Resistance**. *Crop Science* 2015, **55**(2):790-799.
83. Prom LK, Perumal R, Erattaimuthu SR, Little CR, No EG, Erpelding JE, Rooney WL, Odvody GN, Magill CW: **Genetic diversity and pathotype determination of Colletotrichum sublineolum isolates causing anthracnose in sorghum**. *European Journal of Plant Pathology* 2012, **133**(3):671-685.
84. Cuevas HE, Prom LK, Erpelding JE: **Inheritance and molecular mapping of anthracnose resistance genes present in sorghum line SC112-14**. *Molecular Breeding* 2014, **34**(4):1943-1953.
85. Patil NY, Klein RR, Williams CL, Collins SD, Knoll JE, Burrell AM, Anderson WF, Rooney WL, Klein PE: **Quantitative Trait Loci Associated with Anthracnose Resistance in Sorghum**. *Crop Science* 2017, **57**(2):877-890.
86. Cuevas HE, Prom LK: **Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan Sorghum Core collection**. *BMC Genomics* 2020, **21**(1):88.
87. Cuevas HE, Prom LK, Cooper EA, Knoll JE, Ni X: **Genome-Wide Association Mapping of Anthracnose (Colletotrichum sublineolum) Resistance in the U.S. Sorghum Association Panel**. *Plant Genome* 2018, **11**(2).
88. Ibraheem F, Gaffoor I, Chopra S: **Flavonoid Phytoalexin-Dependent Resistance to Anthracnose Leaf Blight Requires a Functional yellow seed1 in Sorghum bicolor**. *Genetics* 2010, **184**(4):915-926.
89. Ibraheem F, Gaffoor I, Tan QX, Shyu CR, Chopra S: **A Sorghum MYB Transcription Factor Induces 3-Deoxyanthocyanidins and Enhances Resistance against Leaf Blights in Maize**. *Molecules* 2015, **20**(2):2388-2404.

90. Snyder BA, Nicholson RL: **Synthesis of phytoalexins in sorghum as a site-specific response to fungal ingress**. *Science* 1990, **248**(4963):1637-1639.
91. Nicholson RL, Kollipara SS, Vincent JR, Lyons PC, Cadenagomez G: **PHYTOALEXIN SYNTHESIS BY THE SORGHUM MESOCOTYL IN RESPONSE TO INFECTION BY PATHOGENIC AND NONPATHOGENIC FUNGI**. *Proceedings of the National Academy of Sciences of the United States of America* 1987, **84**(16):5520-5524.
92. Morris GP, Rhodes DH, Brenton Z, Ramu P, Thayil VM, Deshpande S, Hash CT, Acharya C, Mitchell SE, Buckler ES *et al*: **Dissecting Genome-Wide Association Signals for Loss-of-Function Phenotypes in Sorghum Flavonoid Pigmentation Traits**. *G3-Genes Genomes Genetics* 2013, **3**(11):2085-2094.
93. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S *et al*: **QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations**. *Plant J* 2013, **74**(1):174-183.
94. Hisano H, Sakamoto K, Takagi H, Terauchi R, Sato K: **Exome QTL-seq maps monogenic locus and QTLs in barley**. *BMC Genomics* 2017, **18**(1):125.
95. Singh VK, Khan AW, Jaganathan D, Thudi M, Roorkiwal M, Takagi H, Garg V, Kumar V, Chitikineni A, Gaur PM *et al*: **QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea**. *Plant Biotechnol J* 2016, **14**(11):2110-2119.
96. Yoshitsu Y, Takakusagi M, Abe A, Takagi H, Uemura A, Yaegashi H, Terauchi R, Takahata Y, Hatakeyama K, Yokoi S: **QTL-seq analysis identifies two genomic regions determining the heading date of foxtail millet, Setaria italica (L.) P.Beauv**. *Breed Sci* 2017, **67**(5):518-527.
97. Kodama A, Narita R, Yamaguchi M, Hisano H, Adachi S, Takagi H, Ookawa T, Sato K, Hirasawa T: **QTLs maintaining grain fertility under salt stress detected by exome QTL-seq and interval mapping in barley**. *Breed Sci* 2018, **68**(5):561-570.
98. Kurlovs AH, Snoeck S, Kosterlitz O, Van Leeuwen T, Clark RM: **Trait mapping in diverse arthropods by bulked segregant analysis**. *Curr Opin Insect Sci* 2019, **36**:57-65.

99. Li R, Jiang H, Zhang Z, Zhao Y, Xie J, Wang Q, Zheng H, Hou L, Xiong X, Xin D *et al*: **Combined Linkage Mapping and BSA to Identify QTL and Candidate Genes for Plant Height and the Number of Nodes on the Main Stem in Soybean**. *Int J Mol Sci* 2019, **21**(1).
100. Pujol M, Alexiou KG, Fontaine AS, Mayor P, Miras M, Jahrmann T, Garcia-Mas J, Aranda MA: **Mapping Cucumber Vein Yellowing Virus Resistance in Cucumber (Cucumis sativus L.) by Using BSA-seq Analysis**. *Front Plant Sci* 2019, **10**:1583.
101. Aguado E, García A, Iglesias-Moya J, Romero J, Wehner TC, Gómez-Guillamón ML, Picó B, Garcés-Claver A, Martínez C, Jamilena M: **Mapping a Partial Andromonoecy Locus in Citrullus lanatus Using BSA-Seq and GWAS Approaches**. *Front Plant Sci* 2020, **11**:1243.
102. Lee SB, Kim JE, Kim HT, Lee GM, Kim BS, Lee JM: **Genetic mapping of the c1 locus by GBS-based BSA-seq revealed Pseudo-Response Regulator 2 as a candidate gene controlling pepper fruit color**. *Theor Appl Genet* 2020, **133**(6):1897-1910.
103. Liang T, Chi W, Huang L, Qu M, Zhang S, Chen ZQ, Chen ZJ, Tian D, Gui Y, Chen X *et al*: **Bulked Segregant Analysis Coupled with Whole-Genome Sequencing (BSA-Seq) Mapping Identifies a Novel pi21 Haplotype Conferring Basal Resistance to Rice Blast Disease**. *Int J Mol Sci* 2020, **21**(6).
104. Hoffman GE: **Correcting for population structure and kinship using the linear mixed model: theory and extensions**. *PLoS One* 2013, **8**(10):e75707.
105. Astle W, Balding DJ: **Population Structure and Cryptic Relatedness in Genetic Association Studies**. *Statist Sci* 2009, **24**(4):451-471.
106. Platt A, Vilhjálmsson BJ, Nordborg M: **Conditions under which genome-wide association studies will be positively misleading**. *Genetics* 2010, **186**(3):1045-1052.
107. Kaler AS, Purcell LC: **Estimation of a significance threshold for genome-wide association studies**. *BMC Genomics* 2019, **20**(1):618.
108. Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nature Genetics* 2006, **38**(2):203-208.

109. Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu JM, Arnett DK, Ordovas JM *et al*: **Mixed linear model approach adapted for genome-wide association studies**. *Nat Genet* 2010, **42**(4):355-U118.

110. Liu X, Huang M, Fan B, Buckler ES, Zhang Z: **Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies**. *PLoS Genet* 2016, **12**(2).

111. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z: **BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions**. *Gigascience* 2019, **8**(2).

112. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nature Genetics* 2006, **38**(8):904-909.

113. Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, Gilding E, Cruickshank A, Prentis P, Wang J *et al*: **The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes**. *BMC Plant Biology* 2014, **14**(1):253.

114. Nida H, Girma G, Mekonen M, Lee S, Seyoum A, Dessalegn K, Tadesse T, Ayana G, Senbetay T, Tess T, Ejeta, G. *et al*: **Identification of sorghum grain mold resistance loci through genome wide association mapping**. *Journal of Cereal Science* 2019, **85**:295-304.

115. Prom LK, Waniska RD, Kollo AI, Rooney WL, Bejosano FP: **Role of chitinase and sormatin accumulation in the resistance of sorghum cultivars to grain mold**. *Journal of Agricultural and Food Chemistry* 2005, **53**(14):5565-5570.

116. Menkir A, Ejeta G, Butler LG, Melakeberhan A, Warren HL: **Fungal invasion of kernels and grain mold damage assessment in diverse sorghum germ plasm**. *Plant Disease* 1996, **80**(12):1399-1402.

117. Gilbert J, Fernando WGD: **Epidemiology and biological control of Gibberella zeae Fusarium graminearum**. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie* 2004, **26**(4):464-472.

118. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species**. *Plos One* 2011, **6**(5).

119. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES: **TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline**. *Plos One* 2014, **9**(2).

120. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu SQ, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B *et al*: **The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization**. *Plant Journal* 2018, **93**(2):338-354.

121. Alexander DH, Lange K: **Enhancements to the ADMIXTURE algorithm for individual ancestry estimation**. *Bmc Bioinformatics* 2011, **12**.

122. Lipka AE, Tian F, Wang QS, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang ZW: **GAPIT: genome association and prediction integrated tool**. *Bioinformatics* 2012, **28**(18):2397-2399.

123. **R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.** 2019.

124. Turner SD: **qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots**. *bioRxiv* 2014:005165.

125. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. *Nature Biotechnology* 2011, **29**(1):24-26.

126. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. *Briefings in Bioinformatics* 2013, **14**(2):178-192.

127. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences**. *Journal of Computational Biology* 2000, **7**(1-2):203-214.

128. Wang GF, Wang G, Zhang XW, Wang F, Song RT: **Isolation of High Quality RNA from Cereal Seeds Containing High Levels of Starch**. *Phytochemical Analysis* 2012, **23**(2):159-163.

129. Liu HJ, Du YG, Chu H, Shih CH, Wong YW, Wang MF, Chu IK, Tao YZ, Lo C: **Molecular Dissection of the Pathogen-Inducible 3-Deoxyanthocyanidin Biosynthesis Pathway in Sorghum**. *Plant Cell Physiol* 2010, **51**(7):1173-1185.

130. Fox G, Nugusu Y, Nida H, Tedessa T, McLean G, Jordan D: **Evaluation of variation in Ethiopian sorghum injera quality with new imaging techniques**. *Cereal Chemistry* 2020, **97**(2):362-372.
131. Boddu J, Svabek C, Ibraheem F, Jones AD, Chopra S: **Characterization of a deletion allele of a sorghum Myb gene, yellow seed1 showing loss of 3-deoxyflavonoids**. *Plant Science* 2005, **169**(3):542-552.
132. Brenton ZW, Cooper EA, Myers MT, Boyles RE, Shakoor N, Zielinski KJ, Rauh BL, Bridges WC, Morris GP, Kresovich S: **A Genomic Resource for the Development, Improvement, and Exploitation of Sorghum for Bioenergy**. *Genetics* 2016, **204**(1):21-+.
133. Boddu J, Jiang CH, Sangar V, Olson T, Peterson T, Chopra S: **Comparative structural and functional characterization of sorghum and maize duplications containing orthologous myb transcription regulators of 3-deoxyflavonoid biosynthesis**. *Plant Molecular Biology* 2006, **60**(2):185-199.
134. Lo SCC, De Verdier K, Nicholson RL: **Accumulation of 3-deoxyanthocyanidin phytoalexins and resistance to Colletotrichum sublineolum in sorghum**. *Physiological and Molecular Plant Pathology* 1999, **55**(5):263-273.
135. Du YG, Chu H, Wang MF, Chu IK, Lo C: **Identification of flavone phytoalexins and a pathogen-inducible flavone synthase II gene (SbFNSII) in sorghum**. *Journal of Experimental Botany* 2010, **61**(4):983-994.
136. Xu WJ, Dubos C, Lepiniec L: **Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes**. *Trends in Plant Science* 2015, **20**(3):176-185.
137. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L: **TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana**. *Plant Journal* 2004, **39**(3):366-380.
138. Mizuno H, Yazawa T, Kasuga S, Sawada Y, Ogata J, Ando T, Kanamori H, Yonemaru J, Wu J, Hirai MY *et al*: **Expression level of a flavonoid 3′-hydroxylase gene determines pathogen-induced color variation in sorghum**. *BMC Res Notes* 2014, **7**.

139. Schroeder HW, Christensen JJ: **FACTORS AFFECTING RESISTANCE OF WHEAT TO SCAB CAUSED BY GIBBERELLA ZEAE**. *Phytopathology* 1963, **53**(7):831-&.

140. Mesterhazy A: **Types and components of resistance to Fusarium head blight of wheat**. *Plant Breeding* 1995, **114**(5):377-386.

141. Skinnes H, Semagn K, Tarkegne Y, Maroy AG, Bjornstad A: **The inheritance of anther extrusion in hexaploid wheat and its relationship to Fusarium head blight resistance and deoxynivalenol content**. *Plant Breeding* 2010, **129**(2):149-155.

142. Lu QX, Lillemo M, Skinnes H, He XY, Shi JR, Ji F, Dong YH, Bjornstad A: **Anther extrusion and plant height are associated with Type I resistance to Fusarium head blight in bread wheat line 'Shanghai-3/Catbird'**. *Theoretical and Applied Genetics* 2013, **126**(2):317-334.

143. Liu SY, Hall MD, Griffey CA, McKendry AL: **Meta-Analysis of QTL Associated with Fusarium Head Blight Resistance in Wheat**. *Crop Science* 2009, **49**(6):1955-1968.

144. Srinivasachary, Gosman N, Steed A, Hollins TW, Bayles R, Jennings P, Nicholson P: **Semi-dwarfing Rht-B1 and Rht-D1 loci of wheat differ significantly in their influence on resistance to Fusarium head blight**. *Theoretical and Applied Genetics* 2009, **118**(4):695-702.

145. Jiang Y, Zhao Y, Rodemann B, Plieske J, Kollers S, Korzun V, Ebmeyer E, Argillier O, Hinze M, Ling J *et al*: **Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (Triticum aestivum L.)**. *Heredity* 2015, **114**(3):318-326.

146. Strange RN, Majer JR, Smith H: **ISOLATION AND IDENTIFICATION OF CHOLINE AND BETAINE AS 2 MAJOR COMPONENTS IN ANTHERS AND WHEAT-GERM THAT STIMULATE FUSARIUM-GRAMINEARUM INVITRO**. *Physiological Plant Pathology* 1974, **4**(2):277-290.

147. Kage U, Karre S, Kushalappa AC, McCartney C: **Identification and characterization of a fusarium head blight resistance gene TaACT in wheat QTL-2DL**. *Plant Biotechnology Journal* 2017, **15**(4):447-457.

148. Keller H, Hohlfeld H, Wray V, Hahlbrock K, Scheel D, Strack D: **Changes in the accumulation of soluble and cell wall-bound phenolics in elicitor-treated cell suspension cultures and fungus-infected leaves of Solanum tuberosum**. *Phytochemistry* 1996, **42**(2):389-396.
149. Schmidt A, Scheel D, Strack D: **Elicitor-stimulated biosynthesis of hydroxycinnamoyltyramines in cell suspension cultures of Solanum tuberosum**. *Planta* 1998, **205**(1):51-55.
150. Ishihara A, Hashimoto Y, Tanaka C, Dubouzet JG, Nakao T, Matsuda F, Nishioka T, Miyagawa H, Wakasa K: **The tryptophan pathway is involved in the defense responses of rice against pathogenic infection via serotonin production**. *Plant Journal* 2008, **54**(3):481-495.
151. Nida H, Girma G, Mekonen M, Tirfessa A, Seyoum A, Bejiga T, Birhanu C, Dessalegn K, Senbetay T, Ayana G *et al*: **Genome-wide association analysis reveals seed protein loci as determinants of variations in grain mold resistance in sorghum**. *Theoretical and Applied Genetics* 2021.
152. da Silva JB, Pozzi CR, Mallozzi MAB, Ortega EM, Corrêa B: **Mycoflora and Occurrence of Aflatoxin B1 and Fumonisin B1 during Storage of Brazilian Sorghum**. *Journal of Agricultural and Food Chemistry* 2000, **48**(9):4352-4356.
153. Leslie JF: **Mycotoxins in the Sorghum Grain Chain**. In: *Mycotoxin reduction in grain chains.* Edited by Leslie JF, Logerico AF: John Wiley & Sons, Inc.; 2014: 282-296.
154. Esele JP, Frederiksen RA, Miller FR: **The association of genes controlling caryopsis traits with grain mold resistance in sorghum**. *Phytopathology* 1993, **83**(5):490-495.
155. Girma G, Nida H, Seyoum A, Mekonen M, Nega A, Lule D, Dessalegn K, Bekele A, Gebreyohannes A, Adeyanju A *et al*: **A Large-Scale Genome-Wide Association Analyses of Ethiopian Sorghum Landrace Collection Reveal Loci Associated With Important Traits**. *Front Plant Sci* 2019, **10**.
156. Mengistu G, Shimelis H, Laing M, Lule D, Mathew I: **Genetic variability among Ethiopian sorghum landrace accessions for major agro-morphological traits and anthracnose resistance**. *Euphytica* 2020, **216**.

157. Navi SS, Bandyopadhyay R, Hall AJ, Bramel-Cox PJ: **A pictorial guide for the identification of mold fungi on sorghum grain. Information Bulletin no. 59 (In En. Summaries in En, Fr)**. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics; 1999.

158. Velazco JG, Rodriguez-Alvarez MX, Boer MP, Jordan DR, Eilers PHC, Malosetti M, van Eeuwijk FA: **Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model**. *Theor Appl Genet* 2017, **130**(7):1375-1392.

159. Revelle W: ***psych: Procedures for Psychological, Psychometric, and Personality Research* . Northwestern University, Evanston, Illinois. R package version 1.9.12,** https://CRAN.R-project.org/package=psych. In.; 2019.

160. Bates D, Machler M, Bolker B, Walker S: **Fitting Linear Mixed-Effects Models Using lme4**. *Journal of Statistical Software* 2015, **67**(1):1-48.

161. Wang J, Zhang Z: **GAPIT Version 3:An Interactive Analytical Tool for Genomic Association and Prediction. preprint.** 2018.

162. VanRaden PM: **Efficient methods to compute genomic predictions**. *J Dairy Sci* 2008, **91**(11):4414-4423.

163. Girma G, Nida H, Tirfessa A, Lule D, Bejiga T, Seyoum A, Mekonen M, Nega A, Dessalegn K, Birhanu C *et al*: **A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits**. *The plant genome* 2020, **13**(3):e20055.

164. Quinby JR, Karper RE: **Inheritance of Height in Sorghum**. *Agronomy Journal* 1954, **46**(5):211-216.

165. Mace ES, Jordan DR: **Location of major effect genes in sorghum (Sorghum bicolor (L.) Moench)**. *Theor Appl Genet* 2010, **121**:1339–1356.

166. Hilley J, Truong S, Olson S, Morishige D, Mullet J: **Identification of Dw1, a Regulator of Sorghum Stem Internode Length**. *PLoS One* 2016, **11**(3):e0151271.

167. Li X, Fridman E, Tesso TT, Yu J: **Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis**. *Proc Natl Acad Sci U S A* 2015, **112**(38):11823-11828.

168. Rhodes DH, Hoffmann L, Jr., Rooney WL, Ramu P, Morris GP, Kresovich S: **Genome-wide association study of grain polyphenol concentrations in global sorghum [Sorghum bicolor (L.) Moench] germplasm**. *J Agric Food Chem* 2014, **62**(45):10916-10927.

169. Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Tuinstra M, Morris GP: **Increased Power To Dissect Adaptive Traits in Global Sorghum Diversity Using a Nested Association Mapping Population**. *Genetics* 2017, **206**(2):573-585.

170. Wu YY, Li XR, Xiang WW, Zhu CS, Lin ZW, Wu Y, Li JR, Pandravada S, Ridder DD, Bai GH *et al*: **Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(26):10281-10286.

171. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al*: **Phytozome: a comparative platform for green plant genomics**. *Nucleic Acids Res* 2012, **40**(Database issue):D1178-1186.

172. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR: **Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments**. *Genome Biol* 2008, **9**(1):R7.

173. Hamaker BR, Mohamed AA, Habben JE, Huang CP, Larkins BA: **Efficient procedure for extracting maize and sorghum kernel proteins reveals higher prolamin contents than the conventional method**. *Cereal Chem* 1995, **72**:583-588.

174. Liu Y, Wang L, Xing X, Sun L, Pan J, Kong X, Zhang M, Li D: **ZmLEA3, a Multifunctional Group 3 LEA Protein from Maize (*Zea mays* L.), is Involved in Biotic and Abiotic Stresses**. *Plant and Cell Physiology* 2013, **54**(6):944–959.

175. Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K *et al*: **Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton**. *BMC Genet* 2018, **19**.

176. Chen LQ: **SWEET sugar transporters for phloem transport and pathogen nutrition**. *New Phytologist* 2014, **201**(4):1150-1155.

177. Gebauer P, Korn M, Engelsdorf T, Sonnewald U, Koch C, Voll LM: **Sugar Accumulation in Leaves of Arabidopsis sweet11/sweet12 Double Mutants Enhances Priming of the Salicylic Acid-Mediated Defense Response**. *Front Plant Sci* 2017, **8**.
178. Breia R, Conde A, Pimentel D, Conde C, Fortes AM, Granell A, Gerós H: **VvSWEET7 Is a Mono- and Disaccharide Transporter Up-Regulated in Response to Botrytis cinerea Infection in Grape Berries**. *Front Plant Sci* 2019, **10**.
179. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B: **Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae***. *New Phytologist* 2013, **200**(3):808-819.
180. Li S, Chakraborty N, Borcar A, Menze MA, Toner M, Hand SC: **Late embryogenesis abundant proteins protect human hepatoma cells during acute desiccation**. *Proc Natl Acad Sci U S A* 2012, **109**(51):20859-20864.
181. Taylor J, Bean SR, Ioerger BP, Taylor JRN: **Preferential binding of sorghum tannins with γ-kafirin and the influence of tannin binding on kafirin digestibility and biodegradation**. *Journal of cereal science* 2007, **46**:22-31.
182. Wu Y, Yuan L, Guo X, Holding DR, Messing J: **Mutation in the Seed Storage Protein Kafirin Creates a High-Value Food Trait in Sorghum**. *NATURE COMMUNICATIONS* 2013, **4:2217**.
183. Mace ES, Innes D, Hunt C, Wang X, Tao Y, Baxter J, Hassall M, Hathorn A, Jordan DR: **The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement**. *Theor Appl Genet* 2018, **132**(3):751-766.
184. Winn JA, Mason RE, Robbins AL, Rooney WL, Hays DB: **QTL Mapping of a High Protein Digestibility Trait in Sorghum bicolor**. *Int J Plant Genomics* 2009, **2009**.
185. Meckenstock DH, Gomez F, Rosenow DT, Guiragossian V: **Registration of 'sureno' sorghum**. *Crop Sci* 1993, **33**:213.
186. Song R, Segal G, Messing J: **Expression of the sorghum 10-member kafirin gene cluster in maize endosperm**. *Nucleic Acids Res* 2004, **32**(22):e189.
187. Xu JH, Messing J: **Amplification of Prolamin Storage Protein Genes in Different Subfamilies of the Poaceae**. *Theor Appl Genet* 2009, **119**(8):1397-1412.
188. Xu JH, Bennetzen JL, Messing J: **Dynamic gene copy number variation in collinear regions of grass genomes**. *Mol Biol Evol* 2012, **29**(2):861-871.

189. Shull JM, Watterson JJ, Kirleis AW: **Proposed nomenclature for the alcohol-soluble proteins (kafirins) of Sorghum bicolor (L. Moench) based on molecular weight, solubility, and structure**. *J Agric Food Chem* 1991, **39**:83-87.

190. Esen AA: **Proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (Zea Mays L.)**. *Journal of Cereal Science* 1987, **5**:117–128.

191. Belton PS, Delgadillo I, Halford NG, Shewry PR: **Kafirin structure and functionality**. *Journal of Cereal Science* 2006, **44**:272–286.

192. Xu D, Duan X, Wang B, Hong B, Ho T, Wu R: **Expression of a Late Embryogenesis Abundant Protein Gene, HVA1, from Barley Confers Tolerance to Water Deficit and Salt Stress in Transgenic Rice**. *Plant Physiol* 1996, **110**(1):249-257.

193. Chen ZY, Brown RL, Damann KE, Cleveland TE: **Identification of unique or elevated levels of kernel proteins in aflatoxin-resistant maize genotypes through proteome analysis**. *Phytopathology* 2002, **92**(10):1084-1094.

194. Chen ZY, Brown RL, Damann KE, Cleveland TE: **Identification of Maize Kernel Endosperm Proteins Associated with Resistance to Aflatoxin Contamination by Aspergillus flavus**. *Phytopathology* 2007, **97**(9):1094-1103.

195. Liu J, Li L, Foroud NA, Gong X, Li C, Li T: **Proteomics of Bulked Rachides Combined with Documented QTL Uncovers Genotype Nonspecific Players of the Fusarium Head Blight Responses in Wheat**. *Phytopathology* 2019, **109**(1):111-119.

196. Huang Y, Li L, Smith KP, Muehlbauer GJ: **Differential transcriptomic responses to Fusarium graminearum infection in two barley quantitative trait loci associated with Fusarium head blight resistance**. *BMC Genomics* 2016, **17**:387.

197. Goyal K, Walton LJ, Tunnacliffe A: **LEA proteins prevent protein aggregation due to water stress**. *Biochem J* 2005, **388**(Pt 1):151-157.

198. Hundertmark M, Hincha DK: **LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana**. *BMC Genomics* 2008, **9**:118.

199. Liu Y, Chakrabortee S, Li R, Zheng Y, Tunnacliffe A: **Both plant and animal LEA proteins act as kinetic stabilisers of polyglutamine-dependent protein aggregation**. *FEBS Lett* 2011, **585**(4):630-634.

200. Klemens PA, Patzke K, Deitmer J, Spinner L, Le Hir R, Bellini C, Bedu M, Chardon F, Krapp A, Neuhaus HE: **Overexpression of the vacuolar sugar carrier AtSWEET16 modifies germination, growth, and stress tolerance in Arabidopsis**. *Plant Physiol* 2013, **163**(3):1338-1352.
201. Styles ED, Ceska O: **Pericarp flavonoids in genetic strains of Zea mays**. *Maydica* 1989, **34**:227–237.
202. Lo SC, Nicholson RL: **Reduction of light induced anthocyanin accumulation in inoculated sorghum mesocotyls**. *Plant Physiology* 1998, **116**:979-989.
203. Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC: **The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein**. *Plant Cell* 1999, **11**(7):1337-1350.
204. Zhang B, Schrader A: **TRANSPARENT TESTA GLABRA 1-Dependent Regulation of Flavonoid Biosynthesis**. *Plants (Basel)* 2017, **6**(4).
205. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM: **Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings**. *Plant Journal* 2008, **53**(5):814-827.
206. Hipskind JD, Hanau R, Leite B, Nicholson RL: **Phytoalexin accumulation in sorghum - identification of an apigeninidin acyl ester**. *Physiol Mol Plant P* 1990, **36**(5):381–396.
207. Boddu J, Svabek C, Sekhon R, Gevens A, Nicholson RL, Jones AD, F. PJ, Gustine DL, S. C: **Expression of a putative flavonoid 3 '-hydroxylase in sorghum mesocotyls synthesizing 3-deoxyanthocyanidin phytoalexins**. *Physiol Mol Plant P* 2004, **65**(2):101–113.
208. Emmambux NM, Taylor JRN: **Sorghum kafirin interaction with various phenolic compounds**. *Journal of The Science of Food and Agriculture* 2003, **83**(5):402-407.
209. Awika JM, Rooney LW, Waniska RD: **Properties of 3-deoxyanthocyanins from sorghum**. *J Agric Food Chem* 2004, **52**(14):4388-4394.
210. Ren D, Liu Y, Yang KY, Han L, Mao G, Glazebrook J, Zhang S: **A fungal-responsive MAPK cascade regulates phytoalexin biosynthesis in *Arabidopsis***. *Proc Natl Acad Sci U S A* 2008, **105**(14):5638–5643.

211. Eckardt NA: **Induction of Phytoalexin Biosynthesis: WRKY33 Is a Target of MAPK Signaling**. *Plant Cell* 2011, **23**(4):1190.
212. Kishi-Kaboshi M, Takahashi A, Hirochika H: **MAMP-responsive MAPK cascades regulate phytoalexin biosynthesis**. *Plant Signal Behav* 2010, **5**(12):1653-1656.
213. Li S, Wang W, Gao J, Yin K, Wang R, Wang C, Petersen M, Mundy J, Qiu JL: **MYB75 Phosphorylation by MPK4 Is Required for Light-Induced Anthocyanin Accumulation in Arabidopsis[OPEN]**. *Plant Cell* 2016, **28**(11):2866-2883.
214. Nida H, Lee S, Li Y, Mengiste T: **Transcriptome analysis of early stages of sorghum grain mold disease reveals defense regulators and metabolic pathways associated with resistance**. *BMC Genomics* 2021, **22**(1):295.
215. Rooney WL, Miller FR, Frederiksen RA: **Registration of Tx2911 Sorghum Germplasm**. *Crop Science* 2000, **40**:584.
216. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**. *Nat Protoc* 2012, **7**(3):562-578.
217. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nat Methods* 2015, **12**(4):357-360.
218. Ge SX, Jung D, Yao R: **ShinyGO: a graphical enrichment tool for animals and plants**. *Bioinformatics* 2019.
219. Gibson UE, Heid CA, Williams PM: **A novel method for real time quantitative RT-PCR**. *Genome Res* 1996, **6**(10):995-1001.
220. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative CT method**. *Nature Protocols* 2008, **3**(6):1101-1108.
221. Ward E, Uknes S, Williams S, Dincher S, Wiederhold D, Alexander D, Ahl-Goy P, Metraux J, Ryals J: **Coordinate Gene Activity in Response to Agents That Induce Systemic Acquired Resistance**. *Plant Cell* 1991, **3**(10):1085-1094.
222. Cao H, Bowling SA, Gordon AS, Dong X: **Characterization of an Arabidopsis Mutant That Is Nonresponsive to Inducers of Systemic Acquired Resistance**. *The Plant cell* 1994, **6**(11):1583-1592.

223. Pieterse CM, van Wees SC, van Pelt JA, Knoester M, Laan R, Gerrits H, Weisbeek PJ, van Loon LC: **A novel signaling pathway controlling induced systemic resistance in Arabidopsis**. *The Plant cell* 1998, **10**(9):1571-1580.

224. Lee EJ, Facchini P: **Norcoclaurine Synthase Is a Member of the Pathogenesis-Related 10/Bet v1 Protein Family[W]**. In: *Plant Cell.* vol. 22; 2010: 3489-3503.

225. Stotz HU, Thomson JG, Wang Y: **Plant defensins: Defense, development and application**. In: *Plant Signal Behav.* vol. 4; 2009: 1010-1012.

226. Broekaert WF, Terras FR, Cammue BP, Osborn RW: **Plant defensins: novel antimicrobial peptides as components of the host defense system**. *Plant Physiol* 1995, **108**(4):1353-1358.

227. Cao Y, Liang Y, Tanaka K, Nguyen CT, Jedrzejczak RP, Joachimiak A, Stacey G: **The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1**. In: *eLife.* vol. 3; 2014.

228. Song W, Wang B, Li X, Wei J, Chen L, Zhang D, Zhang W, Li R: **Identification of Immune Related LRR-Containing Genes in Maize (Zea mays L.) by Genome-Wide Sequence Analysis**. *Int J Genomics* 2015, **2015**.

229. Gomez-Gomez L, Boller T: **FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis**. *Mol Cell* 2000, **5**(6):1003-1011.

230. Xu X, Chen C, Fan B, Chen Z: **Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors**. *The Plant cell* 2006, **18**(5):1310-1326.

231. Major IT, Yoshida Y, Campos ML, Kapali G, Xin X, Sugimoto K, de Oliveira Ferreira D, He SY, Howe GA: **Regulation of growth–defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module**. In: *New Phytol.* vol. 215; 2017: 1533-1547.

232. Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA: **Different mechanisms for phytoalexin induction by pathogen and wound signals in Medicago truncatula**. *Proc Natl Acad Sci U S A* 2007, **104**(46):17909-17915.

233. Stone SL, Hauksdóttir H, Troy A, Herschleb J, Kraft E, Callis J: **Functional Analysis of the RING-Type Ubiquitin Ligase Family of Arabidopsis1[w]**. In: *Plant Physiol.* vol. 137; 2005: 13-30.

234. Jambunathan N, McNellis TW: **Regulation of Arabidopsis COPINE 1 Gene Expression in Response to Pathogens and Abiotic Stimuli1**. In: *Plant Physiol.* vol. 132; 2003: 1370-1381.

235. Zou B, Hong X, Ding Y, Wang X, Liu H, Hua J: **Identification and analysis of copine/BONZAI proteins among evolutionarily diverse plant species**. *Genome* 2016, **59**(8):565-573.

236. Neu D, Lehmann T, Elleuche S, Pollmann S: **Arabidopsis amidase 1, a member of the amidase signature family**. *Febs j* 2007, **274**(13):3440-3451.

237. Robert-Seilaniantz A, Grant M, Jones JD: **Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism**. *Annu Rev Phytopathol* 2011, **49**:317-343.

238. Denancé N, Sánchez-Vallet A, Goffner D, Molina A: **Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs**. *Front Plant Sci* 2013, **4**.

239. LeClere S, Tellez R, Rampey RA, Matsuda SPT, Bartel B: **Characterization of a Family of IAA-Amino Acid Conjugate Hydrolases from Arabidopsis**. 2002.

240. Cai D, Kleine M, Kifle S, Harloff HJ, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijn EM, Lange W, Stiekema WJ *et al*: **Positional cloning of a gene for nematode resistance in sugar beet**. *Science* 1997, **275**(5301):832-834.

241. Sivasankar S, Sheldrick B, Rothstein SJ: **Expression of Allene Oxide Synthase Determines Defense Gene Activation in Tomato1**. In: *Plant Physiol.* vol. 122; 2000: 1335-1342.

242. Simpson TD, Gardner HW: **Allene Oxide Synthase and Allene Oxide Cyclase, Enzymes of the Jasmonic Acid Pathway, Localized in Glycine max Tissues**. In: *Plant Physiol.* vol. 108; 1995: 199-202.

243. Qamar A, Mysore KS, Senthil-Kumar M: **Role of proline and pyrroline-5-carboxylate metabolism in plant defense against invading pathogens**. *Front Plant Sci* 2015, **6**:503.

244. Anwar A, She M, Wang K, Riaz B, Ye X: **Biological Roles of Ornithine Aminotransferase (OAT) in Plant Stress Tolerance: Present Progress and Future Perspectives**. In: *Int J Mol Sci.* vol. 19; 2018.

245. Liu X, Yang W, Li Y, Li S, Zhou X, Zhao Q, Fan Y, Lin M, Chen R: **The intergenic region of the maize defensin-like protein genes Def1 and Def2 functions as an embryo-specific asymmetric bidirectional promoter**. *J Exp Bot* 2016, **67**(14):4403-4413.

246. Penninckx IA, Eggermont K, Terras FR, Thomma BP, De Samblanx GW, Buchala A, Metraux JP, Manners JM, Broekaert WF: **Pathogen-induced systemic activation of a plant defensin gene in Arabidopsis follows a salicylic acid-independent pathway**. *The Plant cell* 1996, **8**(12):2309-2323.

247. Nuruzzaman M, Sharoni AM, Kikuchi S: **Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants**. *Front Microbiol* 2013, **4**.

248. Xie Q, Sanz-Burgos AP, Guo H, Garcia JA, Gutierrez C: **GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein**. *Plant Mol Biol* 1999, **39**(4):647-656.

249. Park CJ, Seo YS: **Heat Shock Proteins: A Review of the Molecular Chaperones for Plant Immunity**. In: *Plant Pathol J.* vol. 31; 2015: 323-333.

250. Huang S, Monaghan J, Zhong X, Lin L, Sun T, Dong OX, Li X: **HSP90s are required for NLR immune receptor accumulation in Arabidopsis**. *Plant J* 2014, **79**(3):427-439.

251. Schulze-Lefert P: **Plant immunity: the origami of receptor activation**. *Curr Biol* 2004, **14**(1):R22-24.

252. Jirage D, Tootle TL, Reuber TL, Frost LN, Feys BJ, Parker JE, Ausubel FM, Glazebrook J: **Arabidopsis thaliana PAD4 encodes a lipase-like gene that is important for salicylic acid signaling**. In: *Proc Natl Acad Sci U S A.* vol. 96; 1999: 13583-13588.

253. Zhou N, Tootle TL, Tsui F, Klessig DF, Glazebrook J: **PAD4 functions upstream from salicylic acid to control defense responses in Arabidopsis**. *The Plant cell* 1998, **10**(6):1021-1030.

254. Hunter BG, Beatty MK, Singletary GW, Hamaker BR, Dilkes BP, Larkins BA, Jung R: **Maize Opaque Endosperm Mutations Create Extensive Changes in Patterns of Gene Expression**[W]. In: *Plant Cell.* vol. 14; 2002: 2591-2612.

255. Wu Y, Holding DR, Messing J: **γ-Zeins are essential for endosperm modification in quality protein maize**. In: *Proc Natl Acad Sci U S A.* vol. 107; 2010: 12810-12815.
256. Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiyama C, Dohmae N, Takio K, Minami E, Shibuya N: **Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor**. In: *Proc Natl Acad Sci U S A.* vol. 103; 2006: 11086-11091.
257. Kishi-Kaboshi M, Okada K, Kurimoto L, Murakami S, Umezawa T, Shibuya N, Yamane H, Miyao A, Takatsuji H, Takahashi A *et al*: **A rice fungal MAMP-responsive MAPK cascade regulates metabolic flow to antimicrobial metabolite synthesis**. *Plant J* 2010, **63**(4):599-612.
258. Mao G, Meng X, Liu Y, Zheng Z, Chen Z, Zhang S: **Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis**. *Plant Cell* 2011, **23**(4):1639-1653.
259. Qiu JL, Fiil BK, Petersen K, Nielsen HB, Botanga CJ, Thorgrimsen S, Palma K, Suarez-Rodriguez MC, Sandbech-Clausen S, Lichota J *et al*: **Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus**. *Embo j* 2008, **27**(16):2214-2221.
260. Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B: **Plant 'pathogenesis-related' proteins and their role in defense against pathogens**. *Biochimie* 1993, **75**(8):687-706.
261. Finkina EI, Melnikova DN, Bogdanov IV, Ovchinnikova TV: **Plant Pathogenesis-Related Proteins PR-10 and PR-14 as Components of Innate Immunity System and Ubiquitous Allergens**. *Curr Med Chem* 2017, **24**(17):1772-1787.
262. Wu J, Kim SG, Kang KY, Kim JG, Park SR, Gupta R, Kim YH, Wang Y, Kim ST: **Overexpression of a Pathogenesis-Related Protein 10 Enhances Biotic and Abiotic Stress Tolerance in Rice**. In: *Plant Pathol J.* vol. 32; 2016: 552-562.
263. Park CJ, Kim KJ, Shin R, Park JM, Shin YC, Paek KH: **Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway**. *Plant J* 2004, **37**(2):186-198.
264. Bufe A, Spangfort MD, Kahlert H, Schlaak M, Becker WM: **The major birch pollen allergen, Bet v 1, shows ribonuclease activity**. *Planta* 1996, **199**(3):413-415.

265. Legrand M, Kauffmann S, Geoffroy P, Fritig B: **Biological function of pathogenesis-related proteins: Four tobacco pathogenesis-related proteins are chitinases**. *Proc Natl Acad Sci U S A* 1987, **84**(19):6750-6754.

266. Lay FT, Anderson MA: **Defensins--components of the innate immune system in plants**. *Curr Protein Pept Sci* 2005, **6**(1):85-101.

267. Terras FR, Eggermont K, Kovaleva V, Raikhel NV, Osborn RW, Kester A, Rees SB, Torrekens S, Van Leuven F, Vanderleyden J *et al*: **Small cysteine-rich antifungal proteins from radish: their role in host defense**. *Plant Cell* 1995, **7**(5):573-588.

268. Sels J, Mathys J, De Coninck BM, Cammue BP, De Bolle MF: **Plant pathogenesis-related (PR) proteins: a focus on PR peptides**. *Plant Physiol Biochem* 2008, **46**(11):941-950.

269. Mengiste T: **Plant immunity to necrotrophs**. *Annu Rev Phytopathol* 2012, **50**:267-294.

270. Lin P, Wong JH, Ng TB: **A defensin with highly potent antipathogenic activities from the seeds of purple pole bean**. *Biosci Rep* 2009, **30**(2):101-109.

271. Pieterse CM, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SC: **Hormonal modulation of plant immunity**. *Annual review of cell and developmental biology* 2012, **28**:489-521.

272. Fu F, Girma G, Mengiste T: **Global mRNA and microRNA expression dynamics in response to anthracnose infection in sorghum**. *BMC Genomics* 2020, **21**(1):760.

273. Yazawa T, Kawahigashi H, Matsumoto T, Mizuno H: **Simultaneous transcriptome analysis of Sorghum and Bipolaris sorghicola by using RNA-seq in combination with de novo transcriptome assembly**. *PLoS One* 2013, **8**(4):e62460.

274. Ahuja I, Kissen R, Bones AM: **Phytoalexins in defense against pathogens**. *Trends Plant Sci* 2012, **17**(2):73-90.

275. Jeandet P: **Phytoalexins: Current Progress and Future Prospects**. In: *Molecules.* vol. 20; 2015: 2770-2774.

276. Klein AP, Anarat-Cappillino G, Sattely ES: **Three Cytochromes P450 are Sufficient to Reconstitute the Biosynthesis of Camalexin, a Major Arabidopsis Antibiotic****. *Angew Chem Int Ed Engl* 2013, **52**(51).

277. Nielsen LJ, Stuart P, Pičmanová M, Rasmussen S, Olsen CE, Harholt J, Møller BL, Bjarnholt N: **Dhurrin metabolism in the developing grain of Sorghum bicolor (L.) Moench investigated by metabolite profiling and novel clustering analyses of time-resolved transcriptomic data**. In: *BMC Genomics.* vol. 17; 2016.

278. Havko NE, Major IT, Jewell JB, Attaran E, Browse J, Howe GA: **Control of Carbon Assimilation and Partitioning by Jasmonate: An Accounting of Growth-Defense Tradeoffs**. *Plants (Basel)* 2016, **5**(1).

279. Qi TC, Song SS, Ren QC, Wu DW, Huang H, Chen Y, Fan M, Peng W, Ren CM, Xie DX: **The Jasmonate-ZIM-Domain Proteins Interact with the WD-Repeat/bHLH/MYB Complexes to Regulate Jasmonate-Mediated Anthocyanin Accumulation and Trichome Initiation in Arabidopsis thaliana**. *Plant Cell* 2011, **23**(5):1795-1814.

280. Xiaochen X: **Identification and mapping of anthracnose resistance genes in sorghum [S*orghum bicolor (*L.) Moench]**. West Lafayette, IN, USA: Purdue University; 2019.

281. Xin Z, Velten JP, Oliver MJ, Burke JJ: **High-throughput DNA extraction method suitable for PCR**. *Biotechniques* 2003, **34**(4):820-824, 826.

282. Thorvaldsdóttir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. *Briefings in Bioinformatics* 2013, **14**(2):178-192.

283. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE: **The Phyre2 web portal for protein modeling, prediction and analysis**. *Nature Protocols* 2015, **10**(6):845-858.

284. Bianchet C, Wong A, Quaglia M, Alqurashi M, Gehring C, Ntoukakis V, Pasqualini S: **An Arabidopsis thaliana leucine-rich repeat protein harbors an adenylyl cyclase catalytic center and affects responses to pathogens**. *J Plant Physiol* 2019, **232**:12-22.

285. Jupe F, Pritchard L, Etherington GJ, Mackenzie K, Cock PJ, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JD *et al*: **Identification and localisation of the NB-LRR gene family within the potato genome**. *BMC Genomics* 2012, **13**:75.

286. Yi H, Richards EJ: **A cluster of disease resistance genes in Arabidopsis is coordinately regulated by transcriptional activation and RNA silencing**. *Plant Cell* 2007, **19**(9):2929-2939.

287. Michelmore RW, Meyers BC: **Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process**. *Genome Res* 1998, **8**(11):1113-1130.

288. Christie N, Tobias PA, Naidoo S, Külheim C: **The Eucalyptus grandis NBS-LRR Gene Family: Physical Clustering and Expression Hotspots**. *Front Plant Sci* 2015, **6**:1238.

289. Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB: **Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula**. *Plant Physiol* 2008, **146**(1):5-21.

290. Wang T, Jia ZH, Zhang JY, Liu M, Guo ZR, Wang G: **Identification and Analysis of NBS-LRR Genes in Actinidia chinensis Genome**. *Plants (Basel)* 2020, **9**(10).

291. DeYoung BJ, Innes RW: **Plant NBS-LRR proteins in pathogen sensing and host defense**. *Nat Immunol* 2006, **7**(12):1243-1249.

292. Gehring C: **Adenyl cyclases and cAMP in plant signaling - past and present**. *Cell Commun Signal* 2010, **8**:15.

293. Zhao J, Guo Y, Fujita K, Sakai K: **Involvement of cAMP signaling in elicitor-induced phytoalexin accumulation in Cupressus lusitanica cell cultures**. *New Phytologist* 2004, **161**(3):723-733.

294. Świeżawska B, Jaworski K, Pawełek A, Grzegorzewska W, Szewczuk P, Szmidt-Jaworska A: **Molecular cloning and characterization of a novel adenylyl cyclase gene, HpAC1, involved in stress signaling in Hippeastrum x hybridum**. *Plant Physiology and Biochemistry* 2014, **80**:41-52.

295. Wang W, Feng B, Zhou JM, Tang D: **Plant immune signaling: Advancing on two frontiers**. *J Integr Plant Biol* 2020, **62**(1):2-24.

296. Blanco E, Fortunato S, Viggiano L, de Pinto MC: **Cyclic AMP: A Polyhedral Signalling Molecule in Plants**. *International journal of molecular sciences* 2020, **21**(14):4862.