# GENETIC CONTROL OF ANTHRACNOSE RESISTANCE IN SORGHUM

by

**Demeke Mewa Bayable**


**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*


**Doctor of Philosophy**
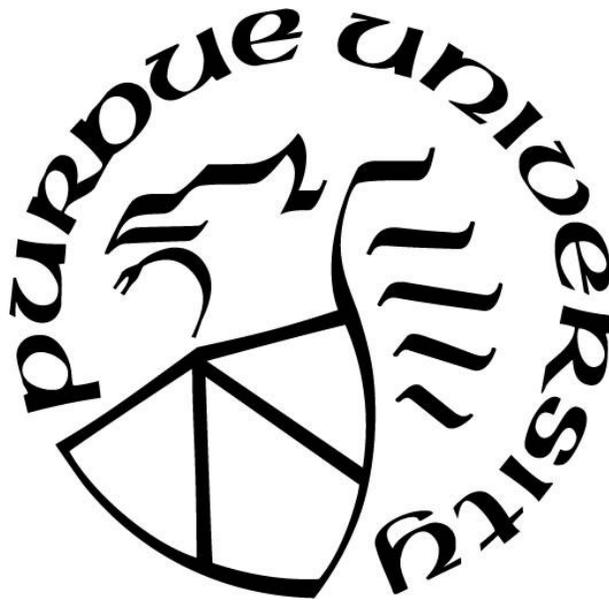


Department of Botany and Plant Pathology

West Lafayette, Indiana

December 2020

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Sorghum [*Sorghum bicolor* (L) Moench] is a major cereal crop of global importance. The crop is grown for food, feed, biofuel and other applications. Among major production constraints of the crop worldwide is sorghum anthracnose that is caused by the fungal pathogen *Colletotrichum sublineolum*. Despite the extensive diversity for genetic resistance to the pathogen in the sorghum natural variation, specific resistance genes have not been identified. Consequently, the molecular mechanisms of resistance to the disease are poorly understood. The paucity of knowledge on resistance genes has hampered crop improvement and studies on mechanisms of resistance to the resistance. In this study, we report the identification of two *ANTHRACNOSE RESISTANCE GENES* (*ARG2* and *ARG3)* that control sorghum resistance to anthracnose. Initially, by screening a collection of sorghum natural variants, we identified genotypes that carry race specific resistance to strains of the pathogen. Bulked Segregant Analyses of genomic DNA raw-reads (BSA-Seq) in biparental mapping populations followed by recombination and comparative genomic analyses were conducted to determine the specific sequence variation in *ARG2* and *ARG3* loci that underpin the differences in disease phenotype. The identity of the *ARG* genes and their genetic link to resistance were validated using independent alleles from distinct germplasm sources. The *ARG2* locus harbors a cluster of duplicate NB-LRR genes with complex structural and functional relationships in different sorghum lines. *ARG2* encodes a canonical resistance protein with CC-NBS-LRR structure and confers complete resistance to some strains of the pathogen. Near-isogenic lines (NILs) that differ in the *ARG2* locus showed that the expression of defense response genes is associated to the resistance and provided further insight to the effect of *ARG2* locus on the growth of the resistant lines. The *ARG3* locus has significant genomic rearrangements in different sorghum lines and a unique genomic feature in the resistant line. *ARG3* is likely a novel resistance gene with no homolog identified in the plant kingdom. Intriguingly, *ARG3* is located within a large genomic region showing striking synteny of genes between rice and sorghum. *ARG3* confers resistance to a strain of the pathogen that is virulent on sorghum lines that are known for broad-spectrum resistance. The identification of *ARG2* and *ARG3* will contribute to breeding for sorghum anthracnose resistance and to studies on the genetic control of anthracnose disease. Additional implications of these findings, future studies and the advances made over previous studies are discussed.

# CHAPTER 1.     SORGHUM ANTHRACNOSE DISEASE RESISTANCE AND LESSONS FROM RELATED PATHOSYSTEMS

## 1.1     Introduction

To satisfy the growing demand for food, global agricultural production must increase by 70% by 2050 (1). The food and agricultural organization of the UN (FAO) estimated that 20 to 40 percent of global crop production are lost to pests, and diseases that cost the global economy around $220 billion per year (2). In maize, the loss of attainable yield due to diseases was estimated at 16% (3), and fungal pathogens account for the lion's share of losses due to plant diseases (4). Yield losses to pathogens hugely vary by pathogen and crop combinations. Environmental factors and agricultural practices are major factors in changing the incidence and the severity of plant diseases. Climate change has been predicted to bring an additional layer of factors in plant disease dynamics by influencing pathogen virulence, ecological range of pathogens, host physiology and host range (5, 6). There is a need for smarter disease control strategies, in the face of projected changes in climate and land-use patterns which may alter the proportion of potential plant pathogens, emergence of new pathogens and reemergence of others worldwide. The contribution of genetics to crop improvement and the importance of host resistance and genetic diversity to crop security has been enormous (7-9). Genetic resistance in the context of a warming environment will be an even more significant task in the future of plant disease resistance.

Sorghum (*Sorghum bicolor* Moench) is a major cereal crop of global importance. It is a source of food, feed and biofuel. Productivity of the crop has long been limited by diseases (10), of which, anthracnose caused by *Colletotrichum sublineolum* is economically important in many sorghum growing regions (11, 12). The importance of sorghum anthracnose disease dates back long before Cs was recognized as a species that is distinct from Cg (13, 14). Previously, Cs was considered as part of *Colletotrichum graminicola* and was reclassified (13) recently as a species by its own. *S. bicolor* is known as the primary host to the pathogen though some other alternate hosts are reported (13, 15-20). Humidity and high rainfall promote infection (21), increase severity of the disease resulting in reduced grain size (22) and up to 50% or more loss of grain and forage yield (22). The main economic loss and the elaborated disease phenotype is manifested on the leaf tissues though old literatures associate the pathogen effect to most other areal plant parts (23, 24). This foliar disease is often referred to as Sorghum Anthracnose disease (SA). Despite the

significance of the disease, there has not been concerted efforts to dissect the genetic and molecular mechanisms of pathogenesis and virulence in the fungus.

This chapter reviews previous observations, current state of knowledge, highlights knowledge gap and point to future directions in the anthracnose disease and genetic resistance as a background to the current research. Studies in related pathosystems are highlighted to leverage the knowledge with a bearing on our understanding of host-pathogen interaction in sorghum. Current concepts in genetics disease resistance, and the molecular basis of host responses are highlighted. The knowledge base on *Colletotrichum sublineolum* (Cs) is not extensive, but I have summarized the available resources about some common features and interspecific relatedness in the cosmopolitan *Colletotrichum* species. Cs Pathogen variability, host range and lifecycle with implications in understanding host resistance mechanisms are covered with a genus-wide perspective. A summary of molecular, cellular and phenotypic responses on both the pathogen and the host are described. The genetics, genomics, breeding, and molecular biology of SA host-resistance is emphasized and contextualized with the current project. The genetics and genomics of host-resistance in closely related pathosystems is highlighted to address common and specific aspects. Finally, non-genic control of SA and its likely implications is noted for completeness.

## 1.2    Pathogenic lifestyle

A micro-organism that associates with another organism for a net negative effect on the host, usually for carbon source, is regarded as a pathogen. Two distinct modes of pathogen nutrition have been characterized. A pathogen that takes aggressive virulence strategies to destruct the host tissue and grow on nutrients obtained from dead tissue is known as necrotroph. Many necrotrophs that infect a limited number of hosts produce notable host-selective toxins (25) whereas necrotrophs that infect a broad range of host species may rely on a plethora of disease factors such as cell wall degrading enzymes and toxins (26). Biotrophic pathogens tend to minimize the destructive phase and acquire nutrition clandestinely from live host tissue with specialized and intricate molecular mechanisms (27, 28). Thus, biotrophy seems a result of the host-pathogen co-evolution for co-existence that may limit the range of hosts to a pathogen. In nature, exclusive necrotrophy or biotrophy may not be common, and the mode of nutrition of a pathogenic species may have been ascribed to the notable and discernable mode of nutrition. Pathogenic species that are recognized by both biotrophic and necrotrophic modes of nutrition are

known as hemi-biotrophs. Hemi-biotrophy begins with biotrophy and switch to necrotrophy though difficulties to delineating the former from the symptomatic stage of a typical necrotrophy may exist (27).

## 1.3   Anthracnose disease and *Colletotrichum species*

The genus *Colletotrichum* (Ct) encompasses a diversity of cosmopolitan fungi (29, 30). Pathogenic Ct species are among the top 10 economically important plant pathogens worldwide (31, 32). The fungi cause diseases on animals (33) and human (34) as well. Recent disease reports show the spread of the pathogens where they had not been known (35). Ct pathogens cause damage to most plant parts including roots, stems, leaves, flowers and fruits (36), but they are often highly specific to individual tissues (37). Many pathogenic Ct species are also specific to particular plant species and cultivars (37). The crop losses occur pre-harvest and post-harvest (31) in both temperate and tropical environment (37).

### 1.3.1   Host range

The fungi of the genus Ct are amorphous (38) and evolutionary relationships among species remains poorly resolved (39); a recent review reported that the genus Ct consists of 29 to 700 species (40). The fungi infect an extremely wide range of plant species (37): from a diverse species of grasses (41) to legumes (42), and from various fruit crops (43) to the rubber tree (44). A species-complex approach is widely employed towards the dissection of phylogenetic relationships and pathogenic traits. Thirteen Ct species-complexes have been reported based on multi-locus markers (39, 45): *C. destructivum* (46), *acutatum* (47), *truncatum* (48), *orbiculare* (49), *boninense* (50), *gigasporum* (51), *siamense (45) or gloeosporioides* (52), *dracaenpohilum*, *magnum* and *orchidearum* (53), *graminicola*, *dematium* and *spaethianum*. The clusters were named after the first identified or the more known species within each species complex (39). In contrast to most of the species complexes, the placement of grass-infecting Ct species has not been satisfactory. The set of molecular markers that provide considerable resolution to most of the species complexes hardly resolved that of *C. graminicola* (39).

In a recent report, 14 closely related graminicolous Ct species were clustered in a *C. graminicola* (Cg) species complex (54). A morphology-based study supported *C. caudatum,* a

species closely related to the Cg species complex, as a single species with a broad host range whereas multi-locus phylogenetic study did not (55). Accordingly, this study treated *C. caudatum* as a species complex by its own. Several morphologically cryptic species were identified as closely related to *C. caudatum,* and the sister taxon placement of *C. caudatum* remained unclear. In addition, phylogenetic analyses suggest *C. cereale* is distinct from the Cg species complex (55-57).

Species in the Cg and *C. caudatum* species complexes are pathogenic to the warm-season grasses (39, 58), and the host range of grass infecting individual species is ambiguous. The well-known host of Cg, as a single distinct species, is *Zea mays* though Cg had been in use for other Ct species within the Cg species complex. Cg was shown to colonize the roots and leaves of sorghum and reported as pathogenic to human (54) as well. The Cg species complex encompass *C. navitas and C. sublineolum. C. navitas,* a sister taxon to Cg, causes anthracnose on the perennial tallgrass prairie switchgrass (59) whereas *C. caudatum* infects numerous C4 grasses including *Eragrostis* and unspecified prairie grasses (55). On the other hand, *C. eremochloae*, a sister taxon shares common phenotype with *C. sublineolum* that causes anthracnose in the widely grown lawn grasses (38, 39). The cool-season adapted *C. cereale* was reported to cause anthracnose outbreaks in wheat, oat, barley, rye and turfgrass in US over the 20th century which was later known to survive as an endophyte (60).The recent genomic analysis of many Ct species (61) may help to lay a platform for a profound study of relationships especially among those pathogenic to the warm-season grasses.

### 1.3.2 Lifestyle

The lifestyle of Ct species have been broadly-categorized as necrotrophic, hemibiotrophic, latent and endophytic (62). The endophytic and hemibiotrophic lifestyles are common (63). The pathogenic Ct are widely known for the hemi-biotrophic mode of nutrition, although a wide range of variation in the duration and distinctness of biotrophy to necrotrophy exists. The difference in lifestyle depends on Ct species, host species, physiological state of the host and environmental conditions, and particularities within the species complexes exist (62-65). There are three major variants of Ct species regarding the duration of the biotrophic phase (66). However, we did not find a report regarding a completely biotrophic or a necrotrophic Ct lifestyle. The anamorphic *C. gloeosporioides* showed subcuticular necrotrophic feature at a higher concentration of $H_2O_2$ and

intracellular hemi-biotrophic when $H_2O_2$ decreases (67). A study on a set of several fungal pathogens in which Ct was represented by any species showed a minimum of four shifts from endophytic to necrotrophic lifestyle whereas biotrophs did not contain endophytes or necrotrophs (68).

The term endophyte encompass commensalistic, latent pathogenic, latent saprophytic and mutualistic relationships (69). A study described latent but effective colonization of the sorghum leaf tissue by Ct (70). Interestingly, many endophytic isolates belong to important pathogenic species(62) whereas many others contribute to host-resistance (68). The endophytic *C. tofieldiae* that colonized the roots of *A. thaliana* was shown to spread systemically to the shoot and promote growth during phosphate starvation (71) whereas host defense response was activated under phosphate-sufficient conditions (72). However, in some localized areas, the epidermal and cortical cells were packed with black microsclerotia (71), which undermines the distinctions and transitions among the mutualistic, latent and biotrophic lifestyles. *Harpophora oryzae* showed a shift from a pathogenic to a mutualistic endophyte and conferred systemic resistance (69, 73). Interestingly, we did not find a report on endophytism in the extensively studied *Magnaporthe oryzae*. The current -omics technologies have started to elucidate the genetic, physiological and environmental factors that account to the lifestyle switch (40, 74) and mutualistic versus pathogenic interaction (75). More inclusive and varied studies may provide a deeper insight to the alternating lifestyle (76). The potentially exposure of the cosmopolitan endophytic and hypervariable Ct to the broad environmental conditions where tropical grasses grow (30) may have conditioned the flexible lifestyle.

### 1.3.3    Mechanisms of pathogenesis and virulence

Understanding the sources and processes that determine successful pathogenesis is a prerequisite for detailed biochemical and molecular work and increasingly relevant to the practical sides of plant pathology (37). Recently, however, the genetics and molecular biology of infection strategies and pathogen virulence had not received sufficient attention in the Ct pathosystems. The recently sequenced genomes of 25 Ct species has become a milestone (22, 61), and the analysis on genomic signatures and endophytism in five grass-infecting Ct species (77) has reinforced the background for molecular studies of host-pathogen interaction to anthracnose diseases. The dispensability of a chromosome in *C. higginsianum* (78) may facilitate the dissection of interesting

molecular pathways. Specific gene family losses and gains were revealed in a diverse species of Ct that have been associated to the host-pathogen arms race whereby pathogens minimize the chance of recognition by the host cells surveillance system (79). A positive selection acting on secreted and nuclear proteins may condition adaptation to specific hosts and ecological niches (79).

Fungal pathogens secrete an enormous, small-proteins that may serve as effector-candidates. Recently, in Ct species, several small secreted proteins (SSP) have been predicted through comparative genomic analyses. Many genes that were predicted to encode SSP lack orthologs in Cs and Cg (80, 81). Similarly, the analysis of nine Ct species has captured 288 Cs and 247 Cg effector candidates and 12 conserved domains; eight of them were induced during the biotrophic infection phase of the bean pathogen *C. lindemuthianum*a (74). As opposed to the meager knowledge base in Ct, over 30 genes that condition rice blast disease have been identified in *M. grisea* (82). Recently, unique effectors targeting host-proteins involved in defense response to a necrotrophic pathogen has also been characterized in wheat (83).

The molecular basis of pathogenesis has been examined through comparative studies. The aggressive necrotrophic phase that dismantles the cuticle of the host plant tissue is universal among the pathogenic Ct species (32). This commonality was associated to the retention of the cutinase domain both in structure and arrangement in 34 Ct species (32). A tailored profile of carbohydrate-degrading enzymes according to their infection lifestyle, and lineage specific expansion of necrosis and ethylene-inducing peptide 1 Nep1-like protein families contribute to pathogen success (79, 84). Additional pathogenesis related enzymes includes alkaloids, terpenoids, polyketides and non-ribosomal peptides (85). Secretion of ammonia enhanced pH and pathogenicity (86). Many specialized secondary metabolites-associated proteins that lack orthologs in Cs and Cg, and common groups of non-conserved proteins including transporters, transcription factors, and CAZymes were reported (80, 81). The recent comparative genomic information based on many Ct species may facilitate the identification of many more secondary metabolites (85).

The molecular basis of evasion of recognition and neutralization of host defense response has also been examined. The surface modification of the fungal cell walls by chitin de-$N$-acetylation protected the fungal cell walls from enzymatic hydrolysis by host chitinases and auto-catalytic defense response system in the invaded host tissue (87). Similarly, some fungal endophytes produce chitin modifying enzymes (88). The dispensable chromosome in *C. higginsianum* is required to suppress defense response via tryptophan-derived metabolites in

*Arabidopsis* (78). A study on rice blast showed glutathione reductase plays key role to neutralize plant generated ROS, and thioredoxin protein contributes to cell-wall integrity; thereby facilitate biotrophic colonization of host cells by *M. oryzae* (89).

The consistence between pathogen lifestyle and host range calls for a diversified approach to better understand Ct species. The hemi-biotrophy lifestyle in many Ct species tends to correlate with host specificity whereas the subcuticular phase tend to associate to a wider host range (85). Ct species with narrow host range tend to show lineage specific loss of genes that encode carbohydrate-active enzymes and proteases whereas Ct species with wider host ranges tend to retain multiple copies of these families of genes (84). Five of the 13 Ct species complexes have pathogenic species to peppers (90), which may mirror either a widely conserved mechanism or a panel of mechanisms of pathogenesis. Similarly, the wide variety of monocot and dicot hosts, including sorghum, to *C. gloeosporioides* may reflect an interesting communalities of pathogenesis (52, 91, 92)*; C. incanum* also infects monocots and dicots *(79)*. Plotting the mechanisms behind the pathogenicity of Cs across grasses of at least the three genera (*Sorghum*, *Rottboellia, Zizania*) may also help elucidate the evolution of host range. In general, the functional diversity of pathogenic Ct species to a single host, and inversely the diversity of hosts infected by a single pathogenic species vis-à-vis the triggers to biotrophy-necrotrophy switch, may help to figure out commonalities and variations in the molecular mechanisms of host-pathogen interactions.

Non-canonical infection processes were reported in some Ct species. A study described latent but effective colonization of cowpea leaf tissue by unidentified Ct taxa (70). In contrast to the common mechanisms by Ct pathogens, this fungus did not penetrate the epidermis, rather shown to find its way to the leaf tissue through the stomatal openings. The development of anthracnose disease symptom was slow (70). The endophytic *C. tofieldiae* infecting *Arabidopsis* root epidermal cells directly by means of undifferentiated hyphae rather than appressoria that spread systemically into shoots was shown to promote growth under phosphate starved condition (71) whereas defense responses were activated under phosphate-sufficient condition (72). *C. tofieldiae* was shown to invade the roots via the junctions between epidermal cells as well and form intercellular hyphae (71).

## 1.4    *Colletotrichum sublineolum* (Cs): variation, life cycle and host range

Cs shows enormous physiological and genetic variation. Variations in color, mycelial growth and conidia type and size is common (93, 94), and such variations including the rate of conidia production was observed in our lab. The extent of variation was explained as monoconidial isolates from a single lesion and from monoconidial cultures and sub-cultures were separated into different virulence (95) and field populations pathotypes that appeared to vary from year to year (96). Sexual reproduction enhances variability by facilitating genetic recombination. The mating type genes in Cs suggested the presence of both idiomorphs. However, no fertile strain has been identified (93). Genetic recombination without parameiosis was shown in Cs, and aneuploid and haploid recombinants were obtained from heterokaryons (97). Asexual reproduction is thought to be the major determinant of pathotypes (98). Reportedly, the species exhibits high genetic instability due to the spontaneous release of sectors during vegetative growth as observed in growth media  (93), and aneuploidy was found associated with the sectors (97).

Cs is known to occur in the mitosporic form (93). The conidia that is used as inoculum is produced in the fruiting bodies, and the specific structure that produces the conidia may determine the mechanism of dispersal. Conidia that are produced in the apices of setae may facilitate dispersal by wind whereas the conidia that come out from the conidiophore are dislodged and dispersed usually by rain droplets (99). Conidia production in the apices of setae was observed in many Ct species, and in Cs, an occasional production of conidia by the setae was reported (99). By contrast, , the production of creamy masses of spores among the setae under suitable environments was reported (100). The pathogen is thought to survive as mycelium, conidia and sclerotia in crop debris, seed and alternate hosts (101). The sclerotia of the root-infecting *Colletotrichum coccodes* was kept in soil for more than a year whereas the survival of the conidia was declined rapidly (102). No empirical evidence was found regarding the longevity of the survival structures in most Ct species.

Cs is known to cause anthracnose disease in a relatively narrow range of plant species. Cs was reported to infect most species in the genus *Sorghum* (15), although evidences are found only for the johnsongrass (*S. halepense*). Johnsongrass serves as an alternate host (16), and differences in virulence, host-tissue specialized races of Cs and defense responses by johnsongrass were reported (17, 18). A study on the pathogenicity of Cs isolates from sorghum and johnsongrass revealed sub-populations of the isolates with respect to the host species from which the pathogen

was collected (19). Beyond the genus *Sorghum,* Cs was causes a typical anthracnose symptom on the wild rice *Zizania aquatica* (20); has pathogenicity to the itchgrass *Rottboellia cochinchinensis* (13), and infects intact maize stalk (103). The genus *Eremochloa* was reported as a common host to Cs and *C. caudatum* (60). On the other hand, Cg was shown to effectively colonize the roots and the leaves of sorghum (103), and a moderately prevalent sorghum anthracnose was reported by the relatively distant *C. gloeosporioides* (91, 92).

Several reports indicated that Cs affects all the major organs of the sorghum plant (36, 104-106). These reports are based on old literatures before the two species (Cg versus Cs) were split, and that was found associated to the sorghum root, stalk, panicle and grain (107, 108). These old literatures referred to the pathogenic effect as anthracnose stalk rot (109), head blight and red stalk rot (100, 106, 110). Our observation of Cs on the stalk of sorghum lines that are susceptible to the leaf tissue suggest the possibility of important effects that needs well refined evaluation. The observation was done based on early indication of stalk rot via the inoculation of seedling roots (111) (Figure 1-1).
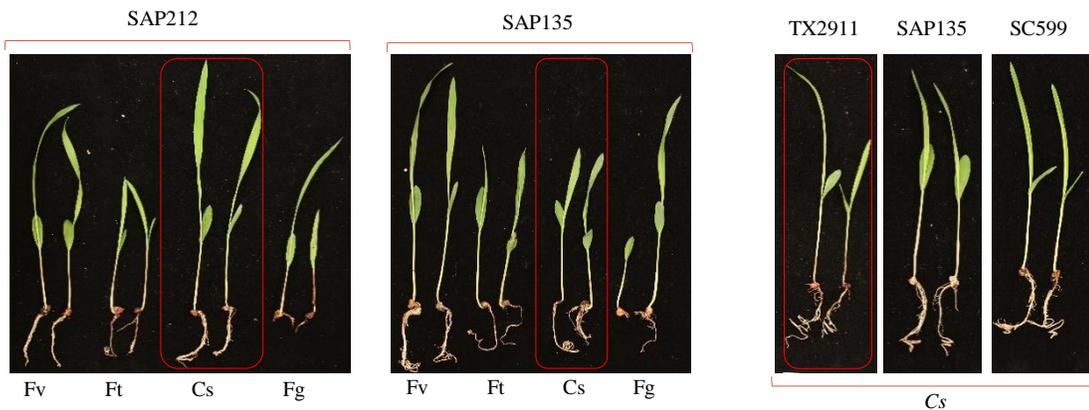


Figure 1-1. Early phenotypic effect of *Colletotrichum sublineolum* [Cs] as compared to other sorghum stalk rot causing pathogens observed after early-seedling root inoculation.
Sorghum lines that are reported to have resistance [SC599] and susceptibility [SAP25] to stalk rot were used. Sorghum anthracnose broad-spectrum resistant [SAP135 and SAP112] and susceptible [Tx2911] were also included. The evaluation was carried out as a preliminary observation about the association of CS, anthracnose, and stalk rot, *Fv = F. verticilloides*, Ft = *F. tapsinum*, Cs = *Colletotrichum sublineolum*, Fg = *F. graminicola*.

The sorghum foliar disease that is caused by Cs has been described extensively and commonly referred to as **S**orghum **A**nthracnose disease (SA). It is important to note that the sorghum stalk rot and panicle blight diseases were described in old literatures that used the species name Cg rather than Cs to the pathogenic species of Ct on sorghum. Thus, considering the reports on the effect of Cg (103) and *C. gloeosporioides* (91, 92) the association of Cs versus other species of Ct to most

tissue types and growth stages of the sorghum plant suggest multi-dimensional coevolution of sorghum – Ct interaction that might include different Ct lifestyle.

The scarcity of information about the effect of Cs on the sorghum plant parts other than the leaf may not imply the pathogen does not infect particularly on the stalk and panicle. Nevertheless, as a seed borne and soil borne pathogen, understanding the association of Cs to the different parts of the sorghum plant with respect to the routes to the pronounced foliar infection is imperative. Stalk rot preceded by leaf anthracnose was reported, although in some instances little foliar disease was evident (100). The maize anthracnose disease at seedling and anthesis mirrors SA at seedling stage and stalk rot at the later stage (112). Cg infects maize roots and produces runner hyphae, hyphopodia, and microsclerotia and inoculation of the root led to an asymptomatic systemic colonization of the plants (113). Similarly, the colonization of the rice root by *Magnaporthe grisae* leads to systemic invasion and the development of canonical disease symptoms on the aerial parts of the plant (82). The gene-for-gene type of host resistance to rice blast that is effective on the leaves of rice operates in roots (82), and such a specialized host-pathogen interaction may occur at any actively growing host plant tissue/organ.

## 1.5   Infection biology and symptom

Cs is generally considered a hemi-biotrophic pathogen (114, 115) though the infection related morphogenesis and the detail mode of nutrition are relatively poorly understood. However, the events leading to the sorghum-Cs interactions were characterized histologically. Inoculated spores germinate in nine hours, and at 18 hours, 100% of the germinated spores had produced mature dark brown, melanized and irregularly shaped appressoria (114, 116). In rice blast, the turgor pressure due to the accumulation of glycerol and melanization of the appressorial cell wall of *M. oryzae* aids to breach the plant cuticle (117). Inward deformation of the cuticle and localized staining of the host cells around the infection peg was followed by the invagination of the host plasma around infection vesicles and primary hyphae, and was oppressed tightly to the fungal cell wall, with no detectable matrix layer at the interface (115). In a resistant host, however, the fungal development was restricted to the single penetrated epidermal cells that was shown to collapse (114); encrusted with electron-opaque material and appeared dead (115). Both virulent and avirulent races penetrate the leaf epidermis and develop infectious vesicle, which may partially justify the biotrophic stage, but not the virulence of a pathogen strain. Cs races are virulent when

the host does not have an NLR protein that recognizes effectors to trigger HR. And the degree of virulence on sorghum genotypes seems to be a function of both the ability of a pathogen strain to rapidly colonize the host tissue and the quantitative resistance in the host.

The biotrophic phase lasts for about 24 hours, during which intracellular infection vesicles and primary hyphae colonize epidermal cells. Subsequent development of secondary hyphae is associated to the death of the host cell and marks the onset of a necrotrophic phase. In 90 hours, the secondary hyphae proliferates throughout the leaf sheath and accompanied by necrotic lesion, and fruiting bodies were observed in 114 hours (114) (Figure 1-2 a and b). Consistent results were shown in one, three and five days interval after inoculation and the number of appressoria, and primary and secondary hyphae were smaller and these fungal structures developed slowly in accordance with the level of host-resistance (116). Similarly, *C. higginsianum* has been known to establish a large biotrophic primary hyphae in the first infected cells. Shortly thereafter, a switch to necrotrophic growth occurs leading to the invasion of neighboring cells by secondary hyphae (78).

[a]    [b]    [c]
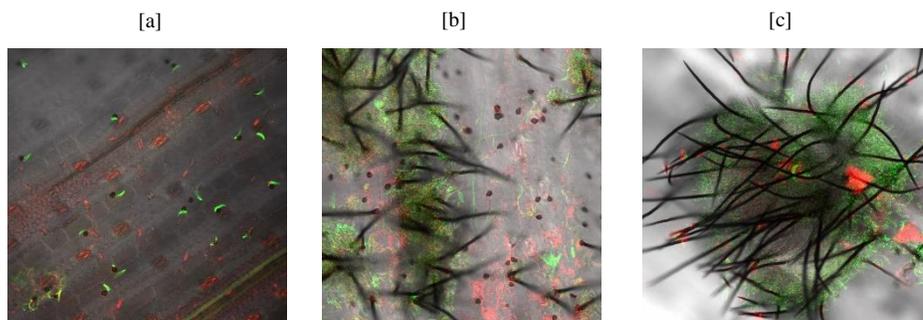


Figure 1-2. Fungal growth [Cs pathogen Csgl1] in the susceptible TAM428 cultivar in detached leaf disease assay.
The images show that the pathogen structures [a] at three days post inoculation with no acervuli [dpi], [b] Initial stages of acervuli at six dpi and [c] the well-developed acervuli at nine dpi. The image was taken under confocal microscope. The hyphae were stained green by WGA-AF488. Propidium iodide stained leaf tissue [red] shows the dead cells.

The SA disease phenotype varies with the pathogen strain and the host genotype. Typical symptoms are small, circular, elliptical or elongated spots usually about 5mm in diameter with small straw-colored centers and wide purple, red or tan margins (24). Many lesions may develop close together, coalesce and kill large portion of the leaf (24) (Figure 1-3). Midrib infection occurs as elongate-elliptical red or purple lesions, and yellowish lesion is also seen in some genotypes.

Moderately affected genotypes produced necrotic spots on the leaf lamina without noticeable acervuli (118). Defense response of resistant lines also shows diversity (119). High humidity is crucial to the successful colonization, and susceptible and resistant genotypes appear resistant under dry conditions (21).

[a]  [b]  [c]
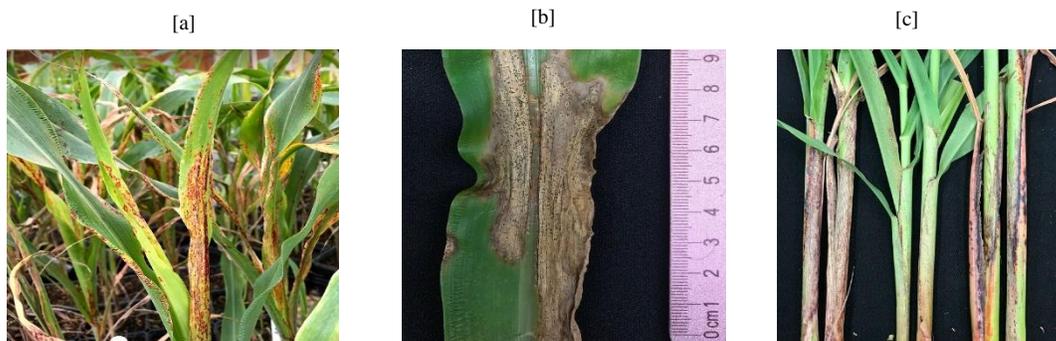


Figure 1-3. Sorghum anthracnose disease symptom.
[a] The disease phenotype on spray inoculated plants under greenhouse condition at 7 dpi. [b] The typical sorghum anthracnose disease [SA] phenotype on the cultivar Macia that is widely grown in Africa as an improved cultivar using the pathogen strain Cs-gl1 in Dr Mengiste Lab, Purdue University. Anthracnose disease phenotype on the leaf sheath in three lines: TAM428 and IS18760 are susceptible to Cs-gl1 whereas SC328C is resistant.

## 1.6   Genetics of plant immunity

Broadly, plant disease resistance has been categorized as qualitative, simply inherited trait or, quantitative, requiring many genes. Perhaps disease resistance is the only trait that the classical mendelian genetics has been widely and routinely applied to the improvement of field crop plants. Following Mendel's discovery of the laws of inheritance, Biffen established that disease resistance is genetically inherited as a Mendelian trait (120). Subsequently, the observations by Stakman on physiological races (121), and the gene for gene hypothesis by Flor (122) ushered an era of studies on the genetics of disease resistance in plants and breeding for disease resistance. Major gene resistance has become the realistic manifestation of Mendelian genetics, that was leveraged for crop improvement. The success of simply inherited disease resistance is apparent in predominantly biotrophic pathogens. Major gene resistance to the different rust species and rice blast disease have been enormously successful reducing disease epidemics with significant losses (123-125). Resistance to host specific necrotrophic pathogens also recorded similar success based on simply inherited resistance genes. These fungi produce host specific toxin which are targeted by host resistance genes as in the case of the maize *Hm1* gene against *C. carbonum* (25). The recent identification of few NLR genes that provide broad-spectrum resistance (126, 127) and non-NLR

major effect resistance gene (128) has elevated the optimism on qualitative genes to improve disease resistance in crop plants.

Knowledge on quantitative resistance (QR) has long been developing in parallel. QR is often durable and broadly effective against multiple pathogens. However, the contribution of individual genes to the resistance phenotype is not highly penetrant that slows down the effective use. Thus, seemingly quantitative disease resistance has been widely deployed as part of a broader breeding objective destined to improve agronomic fitness. The advent of molecular biology has enriched our understanding of genetics and host-pathogen interaction and expedited molecular breeding for various traits including host-resistance by tracking resistance QTL including QR loci (129). Some QR like slow rusting in wheat was bred successfully by stacking several genes (130). In recent years, molecular biology has further expanded the potential dimensions whereby host-resistance can be developed (131).

## 1.7    Immune pathways in plants

Host-pathogen interactions are intricate, and involve both the host and the pathogen genetic factors that modulate the outcome of the interactions (132-135). The nature of the initial host-pathogen interaction determines the battery of molecular and cellular responses such as global changes in transcriptome, regulatory networks, biochemical pathways and secondary metabolites that, ultimately, manifest as resistance or susceptibility (136-139). The nature, timing and magnitude of the plant responses determine the disease progression (139, 140). The environmental conditions at which the host and the pathogen grow, and the interaction occurs influences the readouts. The plant system has various non-specialized to highly specialized mechanisms that perceive and respond to the often-overlapping pathogen pressure and other stresses. However, under experimental systems this complexity is reduced to discernable mechanisms.

Most pathogenic organisms fail to infect most plant species (132, 141). Immunity to non-adapted pathogen is referred to as non-host resistance (NHR) (142), which literally the plant species does not host a pathogenic species, and defense response may not be manifested (143). Some non-host resistance genes are being discovered (144), though the mechanism behind the ubiquitous NHR has been largely mysterious (142). The response of crop plants to non-adapted pathogens would be of high value if its components are known but is not an obvious topic to a grower. For example, rust is of no concern to a tomato grower. However, some plant species host

pathogens that entails the importance of host resistance. In an ideal very wide natural habitat where plants and pathogens adapt, coevolution allows pathogenic races to infect some plant species whereas the plant populations effectively check pathogen outbreak. Intuitively, plant populations in such a habitat harbor a broad range of host resistance and susceptibility in time and space.

Plants have various mechanisms that serve immune functions (145). The passive physical barriers have characteristic features at the tissue level like the leaf epidermal and cuticular layers. As opposed to the vibrant adaptive immune system in animals, plants do not have specialized immunity cells, and the molecular basis of immunity is inbuilt in the individual cells (146). However, plants have a greatly expanded recognition repertoire which may compensate for the lack of an adaptive immune system (147, 148). Two fundamental molecular mechanisms of plant immunity have been thoroughly characterized at cellular level: pathogen associated molecular pattern (**P**AMP) **T**riggered **I**mmunity (PTI) (141) and **E**ffector **T**riggered **I**mmunity (ETI) (136). Immune signals are thought to profuse from the cellular level of pathogen recognition mediated defense response to the different levels of cellular organizations, and tissue and organismal systems (73, 149, 150).

The plant defense response triggered by some classes of the host cell-surface receptors define PTI, the first tier of plant immunity, resulting in quantitative resistance. Majority of these receptors recognize conserved molecular patterns of the pathogen when they come in contact, and hence known as pattern recognition receptors (PRR) (141). PRR are modular transmembrane proteins containing ligand-binding ectodomains, and they function as part of multi-protein complexes at the plasma membrane (151). NHR is likely to be a result of the PTI system (141) that may define the height of a robust PTI. Pathogenic species suppress the robust PTI to get access to the host tissue and gain nutrition.

PTI has been extensively characterized in the host-pathogen interactions. The discovery of the well-known receptor complexes of FLS2 and BAK1 that induce defense response up on recognition of the bacterial flagellin (152), and the FLS2-associated kinase BIK1 that phosphorylates the downstream signaling cascade to plant immunity are among the milestones in the study of the mechanisms to plant immunity (153). Damage associated molecular pattern (DAMP), a result of pathogen/pest damage on the host tissues, is a form of trigger to plant immunity with an outcome of quantitative resistance. Oligogalacturonides DAMPs from damaged plant cell wall elicit immunity in plants (154).

As opposed to many non-adapted pathogens, the typical adapted pathogens may clandestinely bypass host defenses and inject enormous small-secreted-proteins (effectors) into the host cell to modulate the host cellular machinery (133). The gene-for-gene arms race is a widely accepted coevolution of plant resistance and pathogen virulence (122). The recognition of effectors by Nucleotide-binding Leucine-rich-repeat (NLR) proteins trigger a heightened immune response, ETI, which defines the second-tier of plant immunity (155, 156). Thus, ETI confers immunity by guarding and sensitizing pathogen targets including receptors that signal defense responses (28). Host-pathogen protein-protein interaction studies have shown that the molecular recognition of effectors occurs predominantly  inside host cells (156, 157). Usually, a direct or indirect recognition of pathogen effectors by NLR receptors induces numerous defenses, including programmed cell death called hypersensitive response, and restricts pathogens to the infection site (156). Host-resistance to adapted pathogens have been extensively studied (6, 133), and a paradigm of molecular biology to host-pathogen interactions has been established based on studies in limited pathosystems (135, 158, 159).

Many NLR proteins that recognize pathogen effectors have been characterized in model phytopathosystems (159-162). Resistance to bacterial spec disease in tomato occurs when the Pto kinase in the plant physically interacts with *AvrPto* from *Pseudomonas syringae* and elicit a defense response (159). In *Arabidopsis*, the *P. syringae* AvrRPM1 induces defense by activating RPS2 whereas AvrRpm1 in the mutant *rpm1* plants suppress defense responses (162). The case of RPS2-mediates susceptibility in the absence of RPM1, which shows the rare recessive NLR inheritance. No *Avr* protein displays absolute specificity to a particular host-resistance protein, and the gene-for-gene concept by Flor at the point of molecular interaction between the host and the pathogen seems a complex "genes-for-genes" concept (163). More complex gene-for-gene interactions include *Avr* protein recognized by several NLR proteins, and NLR proteins recognizing several *Avr* proteins, one *Avr protein* suppressing recognition of another *Avr protein* by its corresponding NLR protein and two cooperating NLRs both necessary to recognize an *Avr* (133).

PTI and ETI activate both distinct and shared defense response pathways and reinforce the immunity of the plant to keep the pathogen at bay or reduce the pathogen pressure (141). In line with the "genes-for-genes" concept, though more intricate, the immune interactome provides a direct way to describe the molecular mechanisms of specificity and signaling (163). RIN4 interacts

with RPS2 to confers resistance to *P. syringae (164)*, and the interaction of RIN4 with RPM1 confers resistance to another strain of the pathogen *(165)*. Another NLR resistant protein RPS5 shares a third interactor protein NDR1 with both RPM1 and RPS2 (166). Shared interactors may also play opposing role as in the membrane-anchored receptor-like kinase BIK1. BIK1 enhances resistance to a virulent strain of *P. syringae* pv *tomato* whereas its interaction with necrotrophic fungal pathogen causes severe susceptibility (158).

Distinct and shared features of plant immunity can also be viewed with respect to host-pathogen interaction. Pathogens target the highly conserved pattern recognition receptors (PRRs) and the interactors of NLRs and PRRs to effectively modulate the host system and cause disease. The PRR FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB (28). A conserved fungal effector NIS1 suppresses PTI by targeting the well-known plant immune kinases like BAK1 and BIK1, interfering with their essential functions for immune activation upon pathogen recognition (167). The effector AvrRpt2 from *P. syringae* targets RIN4 in *Arabidopsis* and modulates RPS2-mediated resistance to lethality (137). The initiation of RPS2-specified disease resistance is coupled to the AvrRpt2-directed elimination of RIN4 (164). As such, RIN4 is a point of convergence for the activity of at least three unrelated *P. syringae* type III effectors AvrRpm1, AvrB and AvrRpt2 (137).

## 1.8    Sorghum anthracnose resistance

### 1.8.1    Genetics of sorghum anthracnose resistance

Sorghum germplasm harbors resistance to Cs infection. Records show germplasm selection to sorghum anthracnose resistance occurred since 1930's (168) and the determination of inheritance of the trait since 1950's (14). However, studies on characterization of host-resistance and the interaction with the pathogen was spurred after several decades with three developments: the reclassification of Cs from *C. graminicola* (13), the screening method for sorghum anthracnose (169) and the PCR-based molecular markers (170). Since then, the inheritance and molecular mapping of sorghum anthracnose resistance loci has been conducted extensively for more than two decades. Nevertheless, molecular studies of anthracnose resistance in sorghum lags far behind the other prominent cereal crops. Breeding for a stable host resistance to Cs has been difficult due to the hypervariable nature of the pathogen and the incomplete understanding of the host-pathogen

interactions (171, 172). In, 2008, evaluation of 30 commercial hybrids in which six were highly resistant to anthracnose disease (173) was documented. However, we did not find a considerable peer-reviewed research articles regarding the deployment of sorghum anthracnose resistance. Proper analysis and documentation of some selected regional crop variety trials for economic traits like disease resistance is necessary to inform the flow of, and to gain feedback from, the utilization of the knowledge and technologies generated in the basic molecular biology research.

The enormous sorghum germplasm that has been evaluated for anthracnose disease phenotype has revealed huge variability for resistance to the pathogen (174-177), and the sources of SA resistance showed association to the weather pattern in the countries where the sorghum variants originate (176). The variation has also been evaluated extensively in specific agroecologies worldwide. The variation of anthracnose resistance in the temperate-adapted sorghum germplasm has been limited (178), and germplasm enrichment and conservation need has been carried out using core germplasm resources especially from the primary diversity in the East African region (179, 180).

Many sorghum lines have been used for genetic studies of anthracnose resistance that laid some basics of anthracnose resistance in sorghum. Studies on quantitative resistance in crop plants have focused on the underlying genetic architecture rather than to identity mechanisms of individual genes (181). Quantitative resistance to SA is partly manifested by the widely reported defense chemicals including phytoalexins (182). A half-diallel mating analysis of nine anthracnose resistant sorghum lines revealed significant general and specific combining ability effects for anthracnose disease phenotype indicating the role of both additive and non-additive gene actions (183). The loci that are responsible for major QTL effects of a quantitatively inherited trait can only be captured using genome-wide analysis of mapping populations that are not subject to systematic selection. Several genome-wide studies that are summarized in this review have reflected on the quantitative inheritance of anthracnose resistance; including one extensive and well refined QTL map based on a biparental mapping population (184) whereas qualitative inheritance has been widely studied since 1950 (14). Relying on highly penetrant disease phenotypes and simplified evaluation/clustering of a wide range of disease phenotypes and levels of disease severity may have shadowed on the depth of knowledge about quantitative inheritance to sorghum anthracnose.

The majority of genetic studies on sorghum anthracnose resistance revealed dominant inheritance (171) and reports of recessive resistance has been relatively very small (171, 185). A recessive resistance to a specific pathotype was reported to breakdown in a location dependent manner (36) which, later, was reported to be race specific as well (186). The collection of race-specific anthracnose resistant lines of sorghum, known as differentials, have been employed to differentiate Cs races (36, 186). Cs is well described as a foliar pathogen, and foliar resistance by a single dominant gene and midrib resistance by another unlinked recessive gene was reported based on an F2 population of a biparental cross (187). Similar host-tissue specialization of Cs races was shown between the leaf blade and midrib tissues using sorghum and johnsongrass cultivars (17). A study showed 87% correlation between seedling resistance in the greenhouse and adult plant resistance in the field (188). Deterioration of anthracnose resistance of a sorghum line with aging was reported (189) whereas adult plant resistance, which is well known in maize, wheat and rice diseases has not been documented in sorghum anthracnose.

QTL analyses of sorghum anthracnose resistance has been conducted in different experimental setups: from core germplasm collection to biparental mapping population; from single pathogen race to diverse pathotypes to confounding disease pressure; from detached leaf disease assay to evaluation under greenhouse to different environments under field conditions. A preliminary genetic map to the recessive resistance in SC326-6 was developed long ago using the old PCR-based molecular markers (190), and RAPD and SCAR markers were available for marker assisted selection (190, 191). Nevertheless, we did not find an improved mapping report on the recessive resistance in SC326-6 using the up-to-date marker system.

The sorghum line SC748-5 has robust anthracnose resistance to a broad-spectrum of pathogenic races and environments (171, 186) and the genetics of this resistance has been studied for two decades. A decade ago, AFLP markers were used to develop a genetic map of the resistance loci in SC748-5 that includes a major effect QTL at the distal end of chromosome 5 and SSR markers were designed to integrate the resistance in sorghum improvement programs (192). Subsequent mapping effort based on SNP in a biparental mapping population suggested the genetic architecture for anthracnose resistance in SC748-5 was not under the control of one gene but, more likely, a linkage block containing several resistance genes at a single locus (15). The mapping effort has progressed further to the identification of a major effect non-NLR resistance locus in a well refined genome-wide map of hundreds of markers that were reported to be within or linked

to the relatively comprehensive quantitative inheritance of the anthracnose resistance in SC748-5 with respect to the susceptible background BTX623 genome (184).

Broad-spectrum resistance QTL to sorghum anthracnose resistance was also mapped in the line SC112-14 using F2 genotypes and recombinant inbred lines of a biparental mapping population (193). The locus was reported to be distinct, based on comparative analysis, from the resistance locus in SC748-5. Recently, the QTL in SC112-14 was fine-mapped to a 34 Kb genomic interval. This 34 Kb resistance locus harbors several predicted genes containing domains of known functions: ankyrin repeats, F-box, PAP-fibrillin and WD domain, G-beta repeats and domains of unknown function (194). Similarly, three loci that confer stable resistance across environments were reported in the lines SC155-14E and SC414-12E (172). Minor effect and environment specific loci were also reported along with the major loci in all the sorghum anthracnose broad-spectrum resistant lines. The genetic resource base in the sorghum conversion project by USDA included those typical anthracnose resistant lines that originated in East Africa (171).

Several other SA resistance loci have been mapped. Eight loci were defined using association mapping in 242 diverse landrace worldwide mini-core collection, of which seven loci harbor resistance related genes; two of them were NLR genes (195). Another QTL mapping revealed two loci, both loci carry defense related genes (196). QTL mapping of resistance to several sorghum foliar diseases was also conducted in a unit of experimental setup. The resulting QTL was associated with plant color across years and locations. revealed a colocalized QTL associated to plant color (197). Mapping resistance loci based on the confounding disease pressure from multiple races of a pathogen and multiple pathogens may facilitate the mapping of genes that better resonate with the crop improvement needs in sorghum disease hotspot agro-ecologies.

Intraspecific and interspecific comparative studies have also contributed to SA resistance. Two sorghum lines with contrasting SA phenotypes were examined for differences in gene expression profile (198). Down regulation of five candidate genes through VIGS clearly compromised the resistance response. Two of these resistance genes were NLR genes that were deployed in regional sorghum improvement programs. The remaining three genes were a lipid transfer protein, a zinc finger-like transcription factor and a cell death related protein. On the other hand, the johnsongrass-Cs pathosystem has become a valuable comparative platform in the characterization of SA resistance (17, 18, 199) in addition to the potential source of resistance. In

sum, despite extensive genetic studies, there has not been a sorghum gene that was mapped completely and validated for anthracnose resistance.

### 1.8.2   Genomic of sorghum anthracnose resistance

Recent genomic studies in crop plants revealed huge structural variation as an important source of plant genome evolution, maintenance of adaptive traits and disease phenotype (200-203). The 3000 rice pan genome analysis identified 63 million individual structural variant calls that were grouped into 1.5 million allelic variants and the regions of high structural variant frequency were enriched in stress response genes (201). Genes affected by structural variation in *Populus* species showed lower-than-average expression level and higher level of *dN/dS*, suggesting that they are subject to relaxed selection pressure or correspond to pseudogenes (203). Copy number of NLR genes were greatly increased by lineage-specific retroduplication in capsicum species that has played key roles for the massive emergence of NLR genes (202) and gene that were affected by copy number variation showed enrichment in categories related to resistance to stress and pathogens *Populus* species (203). Pangenome analysis in *Brassica napus* showed the absence of 368 NLR gene analogs in the reference genome (204). Genomic analysis of diverse sorghum genotypes revealed approximately 1.9 million indels (205), however, its relevance to disease resistance has not been studied.

The nucleotide binding site leucine rich repeat (NLR) class of genes is one of the most abundant and diverse class of genes in plants (206), and the true extent of intraspecific diversity has been unclear (207). A report in 2016 on the genomic analysis of grasses revealed 88 full-length NLR genes in sorghum, 107 in rice, 24 in maize, and 44 in *Brachypodium* (208). A couple of years earlier, a report that focused on the plasticity of sorghum nucleotide binding site (NBS) resistance genes showed 228 NLR genes in sorghum genome, 37 in maize and 161 in rice (209). In this report, non-LRR NBS genes were included to a total of 346 NBS genes in *S. bicolor* though no disease resistance function has been attached to the non-LRR group (209). A recent pan genome analysis of *Arabidopsis* NLR estimated about 464 NLR ortho-groups with a range of 167 to 251 NLR genes per accession (210). The identity and functional variability of the NLR proteins have been attributed to the LRR domain that is known to detect pathogen molecules thereby trigger the NBS domain to elicit signal for the entire known downstream resistance responses (211).

In the past decade, only few sorghum-focused studies have documented regarding the genomic features of NLR genes (208, 209, 212). Majority of the sorghum NLR genes are distributed around the chromosomal ends (209), that is different from the random chromosomal distribution in rice and *Arabidopsis* (212). A report on the distribution of NLR genes in the Asian cultivated rice chromosome 11 as compared to its wild relatives showed similar pattern (213), and purifying selection has been suggested to account for the aggregation of NLR genes to the chromosomal ends (209, 213). A vast majority of the sorghum NLR genes occur in gene clusters mainly due to local gene duplication, and the number of NLR genes in sorghum, rice, maize and *Brachypodium* is positively correlated with the gene clustering rate (208, 212, 214). The sorghum NLR genes have basal expression levels and show high potential for being miRNA target (208). The proportion of the complete open reading frame (ORF) in the 346 sorghum NBS genes was estimated in the range of 2.17% to 86.95% with the wild and weedy genotypes having higher frequencies of ORF (209).

Studies show majority of NLR genes confer race specific resistance (215). However, NLR heterodimers in which both parties are required to confer disease resistance (211) have been described. Accordingly, a broad-spectrum resistance QTL that is associated to a cluster of NLR genes may likely be conditioned by multiple NLR proteins from the same gene cluster rather than the solitary NLR proteins. Understanding the nature of NLR protein monomers, homodimers and heterodimers encoded by the cluster of NLR genes may significantly shape our understanding of disease resistance and the future of disease resistance deployment. However, the linkage block of resistance genes as suggested for the broad-spectrum resistance in SC748-5 may have distinct functional feature from the NLR gene clusters (15). The resistance by a non-NLR major effect QTL that provides resistance across environments and diverse races of the pathogen.

### 1.8.3 Host-resistance in related pathosystems

Rice blast caused by *Magnaporthe oryzae* has been regarded a model pathosystem for understanding host-pathogen interaction in crop plants (161, 216) and the genetics and molecular mechanisms of resistance to the disease has been relatively well characterized (160, 161). The shared features of the sorghum – Cs pathosystem to the rice blast further suggest some of the observations may be leveraged to expedite our understanding of SA disease and the resistance mechanism to SA. Both pathogens initiate their first phase of infection in the biotrophic lifestyle

to establish infection structures and then shift to the aggressive necrotrophic phase. In our laboratory, a recently identified sorghum anthracnose resistance gene and other evidences show up to near a complete homology to rice blast resistant and putative resistance genes in the cultivated and wild rice (195). The symptomatic infection of wild rice with Cs (20) may also imply relatedness at the overall pathosystem level. Owing to these multiple layers of similarities, some discoveries in related pathosystems, mainly rice blast, is included in this review.

About 500 QTLs (217) and over 100 genes to rice blast resistance have been mapped, and 35 of those genes have been cloned (215). Similarly, the interaction of wheat with the biotrophic fungus *P. graminis*, has resulted in the identification of over 50 catalogued resistant genes in wheat and its wild relative and yellow rust has also been studied extensively (123, 124). Most of the identified genes belong to the NLR gene family and confer race specific resistance. Strong broad-spectrum rice blast resistance has been developed by stacking several race specific resistance genes (125). Pathogen exposure to resistance genes is a major factor to govern co-evolution in rice - *M. oryzae* pathosystem (218). More recently, however, mapping efforts have unraveled several broad-spectrum rice blast resistance NLR genes (219) as opposed to the tenacious gene-for-gene resistance. Some of these broad-spectrum NLR genes have been introduced into elite cultivars and demonstrated the enormous contribution to rice production.

On the other hand, the non-NLR rice blast resistance genes proline-rich *Pi21* (126) and B-lectin receptor-like kinase *Pi-d2* (127) exemplify durable resistance. Similarly, the durable ABC transporter wheat rust resistant gene *Lr34* has long been effective against all the specialized wheat rust pathogen; alleles of Lr34 that confer disease resistance and susceptibility differ by three genetic polymorphisms (128). Interestingly, *Lr34* has increased resistance to rice blast, sorghum rust and sorghum anthracnose (220, 221). Together with the stable non-NLR anthracnose resistant loci that were identified in sorghum, the dependency on a limited sources for durable resistance (193) may be reduced. Likewise, single non-NLR genes may be responsible to the broad-spectrum anthracnose resistances in sorghum lines SC748-5 (184) and SC112-14 (194) and become invaluable additions to the wide effort against disease of the major cereals.

### 1.8.4 Response of sorghum to *Colletotrichum sublineolum*

Preliminary molecular studies of host-resistance to SA have been reported specifically on the kinetics of defense response genes. In a comparative study of cDNA clones using an anthracnose resistant sorghum line, the non-host pathogen *C. heterostrophus* caused a rapid accumulation of PR-10 and chalcone synthase transcripts after the appressoria had become mature whereas the transcript accumulation in response to *Colletotrichum sublineolum* (Cs) was delayed until the host tissue was penetrated and infection vesicles were formed (140). Differential gene expression analysis of Cs-inoculated and non-inoculated tissue on SA resistant line DK18 revealed forty pathogen-inducible genes that are functionally related to cell rescue and defense, signal transduction, abiotic stress, secondary metabolism and protein degradation. In a follow up RT-PCR analysis, all the forty genes were induced by Cs and two non-host pathogens (*Cochliobolus heterostrophus* and *Magnaporthe oryzae*) and nineteen genes showed differential expression between in the host and the non-host cases (222). In the same report, global mRNA was examined in the resistant line SC748-5 and the susceptible line BTx623 was examined after inoculation with Cs. Genes encoding an ABA-responsive protein, a leucine-rich repeat protein, a flavonoid 3′-hydroxylase and a glutathione S-transferase were induced in the resistant line, but their expression was hardly detected in the susceptible line. Earlier induction of chalcone synthase and PR-10 in a resistant cultivar as compared to a susceptible cultivar was also reported (182).

Defense response to infection by Cs was examined in susceptible and resistant lines. Papilla formation is associated to a higher number of conidia and appressoria consistent with the level of resistance. However, the proportion of infection with papilla was higher in the susceptible interaction (116). A similar pattern of cell wall cross-linking and accumulation of callose, 3-deoxanthocyanidin and phenol was also observed (116). The accumulation of callose was also described in the durable non-host resistance to Ct species (223). In sorghum, the accumulation of the phytoalexin 3-deoxanthocyanidin at the site of infection have been linked to quantitative resistance to Cs (182, 224, 225). Studies showed greater and faster accumulation of phytoalexins accompanying the earlier induction of defense-related genes in the resistance than the susceptible sorghum-Cs interaction. Glucan-containing polysaccharide isolates were shown to initiate browning and phytoalexin production (226).

Interestingly, the readout from the sorghum-Cs interaction seems to rely on the temporal regulation of defense response Genes. The kinetics of transcript accumulation of defense response Genes to SA seems rapid in a non-host interaction, a little delayed in host-resistance, and delayed more in the susceptible interaction (80, 114, 140). Early molecular interactions are key to the outcome of infection (141) whereas the relatively extended delay activation of defense response Genes may give a pathogen an advance to rule out the ability of intact cell and tissue systems to defend themselves.

Sorghum-Cs interaction showed a functional metabolic web of phenylpropanoid and flavonoid pathways and a higher amount of apigeninidin luteolinidin and 5-methoxyluteolinidin in the resistance (139, 182). The accumulation of $H_2O_2$ tends to be higher in the resistant sorghum-Cs interactions in the first three days after inoculation whereas massive amount of $H_2O_2$ is generated later in the compatible interaction (116). In lupin anthracnose, the concentration of isoflavones free aglycones was increased during infection, although the increment was dependent on the leaf age and the infection method (227).

The role of some defense response genes has been characterized more in the model pathoysyetems and some other crop plants as compared to sorghum. These studies are becoming more -*omics* technology based that often provides clue to future studies. Abscisic acid was shown to be a substrate of the ABC transporter in the LR34 mediated wheat rust resistance (228) that confers partial resistance to sorghum and rice diseases as well. The epigenetic regulation of rice NLR heterodimer that disabled the resistance to rice blast disease in the reproductive organs has been remarkable (229). The mechanism of NLR mediated rice blast resistance that does not break when genes of the interacting WRKY proteins were knocked-down has not been determined (230).

## 1.9    Genic versus non-genic control of anthracnose

Though the control of anthracnose is almost exclusively through the use of host resistance, breeding for stable resistance has been difficult (171). In a related issue for instance, Cg had long been not a problem in maize crop production in most developed areas due to the deployment of resistant cultivars (60, 231). However, recently maize anthracnose has become a major problem in US as some varieties of engineered maize seem more susceptible to infection and the problem was reported in Portugal and Switzerland as well (54).

In Ct species, the highly variable strains, host range and endophytic nature, the survival in plant debris and seeds, and the continuous supply of inoculum through wind and windblown rain may undermine host-resistance as a sole means to control anthracnose disease for any crop plant at any production condition or scale of crop production. Research and available technologies to non-genetic control of anthracnose disease includes inorganic and organic fungicides, biocontrol agents including endophytic Ct species, and nanoparticles (232-234). Some of the control methods may prove effective as quarantine and postharvest treatment (235, 236). The studies may suggest a judicious use of non-genic options may play key role in the management of anthracnose disease.

## 1.10 Rationale and objectives of this research

The review is developed as a background to the thesis to highlight the history, the status and the prospects of host-resistance to SA. Studies on the inheritance of host-resistance, the pathogenic features of Cs and their interactions were reviewed, and the major knowledge gaps and areas of intervention were suggested. One of the obvious deficiencies that has emerged from reviewing the literature was the lack of identified SA resistance genes which may have spurred both basic research and improvement of the crop for disease resistance. This thesis research was initiated as a direct contribution to this key research gap.

Sorghum germplasm harbors enormous genetic variation for resistance to pathogens specially anthracnose. However, genetic studies in host resistance and identification of resistance genes and their mechanism of functions have been limited. Sorghum lines were previously selected for variable sources of resistance genes (96, 186). The lines originate from different parts of the world and have been used to profile races of Cs and monitor pathotype dynamics. Thus, the lines differentiate pathogen races and known as differential lines. However, the resistance loci and matching effector genes have not been identified. By screening sorghum natural variants and differential cultivars, the aim of this research was to identify specific SA resistance genes and pave the way for future mechanistic studies and agricultural applications. Specifically, the aims were to screen these natural variants for resistance to Cs strains from US and Ethiopia and determine the inheritance of SA resistance in some of the lines, identify the resistance genes and conduct functional analyses.

# CHAPTER 2.    AN NB-LRR RECEPTOR GENE CONFERS RACE-SPECIFIC ANTHRACNOSE RESISTANCE IN SORGHUM

**Abstract**

Anthracnose disease caused by the fungal pathogen *Colletotrichum sublineolum* (Cs) is a widespread and economically important disease of sorghum. Genetic resistance is the only economically feasible disease control strategy in most sorghum producing regions of the world. However, the genetic control and molecular mechanisms of resistance to this fungus are poorly understood slowing the deployment of genetic resistance. Here we describe the *ANTHRACNOSE RESISTANCE GENE 2 (ARG2),* encoding a canonical a nucleotide-binding leucine-rich repeat (NLR) protein, which confers complete resistance to strains of the Cs fungus. The sorghum line SC328C was resistant to isolates of the fungus in a genetic screen of a collection of sorghum natural variants. Bulk-segregant analysis of biparental mapping population using whole-genome DNA sequence located the *ARG2* locus to the near-proximal end of chromosome 5. Subsequent genetic mapping narrowed the *ARG2* locus to a region harboring a cluster of three predicted *NLR* genes. Comparative genomic analysis coupled with gene expression and DNA sequence analyses of the three *NLR* genes resulted in the identification of specific sequence polymorphism and the *ARG2* gene underlying resistance in SC328C. The genetic link between sequence polymorphism in the *ARG2* gene and variation in the disease phenotype was established by identifying genotypes that carry independent mutant allele. Further, the expression of defense response genes flavon-O methyl transferase and chalcone synthase 08 were higher in resistant near isogenic lines (NILs) than the susceptible NILs at the early stages of infection. Interestingly, resistant NILs produced higher shoot biomass than the susceptible NILs in the absence of pathogen pressure, suggesting the absence of growth trade-off associated with resistance. *ARG2* mediated resistance is stable at a relatively high temperature. In sum, we present genetic evidence for the identification of new anthracnose disease resistance gene which will be useful for crop improvement solely or stacked with other resistance genes.

## 2.1    Introduction

Sorghum (*Sorghum bicolor)* is major cereal crop of global importance. It is a source of food, feed and biofuel. The productivity of the crop has long been limited by diseases (1), of which, anthracnose caused by *Colletotrichum sublineolum* (Cs) is economically important in many sorghum growing regions (2, 3). This fungus attacks the aerial part of the plant (4), but the most severe effects arise from foliar infection (5). Humidity and high rainfall promote infection (6) resulting in reduced grain size (7) and greater than 50% loss of grain and forage yield (7).

A continuous supply of inoculum through wind and windblown rain renders cultural management of sorghum anthracnose disease (SA) ineffective (8). The pathogen is known to occur in the mitosporic form (9) and survives as mycelium, conidia and sclerotia in crop debris, seed and alternate hosts (10). Biological (11, 12), chemical (13) and integrated (14) management of the disease have been reported. However, the access to some of these technologies is limited and costs are not feasible in many sorghum producing communities (15). Genetic resistance has been widely considered to be efficient in the management of the disease (16), and expected to be instrumental especially in farming societies where the capacity and access to other disease management options are limited. Sorghum variants that have resistance to different races of the anthracnose pathogen have been identified extensively (17, 18).

Broadly, SA resistance involves many genes that contribute in a quantitative or qualitative manner (19, 20) similar to other pathosystems. Studies on the host-resistance to SA frequently show major gene effect that inherits as dominant allele (16). Qualitative resistance is simply inherited (16) and confers often race-specific resistance (18). The high penetrance of these qualitatively inherited phenotypes makes breeding for qualitative resistance easier. The major drawback of this form of resistance is that it is often overcome by new races or is effective in limited region. Usually *NLR* genes of crop plants mediate qualitative resistance (21, 22), and in sorghum 97% of this class of genes were estimated to occur in cluster (23). Accordingly, broad-spectrum resistance to SA may be often controlled by a linkage block of *NLR* genes rather than a single gene (24). By contrast, quantitative resistance is known to be durable and broad-spectrum but is only partial, allowing for a threshold of pathogen growth and resulting in yield under diseases (25). Many loci that are associated to quantitative resistance to SA has been reported (26). However, studies on quantitative resistance in crop plants have focused on the underlying genetic architecture rather than the identity and mechanisms of individual genes (20). Genetic studies of

SA have been restricted to the identification of QTLs and some candidate resistance genes (19, 27-31), and we did not find a report on the complete identification of anthracnose resistance gene in sorghum despite the genetic diversity in host resistance and pathogen virulence.

The interaction between the rice blast fungus *Magnoportha oryzae,* and rice (*Oryza sativa*) seems to mirror the sorghum interaction with Cs. In both cases, the infection starts with a biotrophic phase followed by an aggressive necrotrophic phase. The genetics and host response mechanisms have been relatively well studied in the rice blast pathosystem. To date more than 30 rice resistance genes and 12 *M. oryzae* fungal effectors have been identified (21). Most of these rice genes encode *NLR* that confer race-specific resistance. Similarly, over 50 resistance genes to wheat stem rust caused by the obligate biotrophic fungus *Puccinia graminis* have been catalogued (22), and yellow rust resistance has been studied extensively (32). In addition, genomic tools in wheat for the transfer of genes from progenitor species have been developed (33).

In this genomics era, the identification resistance genes have catalyzed the characterization of resistance mechanisms, which in turn has opened avenues for improved and novel approaches to control of diseases (34). The molecular mechanisms of disease resistance in crop plants have been studied extensively. In sorghum, the accumulation of phytoalexin 3-deoxanthocyanidin at the site of infection have been linked to quantitative resistance to Cs (35-37). The transcript accumulation of defense response genes and the activation of some physiological processes have also been associated to SA resistance (8, 38, 39). However, our understanding of molecular mechanisms of SA resistance have been incomplete which undermines the potential contribution to breeding for improved host-resistance (40).

This study was initiated to identify and characterize sorghum anthracnose resistance genes and associated host-responses. We report the identification of *ARG2* underlying complete resistance to Cs strains Cs-gl1 and Cs-grg in sorghum genotypes SC328C and SbRio that encodes a canonical NLR protein. Near isogenic lines (NILs) that differ in the *ARG2* locus display the typical resistance and susceptibility disease phenotypes like the parental lines that were used to map *ARG2*. The NILs were used to characterize defense-responses that are mediated by *ARG2*, and impact on plant growth in the absence of the pathogen.

## 2.2 Materials and Methods

### 2.2.1 Plant materials and fungal culture

Sorghum plants were screened for altered responses to five strains of Cs (Table 2-1). The Cs strains Cs-gl1 and Cs-gl2 were originally collected from Indiana (41), Cs-grg was a recent collection from Tiffon - Georgia, USA, and Cs-29 and Cs-27 were from Western Ethiopia. The Cs strains were cultured on potato dextrose agar (24 gm PDA and 15 gm bacto-agar per litter of water). Fungus inoculated petri-plates were kept on lab benches at 21-24$^{o}$c under continuous florescent light that was a meter above the benches. Four of the five strains produce large amount of conidia in two to three weeks of culture. To prepare the inoculum, water was poured on the plates and the conidia was gently scrapped off the media surface and passed through multiple layers of cheesecloth. The inoculum concentration was adjusted to $1 \times 10^6$ conidia per ml, and 2 ml Tween20 per litter of the conidial suspension was added and mixed by gentle shaking for immediate inoculation.

### 2.2.2 Disease assays and evaluation of host response

Sorghum plants were raised in $10 \times 10 \times 10 cm^3$ pots under greenhouse condition and inoculated when the plants develop three to four fully expanded leaves, usually about 30 days old plants. The plants were arranged in trays and sprayed with the inoculum until all the leaves become uniformly wet. The inoculated plants were placed in humidity chambers and incubated at 28-33$^{o}$c and >85% relative humidity for 48 hours. Then, the plants were transferred to greenhouse beds and kept with an overhead misting at 30 minutes interval during the daytime until plant response or disease symptoms become visible. The same range of temperature was maintained in the greenhouse too, and supplemental overhead light was provided for 14 hours per day.

Detached-leaf disease assay was carried out in parallel with the spray inoculation. Fully expanded leaves were detached from plants and placed on wet cheesecloth in plastic trays. Wet cotton balls were placed at the points where the leaves were cut. Twenty µl of the inoculum was drop inoculated at several spots of the flat leaf surfaces such that the droplets stay at the spots. The trays were covered with transparent domes and sealed with tape. The detached-leaf disease assay was carried out on lab benches like the fungal culture. Both spray and detached disease assays

were closely monitored for the best time to evaluate the disease phenotype, which was usually seven to ten days post inoculation (dpi).

The disease phenotype was recorded as clear-cut resistant or susceptible. A genotype was considered resistance when disease symptom was not shown, but anthocyanin pigmentation and highly restricted HR-like lesions appeared in a small proportion of the spray inoculated leaves. In the detached leaf disease assay, however, resistance was characterized by browning of the entire spots of the intact leaf tissue where the inoculum droplets were placed. Inoculated leaves were stained with trypan-blue and wheat germ agglutinin – alexa flour WGA-AF488 to visualize fungal growth on the leaf tissue. Sorghum genotypes that showed the sorghum anthracnose disease symptom and fungal growth were considered susceptible. The current study focused on the resistance in SC328C.

### 2.2.3   DNA extraction and sequencing

Biparental mapping populations were developed by crossing SC328C with TAM428, and bulk-segregant analysis was carried out at F2 using genomic sequence raw-reads. The DNA samples were prepared from 50 resistant and 50 susceptible F2 progenies, using DNeasy plant mini kit. Gel electrophoresis and nanodrop were used to check the quality and concentration of DNA samples. The DNA samples from individual F2 plants that showed single high-molecular-weight band on the gel were used to constitute two pooled-progeny samples, one for the susceptible group and the other for the resistant group. Equal amounts of the individual DNA samples, from the respective 50 progenies, were used to constitute the pooled-progeny DNA samples. The pooled samples were sequenced at 25x coverage using illumina highseq2000. Genomic DNA sample from the resistant parent was also sequenced in parallel, and genomic raw-reads of the susceptible parent was available from previous work.

### 2.2.4   BSA-Seq and recombination analysis

Bulk-segregant analysis using genomic DNA raw-reads (BSA-Seq) was carried out using QTL-seq analytical pipeline to identify the genomic region that confers resistance in SC328C. The pipeline inputs the pooled-genomes raw-reads obtained based on the distinct disease phenotype and yields the corresponding distinct genomic regions that are significantly associated to the

differences in the disease phenotype (42). All default statistical parameters as specified in the pipeline were used. The chromosomal DNA sequences in the *S. bicolor* reference genomes served as a platform to lineup the corresponding sequences from the raw-reads of the pooled samples. The basis of the analysis is the difference between the average frequency of SNPs in the pooled samples (ΔSNP-index) that are inherited from the two parental lines. The random allelic frequency in the F2 pooled-progeny samples is expected to be significantly violated only around the genomic regions that enabled the separation of the progenies based on their disease phenotype. Thus, ΔSNP-indices that result from the random assortment of alleles approximate to zero whereas ΔSNP-indices elevate in the genomic regions that are associated to the disease phenotype and reveal as peak.

The statistics of ΔSNP-index provides a better picture about the potential genomic interval of *ARG2* locus around the peak as revealed through the BSA-Seq. The SNP-indices were estimated as the average frequencies of all SNPs that were identified in two Mb and four Mb sliding windows. The windows slid 50 Kb between successive SNP-index estimates. Thus, the changes in adjacent ΔSNP-indices arise from both the dropped 50 Kb sequence of the preceding window and the subsequently added 50 Kb sequence, depending on the density and distribution of SNPs. Most close-by SNP-indices were estimated from highly overlapping SNP frequencies. This nature of the BSA-Seq statistics leaves imprecise margin to the candidate genomic region around the peak, and an attempt to focus on a smaller candidate genomic region, based on the highest estimates of ΔSNP-index, may mislead be misleading.

DNA size-markers were designed for recombination analysis to define a narrow genomic interval of *ARG2* locus from the BSA-Seq peak genomic region. The BAM files of the parental and the pooled-progeny genome were aligned and examined using the integrated genome viewer (IGV) software to identify insertion-deletion (indel) sites for PCR primers. Accordingly, many DNA markers that contain tens to a few hundred base-pair size-differences between the parental genomes were designed. Intuitively, the frequency of mismatch between the disease phenotypes and the parental origin of a marker-type is expected to increase with distance from *ARG2* locus. Thus, to generate a strong gradient of allelic frequencies, markers were designed beyond the peak genomic region at which the estimates of ΔSNP-indices dropped off to a random recombination. Accordingly, an elaborated bi-directional gradient of allelic frequency was expected to confirm the genomic region where the results of BSA-Seq and recombination analysis overlaps.

In this study, no primer designing tool was used to identify the primer annealing sites. Once a potential-indel site was observed on IGV, preferred primer annealing sites were picked manually to avoid SNPs between the parental genomes that may create a difference in the PCR efficiency due to the parental origin of a template. The manual method enabled the design of optimal indel-to-amplicon size proportion as well, which eases the gel electrophoresis separation of bands. The sequences from the desired annealing sites were copied from IGV and optimized as primers using the NCBI primer-designing-tool. The SbRio reference genome was not integrated into the NCBI database. Therefore, the specificity of primer-pairs was examined using the blast-tool in the phytozome database. Finally, the designed size-markers were validated using amplicons from the parental lines prior to use for the recombination analysis. The markers determined the parental origin of genomic regions in the progenies based on size differences when amplified and run through gel electrophoresis. Since genomic DNA samples from F2 progenies were used for the recombination analysis and the plants were no longer present, all F3 families of recombinant F2 progenies were evaluated to ensure the true-to-type disease phenotype. A recombination analysis was considered complete when the marker-type in a recombinant F2 plant was reconciled to the disease phenotype of its F3 progenies. This cycle of recombination analysis was carried out several times by designing new set of DNA size-markers, in a way tracking closer and closer to the then unknown *ARG2* locus.

### 2.2.5   Generation of *ARG2* near-isogenic lines

Near-isogenic lines (NILs) that differ in the *ARG2* locus were generated to study the molecular biology of *ARG2* resistance by minimizing the background noise. BC5F3 NILs were developed using the susceptible parental line TAM428 as a recurrent mother plant (Figure 2-1). In the development of the NILs, the controlled pollination of the backcross crossing scheme were performed using the plastic-bagging method, which does not completely avoid self-pollination (43, 44). In addition, since the pollen source is heterozygote, the backcross progenies are expected to carry the heterozygote or recessive homozygote *ARG2*. There was no molecular marker to track *ARG2* locus when the NILs were developed, and the F1 progenies of the backcross generations were identified primarily using their disease phenotypes. This F1 screening method necessitated the evaluation of their F2 progenies at each backcross generation to control disease-escape rather

than a true BCF1 resistance. The resistant BCF1 (heterozygote) genotype of all the backcross generations is expected to show the 3:1 disease phenotype at BCF2.
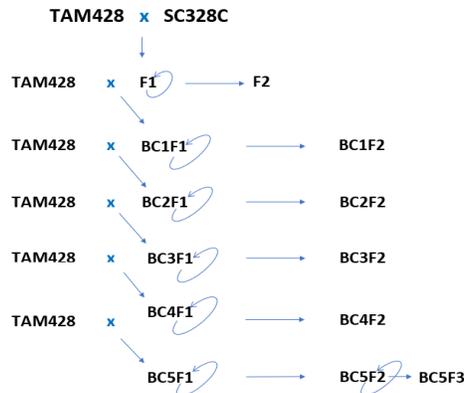


Figure 2-1. Diagrammatic illustration of the development of ARG2 near-isogenic lines.

### 2.2.6  Downstream mapping

Several strategies were pursued to map *ARG2* locus. Once the number of candidate genes were drastically reduced, the fine-mapping employed comparative genomics (CoGe) including the examination of aligned BAM files of many sorghum variants, gene expression and sequence analysis based on targeted sequencing (wide-Seq) of the likely candidate genes and potential gene regulatory segments in the *ARG2* genomic map interval. BLASTp, protein databases including SMART and CoGe were used to evaluate the functional features and conserved coding sequences of the candidate genes.

### 2.2.7  RNA sample preparation

RNA samples were used to study gene expression. Most of the gene expression analyses were carried out using leaf tissue sample that were inoculated at about 30 days after planting, like the growth stage used to evaluate the SA disease phenotype. To prepare tissue samples for RNA extraction, resistant and susceptible *ARG2* NILs were spray-inoculated. The youngest, fully-grown, whole leaves that usually manifest the heavy disease pressure were collected from several plants and bulked as a single tissue sample. The tissue samples were collected prior to inoculation and at

different timepoints after inoculation including mock-treated samples. Since the disease phenotype appears after the tissue samples were collected, before the tissue samples were used for RNA extraction, the success of the disease assay was checked on the remaining leaves of the sampled plants and other control plants that were inoculated but not sampled. Three independent pooled-tissue samples were collected for both NILs and at all time-points, and RNA extraction was conducted separately.

Trizol reagent was used to extract RNA as described in the manufacturer's instruction. The tissue samples were ground using mortar and pestle, and a sample of the ground tissue was collected in an eppendorf tube for the RNA extraction. The entire sample processing was carried out under liquid nitrogen. An equal volume of RNA that was extracted from the independent pooled-tissue samples were mixed to represent the final sample for both NILs and all time-points. Thus, a single sample that represented a genotype (pathogen or mock treated) at a time point was used for cDNA synthesis. Three technical replicates were used for qRT-PCR using actin as a reference transcript level. The kinetics of *ARG2* gene expression was examined in the resistant NIL using the pathogen and mock inoculated samples. For defense response genes, the kinetics of gene expression was compared between the resistant and the susceptible NILs. The mock inoculated resistant samples were used to separate genes that were induced in the pathogen inoculated samples from genes that were presumed to be induced due to the chamber environment where the inoculated plants were kept for 48 hours.

### 2.2.8   Trait association to *ARG2* locus

The *ARG2* NILs were grown in the Purdue University Plant Sciences controlled environment phenotyping facility without the pathogen inoculation to evaluate some agronomic traits that may have association to *ARG2* locus. The plants were supplied with minimal nutrition to reduce the chance of compensation-growth that may otherwise be shown due to the difference in the *ARG2* locus. Shoot biomass was evaluated among the homozygote resistance, heterozygote and homozygote susceptible genotypes. Analysis of variance (ANOVA) was conducted using three replications of the genotypic groups, and each group was represented by three plants.

## 2.3 Results

### 2.3.1 Identification of resistant sorghum lines.

To identify sorghum genotypes with increased resistance to Cs, a genetic screen was conducted. Multiple rounds of disease assays on 25 sorghum lines were conducted using multiple strains of the fungus. The sorghum line SC328C displayed clear-cut resistance to Cs strains Cs-gl1 and Cs-grg but was susceptible to Cs-gl2, Cs-27, CS-29 strains (Table 2-1). SC238C is part of a collection of sorghum differential lines that were used for race identification and its resistance to some Cs races were reported in previous studies (18, 45). Due to the strong resistance phenotype in SC328C, this line was studied in detail and the results are presented in the next sections.

| SN | Sorghum lines | Strains of *Colletotrichum sublineolum* | | | | | Remark |
|---|---|---|---|---|---|---|---|
| | | Cs-gl1 | Cs-gl2 | Cs-grg | Cs-27 | Cs-29 | |
| **1** | **SC328C** | R | S | R | S | S | **Clear-cut disease phenotype** |
| **2** | **IS18760** | S | S | S | R | R | **Easily distinguishable disease phenotype** |
| 3 | P954206 | R | S | R | S | S | |
| 4 | Brandes | R | R | S | R | R | Photoperiod sensitive |
| 5 | PI569979 | R | R | S | R | R | |
| 6 | BTx398 | S | S | S | - | - | - |
| 7 | PI570726 | S | S | S | S | S | - |
| 8 | PI570841 | S | S | - | S | - | - |
| 9 | QL3 | R* | S | S | S | R* | |
| 10 | SC112-14 | S | S | S | S | S | |
| 11 | SC326-6 | R* | S | S | R* | R* | |
| 12 | SC414-12E | S | S | S | S | S | R* = noisy host response |
| 13 | Theis | R* | S | S | R* | R* | [wide spreading pigmentation] |
| 14 | P954130 | S | R* | S | - | S | |
| 15 | P954062 | S | S | S | - | - | |
| 16 | SC211 | R | R* | S | S | S | |
| 17 | BTx623 | S | S | S | S | S | The *S. bicolor* reference genome |
| 18 | TAM428 | S | S | S | S | S | Highly susceptible |
| 19 | SC283 | R | R | R* | S | R | |
| 20 | P9830 | R | S | R | R | S | Previous or current anthracnose resistance mapping studies are associated to these lines |
| 21 | BTx378 | R | R | - | - | R | |
| 22 | SC748 | R | R | R | S | S | |
| 23 | Tx2536 | R | S | S | S | S | |
| 24 | PI525695 03 | S | S | R | - | R | |
| 25 | PI585749 02 | S | S | S | - | R | |

Table 2-1. Response of sorghum variants to five strains of Cs.
The disease phenotypes were established after evaluations in several disease assay. About 30 days old plants were inoculated with the fungal spores and disease scored at seven and ten days after spray-inoculation. "R" denotes disease resistance. Resistance responses range from minute HR-like responses marked by brown dotes to highly localized pigmentation to a highly restricted necrotic tissue. "R*" denotes no typical anthracnose symptom but the inoculation resulted in a widespread purple to red pigmentation on the leaf surfaces. "S" denotes susceptibility with expanding infection that shows the typical disease symptom in multiple leaves per plant. Host-responses that were either ambiguous or not recorded were denoted by the dash sign.

The typical symptom in susceptible genotypes appears usually in five to seven days after inoculation. The pronounced SA symptom usually appears on the leaf sheath when the whole plant was spray-inoculated, and the tissue often tears apart along the venation which may relate to the inoculum load due to runoff. The resistance response among the differential lines was also variable which was often associated to the level of pigmentation of the infected tissues. In some of the lines, the pigmentations cover the entire surface of infected leaf tissue and obscure the typical symptom in the susceptible genotypes. Some plants of the resistant lines were seen collapsing completely when younger plants, before 30 days after planting, were inoculated.
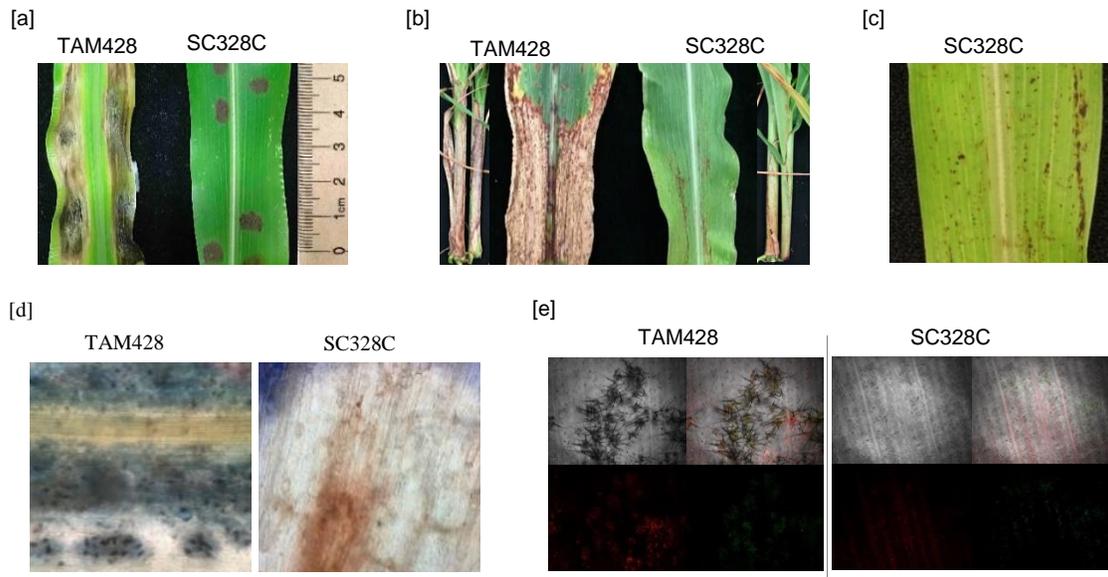


Figure 2-2. Contrasting disease phenotypes and fungal growth on TAM428 and SC328C.
[a] Disease phenotype in detached leaf disease assay. 15-20 µl of a conidial suspension [$1x10^6$ conidia/ml] was drop inoculated at several spots of the leaf surface and incubated on a lab bench at 20-24$^o$c. The images were taken at 10 dpi. [b] The disease phenotype on leaf and leaf sheath at 14 dpi, after the whole plants were spray inoculated. [c] A magnified view of inoculated leaf of the resistant genotype at 7 dpi in spray inoculated plants. [d] Fungal growth revealed by trypan blue staining of spray inoculated tissue. [e] Fungal growth revealed on detached leaf tissue at six dpi observed under confocal microscope. The pathogen acervuli was observed on the susceptible parent, TAM428 whereas no pathogen growth was observed in the resistant parent, SC328C. The fungal hyphae were stained green by WGA-AF488. Propidium iodide stained leaf tissue [red] shows the dead cells.

Visual examination of the disease phenotype was used as the primary method to evaluate the host-response to the pathogen. The accumulation of acervuli with setae as minute black spots that are visible to the naked eye (Figure 2-2), the typical symptom of SA disease, was the major criteria for susceptibility. In spray-inoculated plants, the resistance in SC328C showed HR-like symptoms at the site of attempted infection (Figure 2-2c). The TAM428 disease phenotype shows a widespread disease lesion and enhanced pathogen growth in multiple leaves per plant. The disease

lesion usually covers larger than half the area of actively growing leaves and the entire leaf sheath. At seven to ten days post inoculation (dpi), disease symptoms ranging from widespread necrotic lesions to complete collapse of infected leaves or the entire plant was observed. In the detached leaf assay, browning of the leaf tissue of the resistant genotypes occur at the spots where the inoculum droplets were placed but the tissue remained intact and disease lesion was not observed. By comparison, TAM428 displays enhanced disease symptom with expanding chlorotic and necrotic lesions beyond the site of inoculation. The pathogen proliferates and debilitate the TAM428 leaf tissue.

Staining the inoculated leaf tissue showed the difference in fungal growth between the susceptible and the resistant lines. Trypan blue staining of infected leaf tissue revealed dense masses of mycelia on the TAM428 leaf tissue whereas no pathogen growth was observed on SC328C (Figure 2-2d). Similarly, staining of the inoculated tissue with WGA-AF488, which stains the fungal hyphae (46), clearly revealed the proliferous acervuli (Figure 2-2e) on TAM428 but not on SC328C.

### 2.3.2   Genetic analysis and BSA-Seq

To determine the genetics of the resistance in SC328C to Cs-gl1, SC328C was crossed to two susceptible lines BTx623 and TAM428 (Table 2.1), and the genetics of SA disease resistance was examined in the two populations. The F1 progenies of both crosses were resistant, and the F1 progenies of the reciprocal SC328C x TAM428 cross were also resistant. The F2 progenies of the BTx623 x SC328C cross resulted in 298 resistant and 91 susceptible progenies, which fitted to the classical 3:1 ($X^2$ = 0.5) phenotypic ratio. Similarly, evaluation of 1131 F2 progenies of the TAM428 x SC328C cross resulted in 853 resistant and 278 susceptible plants that fitted to the 3:1 ($X^2$ = 0.10) phenotypic ratio. The 1131 F2 plants of the TAM428 x SC328C cross were evaluated in three rounds and each round of the disease phenotype, as well, satisfied the 3:1 ratio. The disease symptom, on BTx623 appears slower and less severe than on TAM428. Similarly, in most susceptible progenies of the BTx623 population, the pathogen took longer to develop the disease symptom as compared to that of TAM428. Thus, the TAM428 population was pursued to ease the phenotypic evaluation in the subsequent genetic mapping.
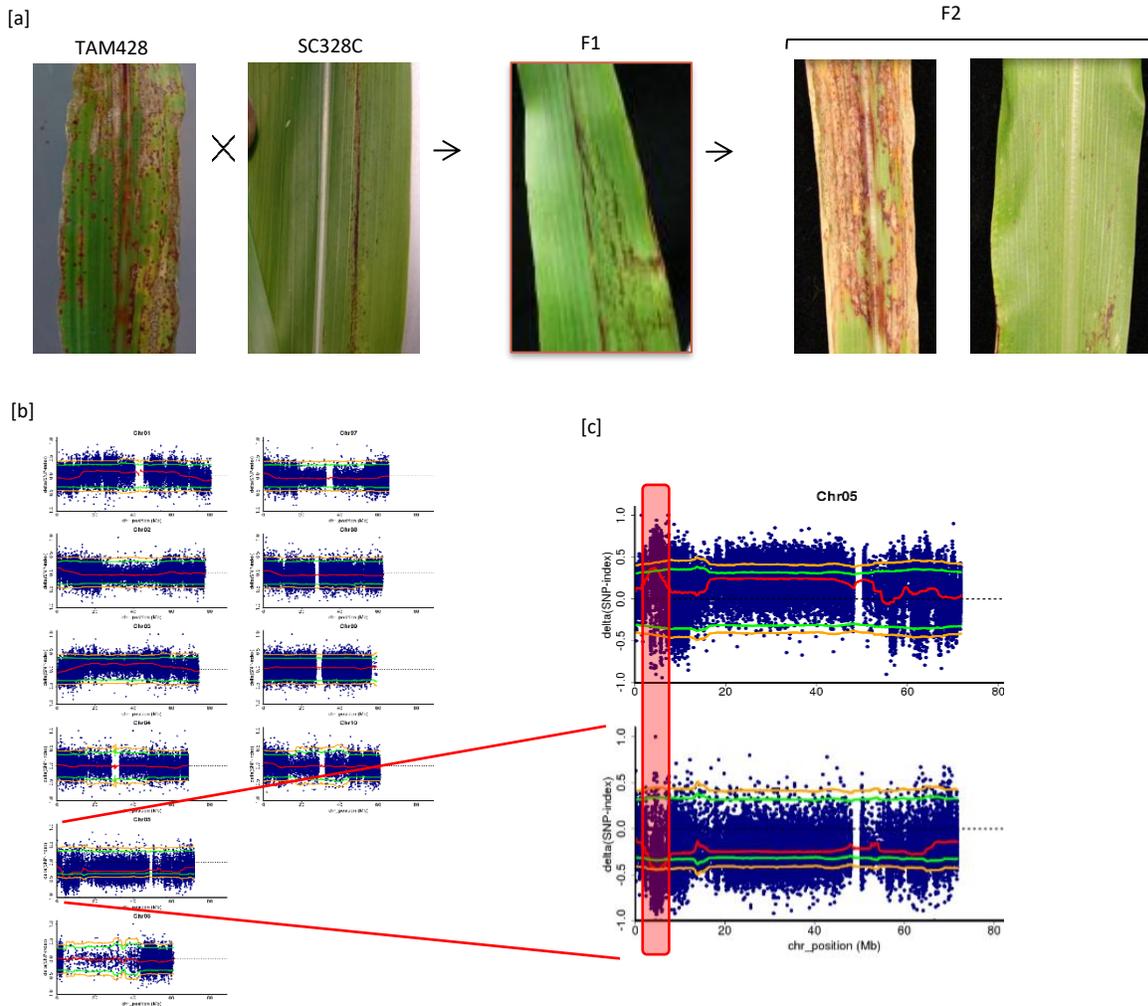
Figure 2-3. The genomic region that determines the monogenic disease phenotype in the ARG2 mapping population was mapped using BSA-Seq.
[a] F1 progenies were resistant, and the F2 progenies showed robust resistance or high susceptibility like the resistant and the susceptible parental lines. [b] BSA-Seq plot on the ten chromosomes of the *Sorghum bicolor* reference genomes. [c] A magnified view of the BSA-Seq plot on chromosome 05, with respect to the resistant [upper] and the susceptible [lower] parental lines. The x-axis refers to the genomic coordinates along each chromosome, and the y-axis refers to the ΔSNP-index estimates. The vertical positions of the x-y axes, as represented by the blue dots – most of them overlapped, represent the ΔSNP-index estimates along the 2 Mb sliding window. The red line shows the regression plots of the ΔSNP-index estimates. The green lines show the significant threshold [p=0.05] of ΔSNP-index estimates along the chromosome and the pink at p=0.01. The vertical rectangular bar [red] shows the genomic position of *ARG2* as located by BSA-Seq.

The third-round phenotypic evaluation was carried out on 336 F2 plants during recombination analysis, and hence, we were able to accompany the genetic analysis with genotyping. The *ARG2* genotypes in the 336 F2 plants cluster to 92 homozygous resistant, 166 heterozygote and 78 homozygous susceptible based on the closest DNA marker that fitted well to

the canonical 1:2:1 ($\chi^2=1.4$) genotypic ratio. The genotypes showed complete correlation with the disease phenotype.

The genetic data on the F1 and F2 progenies suggested that *ARG2* disease phenotype in this *ARG2* mapping population is determined by a single nuclear gene locus, and the resistance is inherited as a dominant allele. Like the parental lines, evaluation of the disease phenotype in the mapping population, and later in *ARG2* NILs that were developed based on the same parental lines, showed a complete consistence in disease phenotype between the strains Cs-gl1 and Cs-grg. Thus, this single dominant locus in SC328C confers resistance to both strains (Table 2-1) and designated as *ARG2*.

I conducted mapping using BSA-Seq to locate the genomic region that carries the *ARG2* locus. In line with the genetic analysis, BSA-Seq analysis of the genomic raw-reads of the pooled-progeny and the parental genomic DNA samples located a single peak genomic region around the proximal end of chromosome 5 (Figures 2-3b, and 2-4 a and b). The statistical determinant of the casual genomic region of the disease phenotype is ΔSNP-index, which is the difference in the average allelic frequencies between the resistant and the susceptible pooled-progeny genomic samples. The significant ΔSNP-index estimates range from 0.32–0.47 with the lowest threshold estimate of 0.31 at 95% CI and 0.42 at 99% CI. Chromosomal coordinates of the significant ΔSNP-index estimates, in the SbRio reference genome, span 2.6–7.05 Mb at 95% confidence interval (CI) and 3.8–6.2 Mb at 99% CI (Figure 2-4 b and c). In the wider significant genomic region (the interval in the 95% CI), smallest depth of the aligned genomic raw-reads and SNP count in the 2 Mb sliding window that slide 50 Kb at a time range 25–29 and 1067–2872 with mean values of 26.8 and 1879, respectively. According to the BSA-Seq statistics, *ARG2* candidate locus spans about 4.5 Mb genomic region around the peak.

### 2.3.3    Recombination analysis

The BSA-Seq analysis mapped *ARG2* to near the proximal end of chromosome 5 (Figures 2-3c and 2-4a). Subsequently, DNA size-marker based recombination analysis was conducted using F2 progenies with respect to their disease phenotypes (Figure 2-4b) to define and narrow the mapping interval of *ARG2* locus. The resistant F2 progenies were not used in the recombination analysis to reduce the time required to resolving the true-to-type phenotypes from the heterozygotes, as *ARG2* resistance is inherited as a dominant allele. The susceptible F2 progenies are expected to harbor the candidate genomic region in the homozygote state like TAM428. Thus, the recombination-based map was established using 72 susceptible F2 plants, which was later consolidated with 336 more F2 plants as discussed earlier.
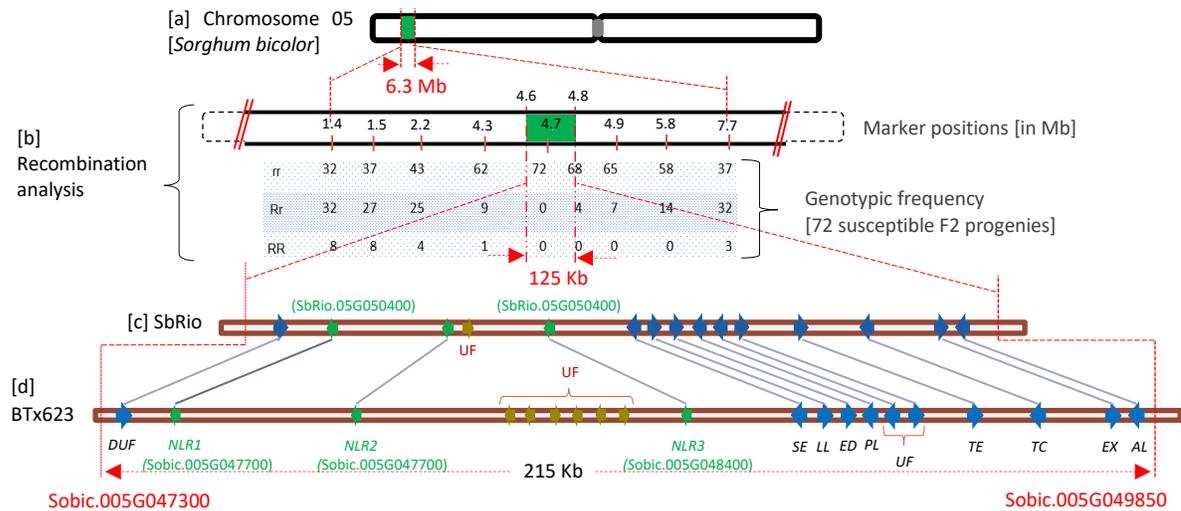


Figure 2-4. A diagrammatic summary of *ARG2* mapping.
[a] The genomic region of *ARG2* as identified by BSA-Seq. *ARG2* locus was mapped to the near-proximal end of chromosome 05 [green shaded]. [b] A magnified view of the 6.3 Mb genome around the *ARG2* genomic region that was identified by BSA-Seq. Recombination analysis spans this 6.3 Mb region using 72 susceptible F2 progenies. The row of numbers on the upper panel shows the position of DNA size-markers with which the genotypic frequencies [lower panels] were determined. The green shaded 125 Kb segment was the genomic region of *ARG2* that was defined by the recombination analysis. [c and d] The genomic view of all the genes available within the 125 Kb *ARG2* genomic region in the SbRio reference genome [c] and in the corresponding genomic region [215 Mb] of the BTx623 reference genome [d], 90 Kb wider in the BTx623 reference genome. The short arrows [most of them shared light blue] along the two *ARG2* genomic region in the two reference genomes show the orientations of individual genes. Predicted homoeologous genes between the two reference genomes were linked by the light-blue lines. All the genes were labelled using the initials of annotated functions as follows: *DUF*: Domain of unknown function; UF: unknown function; *NLR*: *NBS-LRR*; SE: Serine esterase; *LL*: Limkain B Lkap; *ED*: Nad Dependent Epimerase/Dehydratase; *PL*: TRNA-Nucleotidyl Transferase/Poly A polymerase family member; *TE*: TRANSCRIPTION ELONGATION FACTOR B POLYPEPTIDE 3; *TC*: CCR4-NOT Transcription complex subunit; *EX*: Exocyst Subunit EXO70 family protein C1-related; *AL*: Alginate lyase. The three *NLR* genes [green shaded] were the likely candidate genes.

Fifteen markers were designed for the recombination analysis in the 6.3 Mb wide genomic region that ranges from the coordinate 1.4 Mb to 7.7 Mb. The 6.3 Mb region spans additional 1.2 Mb to the left and 0.7 Mb to the right of the significant *ARG2* genomic region, which is about 4.5 Mb as identified by BSA-Seq. Nine-markers were used to show the simplified bidirectional gradient of the recombination frequencies around the candidate genomic region (Figure 2-4b) that pointed at the locus of interest within the genomic region that was identified by BSA-Seq. In addition, this recombination analysis resulted in a well-defined 0.5 Mb *ARG2* genomic region that was flanked by markers at 4.3 Mb and 4.81 Mb.

Later, another recombinant plant was found at 4.69 Mb in the NILs. Accordingly, the recombination analyses resulted in a 125 Kb *ARG2* genomic region. This *ARG2* map is flanked by the markers at 4.69 Mb and 4.81 Mb (Figure 2-4 b and c) and the predicted genes SbRio.05G050200 and SbRio.05G051900 that are located at 4.686 Mb and 4.810 Mb, respectively. The BTx623 and SbRio reference genomes show about 88 Kb genomic size disparity in the *ARG2* genomic region, and this difference appears similar between SC328C and TAM428. The corresponding flanking genes in the BTx623 reference genome, Sobic.005G047300 and Sobic.005G049850, are located about 215 Kb apart (Figure 2-4d).

### 2.3.4  *ARG2* downstream fine mapping

Throughout the marker analysis, the IGV view of the parental genome BAM files showed striking sequence similarities and disparities among the four variants of sorghum (Figure 2-5). SbRio is a sweet sorghum that produce high vegetative biomass (47) whereas SC328C is dwarf and appears grain type. SC328C and SbRio have similar DNA sequences in the 125 Kb *ARG2* genomic region, and so to the similarities of the 215 Kb *ARG2* genomic region between TAM428 and BTx623 albeit not as closely similar as the former. Consistently, SbRio is resistant like SC328C whereas BTx623 is susceptible like TAM428. Thus, the sequence information in the SbRio and BTx623 reference genomes facilitated the analyses of the sequence variation in the parental lines throughout the fine mapping.

[a]

SbRio    SC328C    TAM428    BTx623

[b]

BTx623    TAM428    SC328C    SbRio
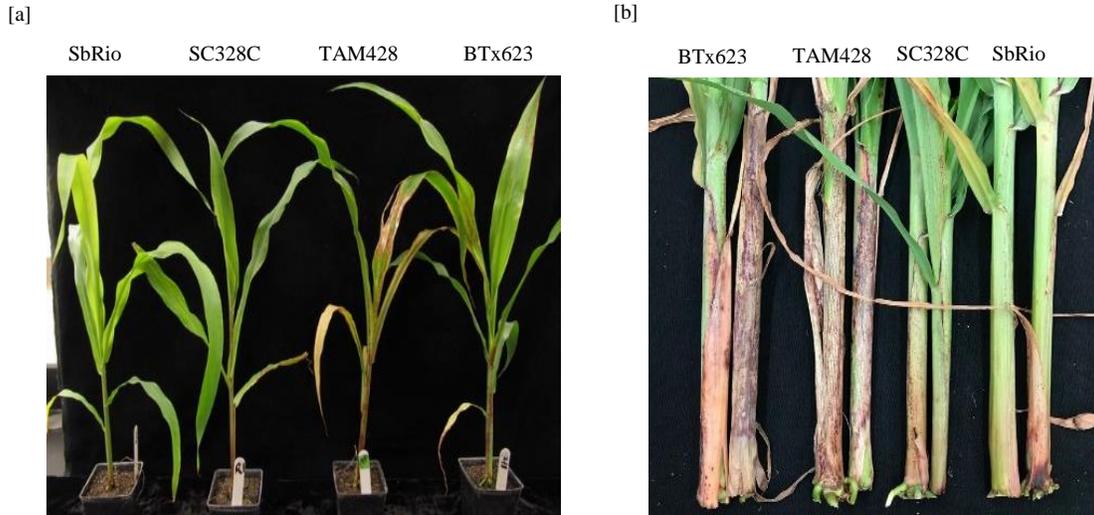
Figure 2-5. *ARG2* disease phenotype on the parental lines as compared to the lines of two S. bicolor reference genome.
[a] Whole plant view of *ARG2* disease phenotype on BTx623, SbRio and the parental lines of *ARG2* mapping population at
7 dpi, and [b] Disease phenotype on leaf sheath at 15 dpi.

The major exclusion of non-candidate genomic regions was achieved through BSA-Seq followed by recombination analyses. The 125 Kb *ARG2* genomic region harbors 15 predicted genes (Figure 2-4c). DNA sequence differences of all the genes, between the parental lines, were examined using the BAM files on IGV and the predicted gene functions. Three of the 15 predicted genes carry *NLR* domain. These *NLR* genes have highly similar DNA sequences that caused many parental DNA raw-reads to map ambiguously at multiple sites. The rest of the genes were nearly identical between the parental genomes and did not show any high impact SNPs. The sequences were examined further in >20 other re-sequenced sorghum variants with known disease phenotypes and showed similar ambiguously mapped raw-reads in the three *NLR* genes loci. Although ambiguous, the alignments in these sorghum variants showed a widespread SNP pattern that was particularly associated to the disease phenotype.

Since *ARG2* is inherited as dominant allele, the disease resistance is likely conferred by a functional allele. Accordingly, first, I wanted to figure out the number of intact genes that are known to play role in disease resistance. However, the overriding feature of the mapping concept is polymorphism between the parental lines. In this regard, majority of the 15 genes do not show high impact SNPs between the parental lines. In addition, these genes do not carry functional domain that is known to play role in disease resistance. Thus, the three *NLR* genes are likely to be

51
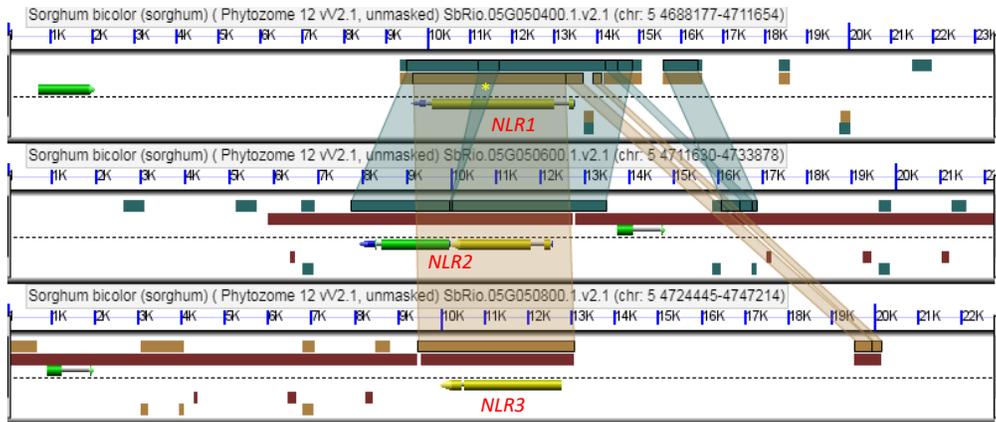
Figure 2-6. CoGe view of the three duplicate NLR genes in the ARG2 locus.
The resistant parent and SbRio reference genome carry the same *ARG2* locus, and the homologous segments of the three duplicate *NLR* genes were viewed on the CoGe platform. The lowest panel shows a big indel that detached the upstream region of *NLR3* including some coding sequences of the gene [as shown by the light pink connector between *NLR1* and *NLR3*]. *NLR2* refers to the middle of the three duplicate *NLR* genes. *NLR2* was predicted as two genes [SbRio.05G050500 and SbRio.05G050600] in the reference genome although the two predicted gene sequences code for a single transcript. The dark-green triangular connector between *NLR1* and *NLR2*, marked by the yellow asterisk [*] shows a 442 bp homologous segment of *NLR1* that was duplicated in *NLR2*. There was no disruption of *NLR1* was observed *NLR1*. CoGe analysis of *NLR1* in related crop species showed the coverage of the complete coding sequence in NLR1 has been conserved.

*ARG2* based on the polymorphisms between the parents and the functional domains of the genes. The small, predicted gene of unknown function (UF) near *NLR2* that is shown only in the resistance genome does not carry ORF (Figure 2-4c).

The major clue about a single *ARG2* candidate gene, from among the three duplicate *NLR* genes, was obtained from CoGe analysis (Figure 2-6). In the resistant parent, *NLR3* carries a >9 Kb transposon-like indel in the upstream region, which may have separated part of the CC-domain together with the upstream regulatory sequence away from the main body of the gene *NLR2* represents the middle *NLR* gene, which was predicted as two genes. However, I amplified a single transcript that spans the two predicted genes, and the coding sequence was verified by sequencing and homology analysis. *NLR2* show a 442 bp duplication in the middle of the gene, which may have caused disruption of the ORF and mis-annotation of the single gene as two genes. Thus, CoGe analysis of the three highly similar *NLR* genes indicated that two of the three *NLR* genes may not be intact in the resistant parent.

Analyses of transcript in the three *NLR* genes supported the CoGe analysis. The attempt to design specific primers to clone the complete transcripts of the individual *NLR* genes was challenging due to the high sequence similarity. Complete transcripts were amplified using specific

primers in the UTR regions, and these PCR products were used as templates to clone the coding sequences of the genes. The complete coverage of the coding sequence of *NLR1* was determined based on the widely conserved homologous genes in monocots. Transcripts were detected from *NLR1* and *NLR2,* but not from *NLR3*. The complete transcript sequence showed presence of a premature stop codon in *NLR2* near the predicted translation start site and downstream of the 442 bp duplicate. Thus, *NLR2* and *NLR3* are non-functional genes in the resistant parent, and *NLR1* is the only likely *ARG2* candidate gene. *NLR1* refers the gene ID SbRio.05G050400 in the *S. bicolor Rio v2.1* and Sobic.005G047700 in the *S. bicolor v3.1.1* reference genomes in the phytozome database. *NLR1* is intact in the resistant parent and regarded as a wild-type allele whereas the gene carries a premature stop codon in the susceptible parent and hence regarded as a mutant non-functional allele (Figure 2-7).
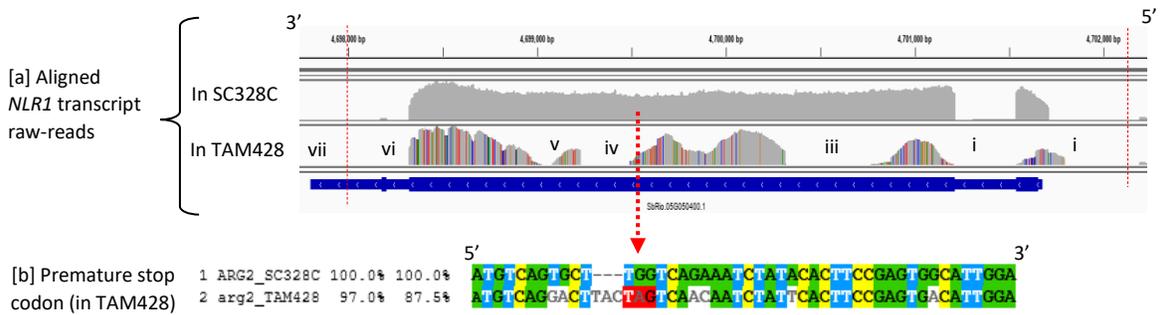


Figure 2-7. The genomic view of NLR1 transcript raw-reads and position of premature stop codon.
[a] The genomic view [IGV] of *NLR1* transcript raw-reads in the parental lines aligned to the SbRio reference genome. The gene is in the reverse orientation. The transcript sequences were generated from single PCR products using wide-Seq. The numbers on top were the genomic coordinates of *ARG2* on chromosome 5 in SbRio reference genome that corresponds to 4.509-4.513 Mb in the BTx623 reference genome. The height of the gray area shows the depth of aligned reads with identical base-pair sequences, range from about 700 reads at the peripheries to >7000 at the highest depth. The reads in SC328C and SbRio reference genome were identical. The red, orange, blue and purple colored vertical lines in the lower panel [gray background] referred to the SNPs in the TAM428 allele. The gene regions on the TAM428 panel that did not show aligned reads [no gray background at all] were a 62 base-pair indel [iv] and other exon regions [iii and v] that carry dense SNPs that prohibited alignment. The other empty regions were intronic regions [i, ii, vi] and the 3'UTR [vii] that was not included in the PCR amplification. The gene structure [blue] below the two panels was the predicted NLR1 gene [the GFF file]: the thicker segments referred to exons and the thinner once were introns and UTRs. [b] DNA sequence alignment [bottom] shows part of the coding sequence in SC328C and TAM428 that carries the premature stop codon in TAM428. This sequence alignment was done for the complete coding sequences, and hence the statistics in the left side of the sequence belongs to the entire coding sequences of *NLR1*.

### 2.3.5   Identification of independent *ARG2* mutant alleles

Sorghum variants were examined to find a mutant *ARG2* allele that is different from the allele carried by the susceptible parent to validate the candidate *ARG2* gene. Greater than eleven susceptible sorghum variants were identified, but transcript was amplified only from 11 variants. The susceptible variants were BTx642, ICSV745, IS8525, IS18760, IS9830, Malisor84-7, SC23 and PQ-434, BTx623 and TAM428. PCR product of the candidate *ARG2* genomic and transcript sequences were not found in IS3121 and greenleaf. Four resistant sorghum lines: SC328C, SbRio, Keller and SC237 were included in this analysis. Figures 2-8 and 2-9b show the disease phenotype on selected sorghum variants that represent the four allele types using two Cs strains that show alternating virulence. The complete coding sequences of all alleles were examined through targeted amplification and sequencing (Supplemental file 1).
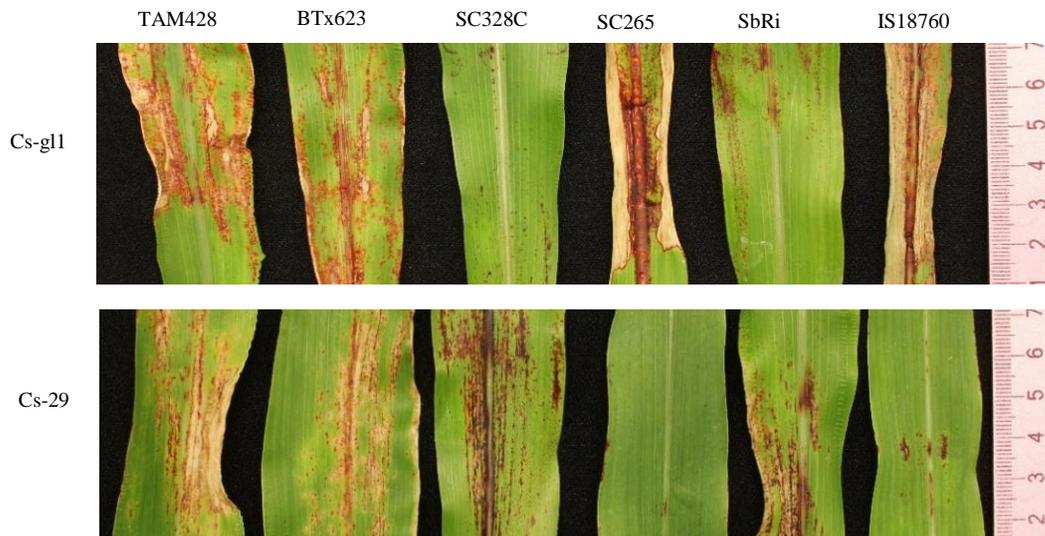


Figure 2-8. Alternating disease phenotypes in four sorghum variants in response to two Cs strains that is used to better show the disease phenotype of the arg2 mutant allele in SC265.
TAM428 and BTx623 are susceptible to both strains of the pathogen. SC328C and SbRio are resistant to Cs-gl1 and susceptible to Cs-29 and vice versa for SC265 and IS18760. The plants were spray inoculated and the images were taken at 7 dpi. SC265 carries a mutant *arg2* allele with a premature stop codon downstream of the premature stop codon in TAM428.

The coding sequences in the total of 15 sorghum lines were clustered to four similarity groups based on general sequence differences, presence of complete ORF and position of premature stop codon. The four resistant variants carry identical *ARG2* coding sequences. The coding sequences in the susceptible variants were clustered in to three: six lines carry the TAM428 allele, four lines carry the BTx623 allele, and SC265 carry a unique allele. The TAM428 and the SC265 alleles carry premature stop codon at different positions.



Figure 2-9. Positions of the premature stop codon in SC265 and TAM428, and phenotypic comparisons. The complete coding segment [dark shaded segments] of *arg2* in SC265 and TAM428 and associated disease phenotypes in six sorghum lines using Cs-gl1 in detached leaf disease assay.

The positions of the premature stop codons in SC265 and TAM428 suggest an independent loss of ARG2 function. The general nucleotide and peptide sequence relationships also support independent trend of divergence of these two alleles to the loss of function (Figure 2-10, Supplemental files 1 and 2). The allele in SC265 carries <8% SNP with respect to the wild-type and a large proportion of these SNPs came from the unique 87 bp deletion whereas the other mutant alleles carry >10% SNP. The codon sequence differences were also compared based on the shortest ORF, which is 1848 bp in the TAM428 allele. The SC265 allele carries an about 4% nonsynonymous codon in this shortest ORF as compared to the wild-type whereas the TAM428 allele carries about 14% nonsynonymous codon differences. This shortest ORF in SC265 carry about 15% nonsynonymous codon as compared to the one in TAM428. Thus, the sequence divergence between the SC265 and the TAM428 alleles was greater than their divergence from

55

the wild-type. Altogether, the mutant allele in TAM428 is distinct from that of SC265. Therefore, the *NLR1* allele in SC265 was regarded as an independent mutant allele to the validation of *NLR1* as *ARG2* gene.
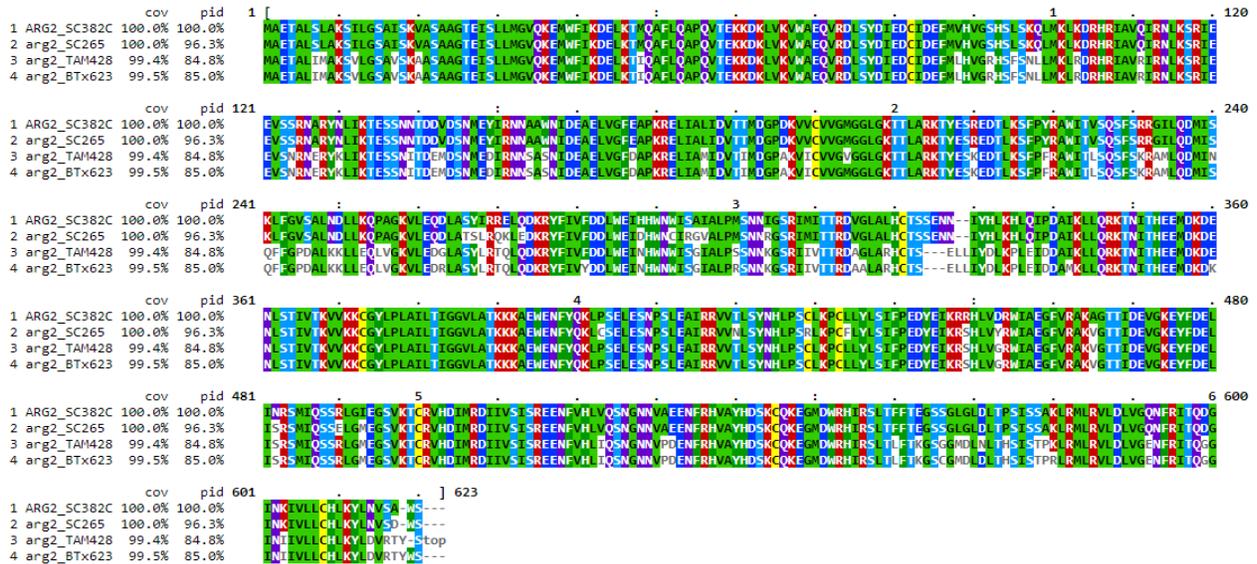


Figure 2-10. Alignment of the deduced amino acid sequences of the four types of ARG2 mutant alleles.
Four types of *ARG2* mutant alleles based on the shortest ORF [in TAM428]: [1] The resistant parent SC328C carries the intact *ARG2* allele, [2 and 3] the two mutant alleles in SC265 and TAM428 carry premature stop codons at different positions and the divergent mutant allele in BTx623 maintains a complete ORF.

The mutant alleles in TAM428 and BTx623 share most of the sequence variation as compared to the resistance allele (Figure 2-10, Supplemental files 1 and 2). The concordance of general sequence divergence in the BTx623 and TAM428 alleles may suggest gradual divergence of these two alleles from the wild-type to the loss of ARG2 function. However, the BTx623 allele maintains a complete ORF. The divergence of *ARG2* allele in BTx623 from the wild-type allele was examined based on the predicted functional domains of the deduced amino acid sequences. The LRR domain that is crucial for pathogen recognition shows about 19% nonsynonymous amino acid substitutions as compared to the resistance allele. The sequence divergence in the LRR domain includes two indel sites (two bp and four bp indels). These two indels flank a 45 bp coding sequence that resulted in a reading frame shift of 15 amino acids. In addition, the amino acid sequences of the two alleles were predicted for specific functional features. The minimum functional module for CC-domain dependent *NLR* homodimerization that triggers cell death in

barley (48) was predicted in the wild-type allele but not in the BTx623 allele. Altogether, the sequence variation in the BTx623 allele as compared to the wild-type likely caused an altered function, and hence loss of function and susceptibility to the fungal strain.

## 2.3.6 Genomic structure of *ARG2* and potential cis-regulatory sequence

The *ARG2* locus seems to be a cluster of three duplicate *NLR* genes especially in SC328C, in which *NLR2* and *NLR3* undergone high impact structural rearrangements (Figure 2-6). However, in TAM428 and BTx623, these duplicate genes exhibit distinct structural and functional features as compared to the wildtype. These genes in TAM428 and BTx623 are found in a wider genomic interval interspersed with several other predicted genes (Figure 2-4 c and d), and no structural rearrangement was observed in *NLR2*. The big indel in the upstream sequence of *NLR3* in the wild-type locus (Figure 2-6) is also observed in TAM428 and BTx623 although the indel size is relatively very small.
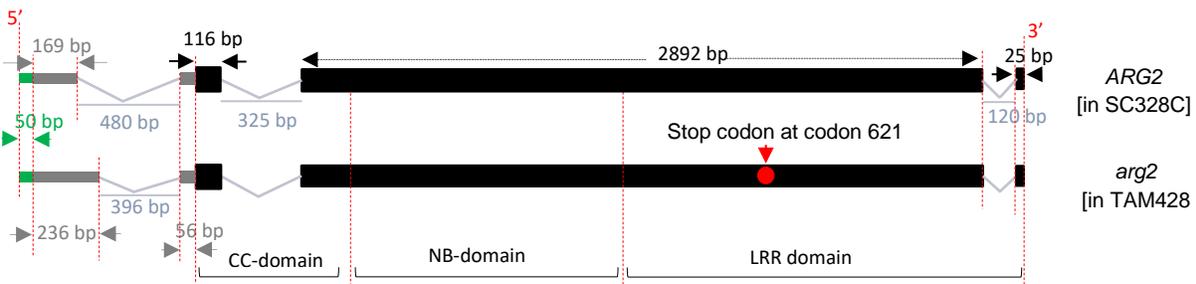


Figure 2-11. *ARG2* gene structure, positions of high impact SNPs in the mutant alleles and associated disease phenotypes in six sorghum lines.
[a] *ARG2* gene structure and potential upstream cis-regulatory elements in the parental lines of the *ARG2* mapping population. The dark Black shaded exons are indicated. ARG2 harbors three exons with a complete ORF in SC328C whereas the allele in TAM428 carries a premature stop codon. The first exon carries a small upstream portion of the CC-domain. The second exon carries most of the coding sequence including the remaining portion of the CC-domain, the entire nucleotide binding domain and nearly the entire LRR domain. The third small exon carries the end of the LRR domain. The green labelled predicted promotor has 50 bp at 697 bp upstream of the coding sequence. According to this promotor prediction, transcription starts inside the promotor sequence. Thus, the 169 bp sequence downstream of the promotor in SC328C is expected to be part of the 5'UTR, that extends from the promotor sequence to the first transcript splice site in the 5'UTR. The 5'UTR splice site in TAM428 is within in the 480 bp sequence that is an intron in SC328C, and hence the 5'UTR at the upstream of the intron splice site is 236 bp [as opposed to the 169 bp in SC328C]. The indels are a single bp and 80 bp [absent in TAM428] and 64 bp [absent in SC328C]. The 56 bp segment shows the remaining 5'UTR that extends up to the translation start site.

In addition, structural and functional differences are available around the *ARG2* cis-regulatory sequence. The wild-type *ARG2* harbors a 480 bp intron in the 5'UTR region, and the transcript carries a 71-codon uORF (Figure 2-11). One of the splice sites of this intron has a SNP, C**T**/CC in the wild-type and C**A**/CC in TAM428, and this splice site is not present in TAM428. However, the 480 bp intron carries many SNPs and three small indel sites, and the TAM428 transcript has a transcript splice site within the 480 bp intron (TGAT in the wild-type, T**AG**GAT in TAM428).

### 2.3.7 ARG2 homologs-
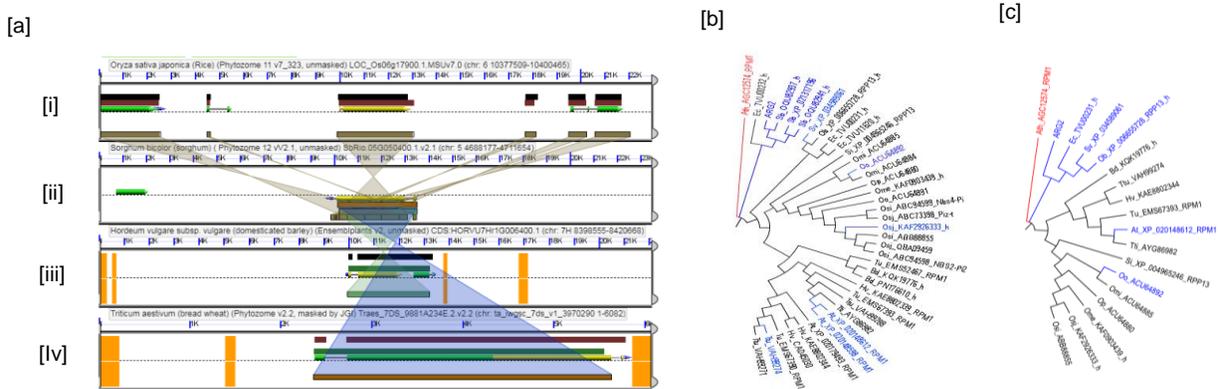


Figure 2-12. A snapshot of the top ARG2 homologs in other monocot species and analyses of phylogenetic tree.
[a] The top rice, barley and wheat homologs identified by CoGe. The different color connectors between the panels show the homologous segment between pairs of genes in the different species. [i] *O. sativa* has a cluster of three highly homologous genes, one of them has gene ID LOC_Os06g17880 [Apoptic ATPase, LRR annotations]. [iii] *H. vulgare* barley homolog has the gene ID HORVU7Hr1G006400. [iv] The top homolog in *T. aestivum* carries 100% sequence coverage, gene ID Traes_7DS_9881A234E [RPM1, Apoptic ATPase, NB-ARC, LRR annotations]. [b] The top ARG2 homologs in 18 monocots identified using BLASTp, and the phylogenetic trees that were constructed using Mega-x and Figtree. The peptide sequences of 37 highest sequence pair [HSP] ARG2 homologs in 18 monocot species and species-groups that were selected from 100 top HSP monocot homologs through NCBI blast search. The *Arabidopsis* RPM1 [homolog of ARG2] was included as a likely outlier for peptide sequence comparison purposes. In species and species-groups that were represented by more than three homologs within the 100 monocot homologs, the upper three HSP homologs were used. [c] The tree construction was repeated using 18 homologs that represented each species only once. The two or three letter prefix refers to the initials of the species name that harbors the peptides: Sb = *Sorghum bicolor*; Sv = *Setaria viridis*; Ttu = *Triticum turgidum;* Ome = *Oryza meyeriana*; Osi = Oryza sativa Indica Group; At = *Aegilops tauschii*; Tti = *Triticum timopheevii*; Osj = Oryza sativa Japonica Group; Oo = *Oryza officinalis*; Op = *Oryza punctate*; Ec = *Eragrostis curvula*; Si = *Setaria italica*; Hv = *Hordeum vulgare*; Omi = *Oryza minuta*, Bd; *Brachypodium distachyon*; Ob = *Oryza brachyantha*; Tu = *Triticum urartu*; Ath = *Arabidopsis thaliana*. The suffix 'h' refers to a hypothetical protein as shown in the blast result.

The homologs of ARG2 were examined in the genomes of monocot and dicot plant lineages. Comparative genome analyses revealed that ARG2 has orthologs in japonica-rice (#49), *Aegilops tauschii* (#51) and bread-wheat (#239). The rice orthologs include a cluster of three *NLR* genes like the *ARG2* locus, and a 100% coverage of peptide sequence was observed with a

predicted RPM1 in *Triticum aestivum* and *T. durum* (Figure 2-12a). BLASTp search revealed ARG2 as the *S. bicolor* RPM1 homolog with widespread homologs in monocots and dicots. The *Arabidopsis* RPM1 homolog confers resistance to multiple strains of *Pseudomonas syringae (49)*. The top 100 monocot and 100 dicot ARG2 homologs showed strong homology to ARG2, E-value=0.0. Many ARG2 homologs of monocots showed a 99% sequence coverage, and the dicot homologs showed >90% coverage.

However, sequence identity of all the homologs to ARG2, except the duplicate genes at the *ARG2* locus were <60%. Nevertheless, the top hit in *Zea mays* showed only 10% sequence similarity with <25% blast quality which may not show anything beyond the overall structural similarity among NLR protein. As compared to the 1010 amino acids in ARG2, nearly all the monocot homologs carry 950-1025 amino acids whereas nearly all the dicot homologs carry 870-920 amino acids. The last 55-70 amino acids of ARG2 at the C-terminal are absent in the dicot homologs, and this small peptide motif showed homology to a protein sequence of bacterial endotoxin in SMART protein database. The best homolog in *Oryza meyeriana and Eragrostis curvula* clustered with ARG2 as opposed to the taxonomic group (Figure 2-12 b and c).

### 2.3.8 *ARG2* gene expression

Analysis of gene expression showed *ARG2* was slightly induced at 24 hpi in response to pathogen inoculation as compared to the mock (Figure 2-13a). At the later timepoints, however, the elevated expression was comparable between Cs inoculated and mock inoculated samples. The expression of *ARG2* in different plant tissues and growth stages was also examined without pathogen inoculation (Figure 2-13b). The flag leaf showed by far the highest level of expression of *ARG2*. The leaf sheath and the leaf blade at the two earlier growth stages showed relatively low accumulation of *ARG2* transcript whereas *ARG2* transcript was hardly detected in the stalk and the panicle.

[a]

[b]
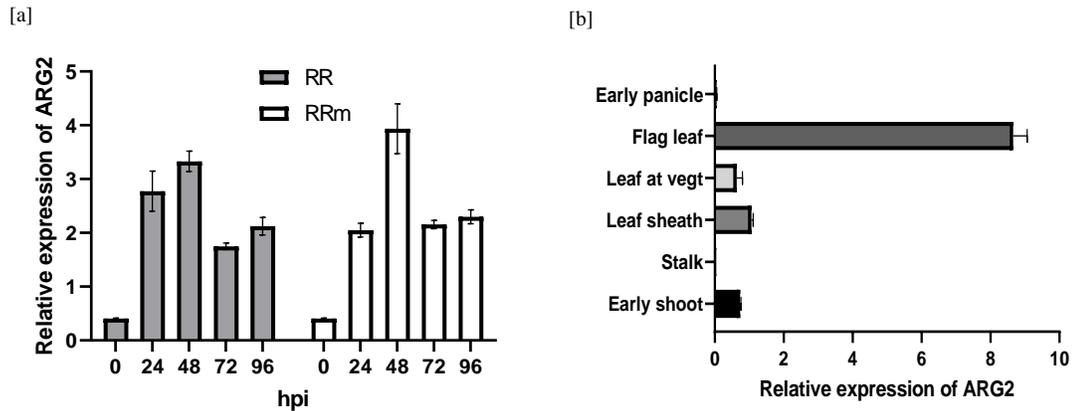


Figure 2-13. ARG2 expression in response to Cs, in different tissue and at different growth stage.
[a] *ARG2* gene expression in response to pathogen inoculation... [b] The expression of *ARG2* in different tissues and growth stages in non-inoculated samples. This tissue sample was collected from the SC382C and TAM428. The tissue samples were the entire shoot at 10 days after planting [dap]; the stalk, leaf sheath and leaf tissues [leaf at vegt] all at 30 dap; and the flag leaf and the panicle at early flowering.

*ARG2* NILs were developed to study the biology of *ARG2* resistance based on a common genomic background. The resistant and the susceptible NILs inherited the clear-cut difference in the disease phenotype (Figure 2-14a) and the proliferous pathogen growth like the difference between the parental lines (Figure 2-14b). The expression kinetics of *ARG2* and potential immune response genes were examined in response to pathogen inoculation to generate molecular information about the resistance mediated by *ARG2*.
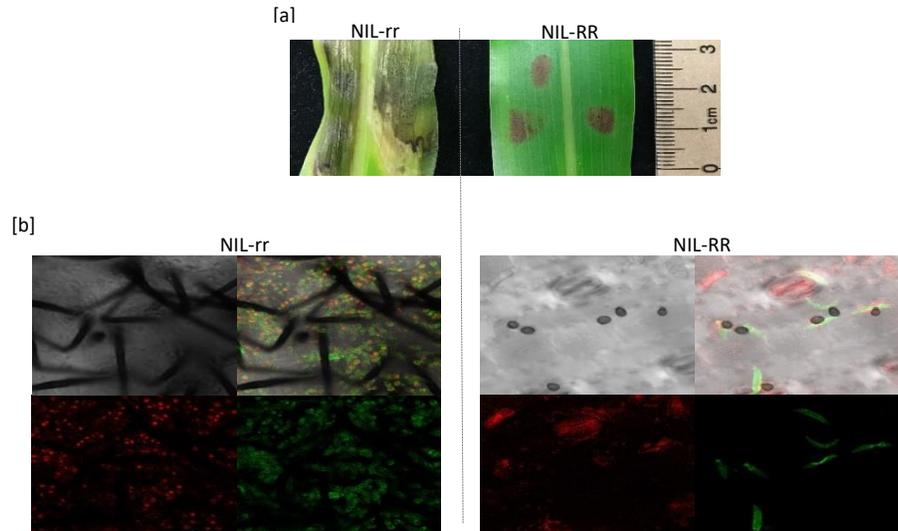
Figure 2-14. The disease phenotype and pathogen growth in ARG2 NILs look exactly like the parental types.
The detail of the development of the NILs is discussed in the methods part. [a] Disease phenotype in NIL-rr [true-to-type mutant] and NIL-RR [true-to-type wild-type] alleles. The image was taken at 10 dpi after detached leaf assays. [b] Fungal growth on inoculated tissue. Infected tissues were observed under confocal microscope. The fungal hyphae were stained green by WGA-AF488. The dead cells were stained red by propidium iodide. The image was taken from tissues at 6 dpi.

*ARG2* disease phenotype was evaluated under two stress conditions. TAM428 and the *ARG2* NILs that predominantly share the same genomic background did not survive at 38$^{o}$c in a controlled growth chamber. The *ARG2* resistant parent was evaluated for disease phenotype at 38$^{o}$c in the growth chamber (Figure 2-15a). The resistance phenotype was not compromised but the susceptible phenotype in BTx623 was delayed by several days. In addition, wounding by needle did not alter *ARG2* resistance response (Figure 2-15b).
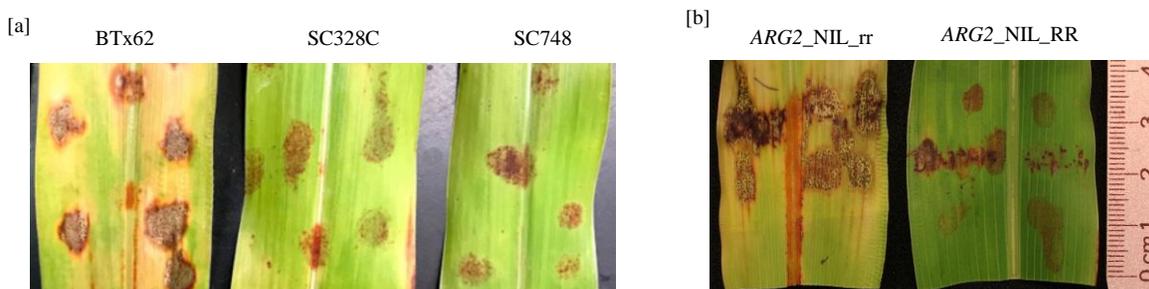


Figure 2-15. Disease responses of sorghum lines carrying different ARG2 alleles at high temperature and after wound-inoculation.
[a] Detached leaves from several sorghum variants were inoculated and kept at 38oc in a growth chamber. BTx623, SC328C and SC748 survived the hot chamber environment whereas the tissue from TAM428 did not survive the hot chamber environment. The widely known resistant line SC748 was used as a control. The appearance of disease symptom was delayed from 7 to 10 days in the lab [21-24oc] and greenhouse to two weeks in this hot chamber. The image was taken at 15 dpi. [b] Leaves from *ARG2* NILs were wounded by needle at many points and the injured spots were inoculated immediately. The mage was taken at 10 dpi.

## 2.3.9 Expression of defense response genes

*ARG2*-mediated expression of eleven defense response genes was examined, the genes were selected based on earlier studies (38, 39). A higher accumulation of flavon O-methyltransferase (*FOMT)* and chalcone synthase 08 (*CHS8)* was associated to the resistance at the earlier timepoints (Figure 2-16 a and b). At the later timepoints, however, their expression level was reversed. *FOMT* and *CHS8* transcript levels dropped in the resistance at the later time points whereas the susceptible NILs showed a higher level of the transcripts with increased pathogen pressure. The accumulation of *PR10* was highly associated to susceptibility at the later timepoints (Figure 2-16c). Heat shock 90, chalcone synthase 5, pathogenesis related thaumatin-like protein, DNAJ heat shock protein, mitogen activated protein kinase, prolyl 4-hydroxylase, glutathione S-transferase 4 and a defensins homolog GEM-like protein 1 (39, 50) did not show differences in gene expression between the NILs in response to pathogen inoculation.

[a]

[b]

[c]
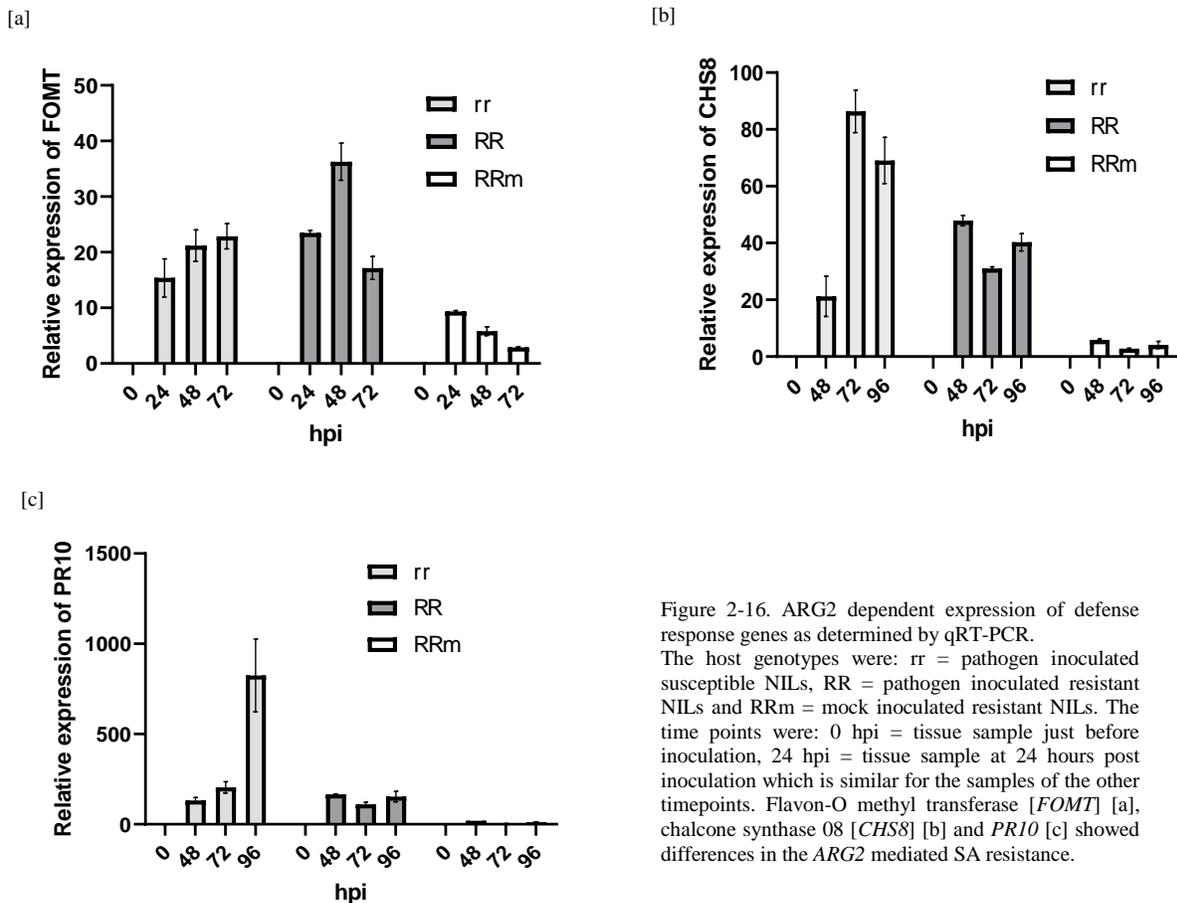
Figure 2-16. ARG2 dependent expression of defense response genes as determined by qRT-PCR.
The host genotypes were: rr = pathogen inoculated susceptible NILs, RR = pathogen inoculated resistant NILs and RRm = mock inoculated resistant NILs. The time points were: 0 hpi = tissue sample just before inoculation, 24 hpi = tissue sample at 24 hours post inoculation which is similar for the samples of the other timepoints. Flavon-O methyl transferase [*FOMT*] [a], chalcone synthase 08 [*CHS8*] [b] and *PR10* [c] showed differences in the *ARG2* mediated SA resistance.

### 2.3.10 Trait association to *ARG2* locus

The shared genomic background in the *ARG2* NILs was leveraged to evaluate the association of *ARG2* locus to plant growth traits. Phenotypic wise, the *ARG2* NILs inherited the architecture of the recurrent parent TAM428, an old improved variety (51). The resistant parental SC328C lacks some major desirable agronomical features. The lower leaves of SC328C turn brownish at the vegetative stage and part of the panicle remains covered in the leaf sheath. Total shoot biomass was measured without pathogen inoculation to evaluate the impact of the *ARG2* gene on plant growth in uninfected plants (52). ANOVA showed the effect of *ARG2* locus on the biomass yield in *ARG2* NILs (Figure 2-17). Total shoot biomass showed significant difference (p<0.002), and panicle biomass showed significant difference (p<0.02). The vegetative shoot biomass was comparable between the true-to-type resistant and susceptible groups whereas the trait value was significantly lower in the heterozygote, p<0.015.
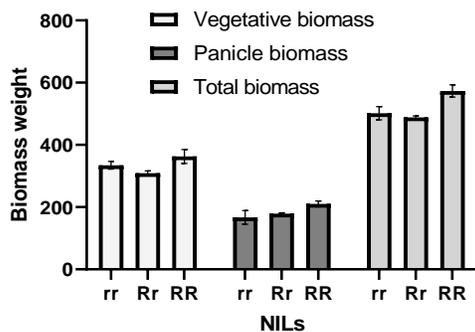


Figure 2-17. Relative biomass yield among ARG2 NILs. The experimental subjects were the true-to-type resistant [RR], the heterozygote [Rr] and the true-to-type resistant *ARG2* NILs. The genotypes were raised from three heterozygote BC5F2 siblings with equal number of plant genotype from each BC5F3 family. Four plants each genotype from each family [a total of 36 plants] were grown in a controlled environment phenotyping facility at Purdue University. Total shoot biomass, vegetative shoot biomass and panicle biomass were measured from dried plants. ANOVA showed panicle biomass and total biomass is higher in the resistant and vegetative biomass is lower in the heterozygote.

### 2.4 Discussions

In this report we document the identification of a sorghum resistance gene and describe its function in disease resistance. *ARG2* was cloned based on the biparental mapping population that was generated using SC283C and TAM428 sorghum lines. Genetic evidence to the identification of *ARG2* is presented based on alleles that confer resistance and alleles that does not confer resistance including an independent mutant allele. *ARG2* that encodes an *NLR* is found in a cluster of three paralogs with the two copies truncated and non-functional in the resistant line SC328C. Interestingly, this race-specific resistance locus appears identical in the sweet sorghum SbRio and SC328C, the latter is dwarf seemingly a high-grain yielding variety. *ARG2* was relatively highly

expressed in the flag leaf and confers resistance even at a high temperature of 38°C. Despite the reported negative impact of resistance genes on plant growth (52), *ARG2* locus appears to contribute to better growth as observed in the NILs.

SC328C was one of the differential cultivars used in the US for race identifications (18, 45). However, the resistance in this cultivar was not mapped previously. We did not find any previous report of anthracnose resistance that is associated to the *ARG2* locus. A report on SA host-resistance linkage map that was developed based on a biparental mapping population between BTx623 and the broad-spectrum resistant SC748-5 showed a resistance QTL marker on chromosome 5 at 4.2 Mb (26). The report marks the closest SA resistance locus to the *ARG2* genomic region. Nevertheless, SC748-5 carries the BTx623 *arg2* mutant allele, and hence the marker is less likely to be associated with *ARG2*. This SNP is located about 3 Mb away from *ARG2* locus which is around the tail of the BSA-Seq peak genomic region that we have identified in the current study. In addition, SC748-5 has broad-spectrum of anthracnose disease resistance environments (18, 26, 40).

### 2.4.1   Organization and evolutionary view of *ARG2* locus

The non-random distribution of *NLR* genes and associated structural variation (SV) has been studied extensively. A recent report showed that 97% of sorghum *NLR* genes exist in clusters, which suggest extensive duplication and evolution of the *NLR* gene family (23). Similarly, the proximity of the three *NLR* genes as observed in the wild-type *ARG2* locus also suggest a cluster of duplicate *NLR* genes. A cluster of three *NLR* genes on chromosome 9 was also associated to SA resistance (53). On the other hand, the three duplicate *NLR* genes are interspersed with several other predicted genes as seen in the BTx623 reference genome, which may contradict the idea of gene cluster. The 3000 rice pangenome project has shown a striking match to the case in the *ARG2* locus. SV was enriched in the promoter region like *NLR3* that disrupted its function (Figure 2-6) whereas shorter SV was enriched in the 5′UTR like the small indels in *ARG2* 5′UTR (54) (Figure 2-11). Other evidence also show the frequent occurrence and active role of SV in driving the evolution of the plant genome (55) and stress response genes (54).

SV and the complexity of *NLR* gene clusters extend to copy number variation (CNV). In the *ARG2* locus, the premature stop codon in *NLR2* and the absence of transcript from *NLR3* lessen the complexity to functional genomic dissection in such a cluster of duplicate genes. The recently

released RTx430 *S. bicolor* reference genome shows a cluster of >4 predicted *NLR* genes at the *ARG2* locus. On the other hand, the absence of PCR product from some of the sorghum lines suggest the complete absence of *ARG2* in some sorghum variants. Thus, the real number of the *ARG2* duplicate genes and the associated SV and CNV might be difficult to establish. Comparative studies reveal > 17,000 CNV in sorghum (56) and hundreds of genome specific loci in rice (57). Similarly, a study in *Capsicum* species revealed the abundance and key role of long tandem repeat transposons in *NLR* CNV, which is in most cases lineage-specific (58). A study on the popular pan-genome showed CNV was enriched in traits that are related to resistance to stresses and pathogens (55). Thus, over the wider sorghum pangenome, a solitary-gene view of *ARG2*-mediated resistance regardless of the SV and CNV in the cluster, and their specific sequence similarities and differences may be misleading.

NLR proteins may dimerize to trigger ETI (48, 59). However, the real extent of divergence that prohibits functional features like dimerization has been unclear (60). In contrast to the wild-type *ARG2*, the CC-domain in the BTx623 allele does not reveal the predicted functional module of homodimerization, which is consistent to the disease phenotype and loss of the *ARG2* resistance function. In this regard, the *ARG2* sequence divergence of the LRR-domain is about 19% that may be more than enough to cause the loss the recognition function, which the wild-type allele confers. Similarly, studies showed a high proportion of large effect SNPs in the LRR domain in sorghum (56). Excess amino acid diversity in the LRR domain was reported in *Arabidopsis (61) and* major effect SNP in the TIR-domain was reported in *Nicotiana* (62)*.* On the other hand, although the line BTx623 was susceptible to all the five Cs strains that were evaluated in the current study and to many more strains in an earlier study (18, 26), the presence of a complete ORF imply the likelihood of disease resistance to another pathogen that the *arg2* allele in BTx623 may confer (63).

ARG2 homologs are widespread in plant genomes. However, despite the proximities of maize to sorghum *and C. graminarum* to Cs did not show in ARG2 sequence, no ARG2 homolog was found in maize. In contrast, ARG2 homologs in rice and wheat shared high sequence homology, including a cluster of three homologs like the *ARG2* gene cluster (Figure 2-12). A similar homology was reported between another SA resistance candidate gene and a rice blast resistance gene (29), and the wheat Lr34 resistance gene was reported to confer resistance to SA and sorghum rust diseases (64). ARG2 is the sorghum homolog of RPM1.

Interestingly, the BLASTp search for ARG2 homolog showed *Oryza officinalis* has the closest homolog from among all the monocot genomes available in the NCBI database. The result was consistent to the shared host responses in wild rice and sorghum to Cs (65), which may suggest a functional similarity coupled to structural similarity. Functional convergence was described between Rpg1-b in soybean and RPM1 in *Arabidopsis* to the effector *AvrB* from *Pseudomonas syringae (66, 67)*. Closely related species share the same genetic bases of traits (68). Nevertheless, the shared pathosystem in the wild rice and sorghum is likely conditioned by the nature of the basal immunity, rather than ETI in which ARG2 is presumed to function.

In bilateral sequence comparison using CoGe (Figure 2-12a) and BLASTp, the widespread ARG2 homologs that showed up to 100% sequence coverage. However, the sequence similarities of all the homologs, except the *NLR* genes cluster in the *ARG2* locus, were <60%. Accordingly, the homologs clustered with respect to their taxonomic group in the multilateral analysis of the sequence relatedness among the selected homologs except *Oryza meyeriana and Eragrostis curvula* (Figure 2-12 b and c).

### 2.4.2   Functional view of ARG2

The boom and bust in host-resistance to Cs has been indicated in both parental lines of ARG2 mapping population over the past few decades. Loss of anthracnose resistance in SC328C was reported in Georgia two decades ago (69), from which the highly virulent Cs strain Cs-grg, but avirulent to SC328C, was collected in 2014. Similarly, studies on *Colletotrichum lindemuthianum* and *Phaseolus vulgaris* showed that most resistance specificities are overcome in sympatric situations (70). Regarding the TAM428, old reports showed SA resistance in TAM428 (51) in a race specific manner (17, 71, 72). However, later reports and the evaluation in the current study show that TAM428 is highly susceptible to all tested strains of the pathogen (18). The consistent disease phenotype, between the Cs strains Cs-gl1 and Cs-grg, on the susceptible and the resistant progenies of the *ARG2* mapping population and NILs strongly suggest the resistance to Cs-grg is also conferred by *ARG2*. The current study shows that *ARG2* confers a typical race-specific resistance that is presumed to recognize a Cs-effector (AvrARG2) from Cs-gl1 and Cs-grg to trigger host-defense response.

Genome-wide analysis of *NLR* in *S. bicolor* revealed that *NLR* genes have basal expression level (73), which is similar to the expression feature of *ARG2*. This basal expression may be

sufficient to the ARG2-mediated SA resistance. Most *NLR* genes that mediate rice blast resistance express constitutively (74), and the marginally induced ones were also regarded as constitutive (75).

Early events of host-pathogen interaction are strongly associated with disease resistance. In rice blast, the invasive hyphae proliferates in the first infected rice cells in 12 hpi (76, 77), and the biotrophic phase in SA infection lasts in about 24 hpi (71). Early recognition of pathogen molecules result in an effective disease resistance (78), and early transcriptional events were reported in maize anthracnose resistance (79). In the current study, the marginal increment in the accumulated *ARG2* transcript at 24 hpi (Figure 2-13a) may suggest the induction of *ARG2* expression by the pathogen although the contribution to the resistance was not determined.

Overexpression of *NLR* genes may result in autoimmunity and fitness costs (80). The intron in the 5'UTR and the uORF in the *ARG2* transcript seems to have a conflicting role over the level of *ARG2* gene expression. An intron in the 5'UTR may enhance gene expression (81) by recruiting RNA polymerase II (82) whereas uORF in the 5'UTR may downregulate gene expression (83). In this regard, the highly elevated expression of *ARG2* in the flag leaf that did not manifest the well-known signs of autoimmunity may be a paradox that requires further study (Figure 2-13b). While the benefit of ARG2 in SA resistance is evident, in light of the higher biomass yield, the potential merit of *ARG2* in the absence of the pathogen may imply a tight regulation of the *ARG2* effect (83-87). The *Arabisopsis* RPM1, the ARG2 homologue, was degraded coincident with the onset of the hypersensitive response that may control the extent of cell death and overall resistance response (88). Recent advances show disease resistance can be achieved without yield penalties (89), or even with yield benefits (74).

The activation of *NLR* mediated resistance is often accompanied by changes in expression of a battery of defense genes with diverse functions (37). We assumed that if a defense response gene shows a significantly higher level of transcript in the resistant NIL, this transcript level should remain lower when the resistant NIL is mock inoculated. Accordingly, the higher accumulation of Flavon-O-methyl transferase (*FOMT*) and Chalcone synthase 08 *(CHS8)* in the resistant NIL at the earlier time points may have contributed to ARG2-mediated resistance (Figure 2-16 a and b). Similar results were found in other studies (38, 39, 90). *CHS* plays a role in the flavonoid pathway although CHS8, that accumulates during SA resistance, does not involve in the flavonoid pathway. *CHS8* is regarded as a defense response marker to SA resistance and proposed to be named as

stilbene synthase 1 (90). In contrast to the current study (Figure 2-16c), an early induction of PR10 in SA resistance was reported (38), which might be genotype dependent. The higher accumulation the transcripts in the susceptible NILs at the later timepoints may suggest host-recognition of the elevated stress due to the higher pathogen growth. A metabolic web of phenylpropanoid and flavonoid pathways, in which *FOMT* plays a role, result in a significantly higher accumulation of apigeninidin and luteolinidin in SA resistance as compared to susceptible genotypes (91).

The resistant *ARG2* NIL produces healthy leaves and grows better under disease pressure, and I thought, sorghum cultivars benefit from *ARG2* when there is pathogen pressure. However, SA disease may not occur every cropping season and throughout the cropping season. In the absence of pathogen pressure, the difference in shoot biomass provided insight to the potential effect of *ARG2* locus in the sorghum plants. NLR-mediated resistance may also depend on temperature (92). Plants preferentially activate ETI signaling at relatively lower temperature, whereas the signaling switches to PTI at a higher temperature (37). In the current study, the series of evaluation of SA disease phenotype showed ARG2-mediated SA resistance was robust and did not vary with temperature. The disease assay was carried out in the lab, greenhouse and growth chamber within 20-38°c.

The functional studies provided insight that *ARG2* may contribute to crop improvement for a wide range of sorghum production regimes especially through pyramiding with other genes. A study showed 87% correlation between seedling resistance in the greenhouse and adult plant resistance in the field (93). The stability of ARG2 at high temperature where the major sorghum productions take place and the likely absence of a major effect resistance-tradeoff suggest that ARG2 confers particularly useful form of SA resistance. However, the reduced quantity of shoot biomass in the heterozygote NILs might suggest undesirable interaction among some of the *NLR* duplicate genes in the *ARG2* locus in the heterozygote state. Thus, the effect of *ARG2* in hybrid sorghum may need special attention. In certain plant hybrids, immunity signaling is initiated when immune components interact in the absence of a pathogen trigger (94). The markers used in the current study may not be useful for marker-assisted selection of *ARG2* locus, and special caution may be necessary to design *ARG2* markers based on the DNA sequence relationships that exist among the duplicate *NLR* genes in the sorghum lines under consideration.

# CHAPTER 3.     RACE SPECIFIC ANTHRACNOSE RESISTANCE IN SORGHUM CONFERRED BY A NON-CANONICAL RESISTANCE GENE

**Abstract**

This study was initiated to identify gene(s) in the sorghum genotype IS18760 that conditions resistance to a highly virulent strain of *Colletotrichum sublienolum,* the causal agent of sorghum anthracnose diseases. IS18760 is resistant to Cs strains Cs-27, and Cs-29 but susceptible to Cs-gl1, Cs-gl2 and Cs-grg strains revealing features of race specific resistance. The avirulent strains are from Western Ethiopia from where the resistant sorghum line IS18760 was also originated whereas the virulent Cs strains were collected from US. BSA-seq analysis based on biparental mapping population located a single highly significant genomic region in IS18760 that is associated to the resistance, designated *ANTHRACNOSE RESISTSNCE 3 (ARG3).* ARG3-mediated resistance is characterized by deregulated cell-death response that covers wide area of the infected leaves. Recombination analysis of the progenies of *ARG3* mapping population using molecular markers and the disease phenotypes defined the *ARG3* locus to a candidate genomic region of about 18.6 Kb that carries a tyrosine protein kinase and a predicted gene of unknown function. The likely sequence difference responsible for the disease responses is found in a genomic region that does not carry predicted gene. At this candidate sequence, a transcript was amplified, and the deduced peptide sequence shows functional motifs like ligand receptor, toxin and anti-microbe. This functional sequence-signature does not show any homology in the plant kingdom. The *ARG3* genomic regions shows big structural rearrangements and is associated to patches of highly repetitive DNA sequences. Interestingly, the unique likely *ARG3* gene is located within a syntenic region conserved between rice and sorghum with striking conservation of gene order. The characterization of *ARG3* and its functional features will be determined in subsequent studies.

## 3.1    Introduction

In plants, different categories of disease resistance genes have been characterized. Resistance genes that are inherited as a dominant allele are common (215) although some recessive resistance genes are known to play key roles in crop improvement (284). Genes that confer major

effect disease resistance have been utilized widely relative to minor effect loci. Genes that encode for NB-LRR (NLR) and pattern-recognition receptor kinases and receptor-like proteins (PRR) predominate the known disease resistance genes (285, 286). Genes that confer resistance to some races of pathogens, a broad spectrum of pathogen races and multiple pathogenic species brought attention from the practical side of agricultural crops (215, 220). The fragility of most disease resistance genes necessitated the attention on durable disease resistance. In most crop plants, comparative and empirical evidence show the abundance of *NLR* genes that frequently confer race-specific and time bound disease resistance (69, 215).

Although disease resistance to specific races of pathogens is usually conferred by *NLR* genes (215), some non-*NLR* genes that confer race-specific resistance also have been studied. The first cloned plant disease resistance gene *Hm1* encodes for the toxin reductase in maize that inactivates the race specific HC-toxin from the pathogen *Cochliobolus carbonum* (287). The race-specific resistance to the bacterial leaf spot disease of pepper was specified by the promotor sequence of the non-*NLR* gene *Bs3* (288). In this pathosystem, the pathogen effector binds to and activates the promotor of the cognate resistance gene, and no other stimulus was found to activate the gene. In contrast to *Bs3*, the *Arabidopsis* FMO1 that shares homology to Bs3 is induced by a variety of stimuli and confer a broad-spectrum disease resistance. The durable barley stem rust resistant gene, *Rpg1,* that encodes a receptor kinase-like protein exhibits race-specificity (289) although no *NLR* gene has been reported to be functionally association with it. Similarly, a race-specific resistance gene to bean anthracnose disease is mapped to a genomic region that does not carry an *NLR* gene (290).

In biotrophic and hemi-biotrophic pathogens, the independence of non-NLR race-specific resistance from NLR-mediated resistance may require in depth molecular biology study. One of the most characterized race-specific disease resistance genes, *Pto,* carries a serine/threonine kinase domain (159). A follow up study unraveled that *Pto* forms a complex with the NLR protein Prf to mediate resistance to the bacterial speck disease of tomato (291). The eLRR-type plant proteins usually interact with some PRR proteins to confer disease resistance, some of them in the typical gene-for-gene kind of interaction as in the case of tomato resistance to *Cladosporium fulvum* (292). NLR and eLRR proteins share the LRR domain, which plays key role in the pathogen recognition, although the mechanism by which they trigger resistance response may vary. In nutshell, the

dependance of disease resistance from NLR gene families seem far too broad (137), and so to the non-*NLR* genes that confer race-specific resistance.

This study was initiated to identify sorghum anthracnose disease (SA) resistance genes and associated immune responses in the sorghum line IS18760. This line carries a major resistance locus, *ARG3*, against the Cs strain that are virulent to *ARG2* mediate resistance (as discussed in chapter two). The molecular and genetic bases of resistance in IS18760 is different from those provided by*ARG2* gene. In the current study, we report a map based cloning and partial characterization of *ANTHRACNOSE RESISTANCE GENE 3* (*ARG3*), which encodes a protein of unknown function that appears to confer race-specific resistance to Cs. Based on analysis of a partial cDNA sequence, *ARG3* encodes for a novel protein. The resistant and the susceptible siblings that share a similar background genome are developed for characterization of ARG3-mediated defense-response.

## 3.2   **Materials and methods**

Sorghum variants were evaluated for resistance to five Cs strains as shown in the *ARG2* study. A biparental mapping population was developed by crossing IS18760 to the susceptible line TAM428. The progenies were evaluated and selected for disease phenotype up to F5 generation using single seed descent method, and BSA-seq was carried out using genomic sequences of the resistant and the susceptible pooled-progenies at F5.

Fungal culture, inoculum preparation, disease assays and the evaluation of disease phenotype was carried out using the methods, conditions and resources that were described in Chapter 2. The tissue sample preparation and processing for the BSA-Seq analysis was also similar. The resistant parent, the resistant progeny-pool and the susceptible progeny-pool tissues whole genome samples were re-sequenced using Illumina NovaSeq 6000 platform with about 25x coverage as opposed to the illumina HiSeq2000 for *ARG2* study. The genome sequence of the susceptible parent TAM428 was available from previous studies. The other methods in *ARG3* study that were distinct from *ARG2* study (chapter 2) are described below.

### 3.2.1 ARG3 mapping

BSA-Seq analysis of the genomic raw-reads in the two pooled-progeny samples and in the two parental lines was used to map the genomic region of *ARG3* using the QTL-Seq analytical pipeline. The assumptions and analytical considerations that were employed to develop DNA size-markers and to carry out the entire recombination analyses during *ARG2* mapping were instrumental to mapping *ARG3* as well. However, based on the lessons from *ARG2* study and the specific features of *ARG3*, the recombination analysis and the downstream mapping had distinct features.

The recombination analysis was followed by genome assembly of the parental lines and many other resequenced sorghum variants. *De novo* genome assembly of the genomic raw-reads of the *ARG3* parental lines and pooled progenies was carried out using the conservative genome assembler SPAdes 3.14.1 (293). The assembled genomes were examined using Bandage software (294). The assembly was carried out primarily to resolve the big indel in the *ARG3* locus, which was expanded later to many other sorghum variants including the publicly available sorghum genomic and transcript raw data to generate additional data. These sorghum variants were selected based on the disease phenotype to enable the association of sequence variation to *ARG3* disease resistance function.

The detail of sequence variation and associated functional features were also examined in the parental lines, reference genomes and other sorghum lines. Gene expression and functional domain analysis were carried out to further resolve the molecular basis of the *ARG3* disease phenotype. The comparative analysis of *ARG3* locus was further expanded to other sorghum genetic resources and related organisms using SNPs, CoGe and BLASTp platforms (295).

## 3.3 Results

IS18760 showed strong resistance to the Cs strains Cs-27 and Cs-29 but susceptible to the other three strains (Table 2.1). The fungal strains Cs-27 and Cs-29 were collected from Western Ethiopia. Cs-29 exhibits stable characteristics on a growth media and usually sporulates profusely producing large amount of conidia whereas Cs-27 did not provide consistent growth characteristics on the media and frequently failed to produce enough conidia for inoculation purpose. Thus, Cs-29 was used to map the resistance in IS18760. As opposed to the resistant parent, IS18760, the

resistant progenies of *ARG3* mapping population show inconsistent disease phenotype between Cs-27 and Cs-29.

### 3.3.1 Genetic analysis

The progenies of TAM428 and IS18760 cross were analyzed to determine the inheritance of the resistance in IS18760 to Cs-29 that was designated as *ARG3*. In contrast to the robust resistance in IS18760 (Figures 3-1a, 3-1b, 3-2b), the F1 progenies were affected by the pathogen to a considerable extent though the SA symptom that typically shows the pathogen fruiting bodies did not appear (Figure 3-1c). The disease phenotypes at F2 were clustered as robust resistance, widespread cell death response that does not support pathogen growth (referred to as unusual disease phenotype) and susceptibility with the proliferous pathogen growth that develops to the typical disease symptom. The disease phenotype that was scored as resistant includes the robust resistance response as in IS18760 (Figure 3-1d) restricted cell death. A large proportion of F2 plants showed the cell-death response that covered wide area of inoculated leaves. Nonetheless, the disease phenotype did not show continuous variation.
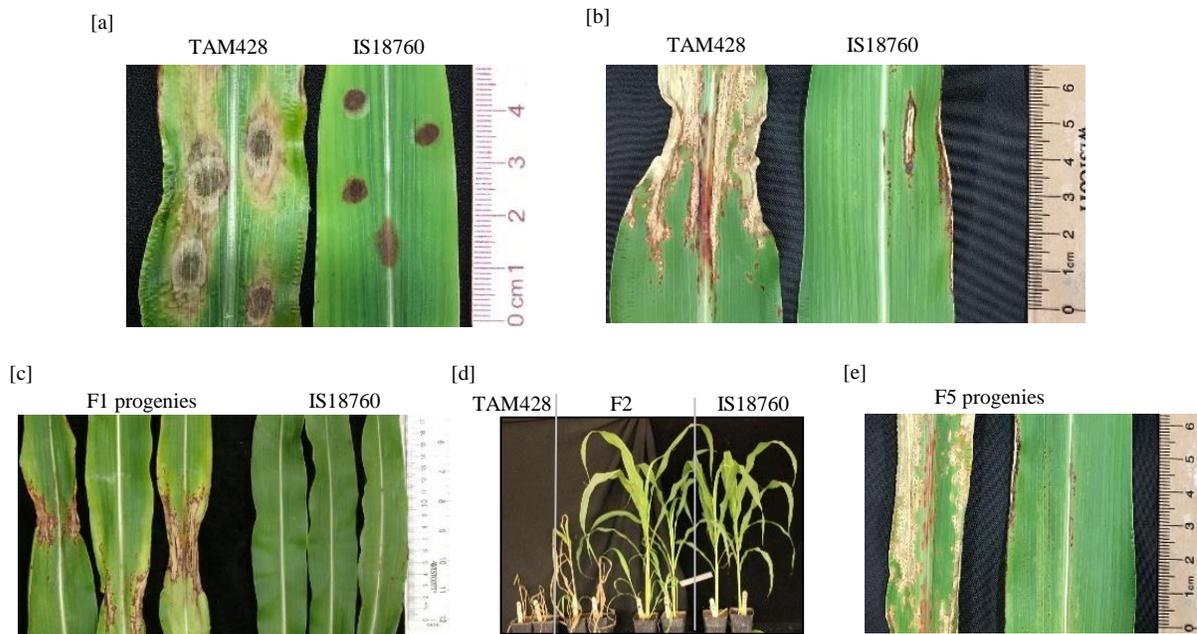


Figure 3-1. *ARG3* disease phenotype.
*ARG3* disease phenotype [a] using detached leaf disease assay and [b] using spray inoculated plants. [c] Disease phenotype of F1 progenies as compared to the resistant parent, IS18760. [d] Extreme F2 progeny disease phenotypes that resemble that of the parental lines as compared to the parental line disease phenotypes [whole plant view]. [e] typical F5 resistant and susceptible disease phenotypes that were used for BSA-Seq. All images were taken at 10 dpi.

Although many F2 plants showed the resistant parent disease phenotype (Figure 3-1d), the proportion of disease phenotypes at the initial evaluation did not result in a meaningful statistics that is determine the genetics of *ARG3* resistance primarily due to the incomplete understanding the specific nature of ARG3 disease phenotype. Thus, the phenotypic evaluation was prolonged over several generations although the phenotypic variation of progeny families at F3 and F4 end up unsuccessfully for a similar reason. Interestingly, the robust resistance disease phenotype in some F2 plants was lost over generations and some other descendants showed more of the unusual disease phenotype. However, after the genomic region that harbors the resistance gene was identified, which is discussed in the next section, correlation analysis of the disease phenotypes to the *ARG3* locus was carried out to determine the host-response especially in the heterozygote state. Multiple F5 progeny-families, in which the disease phenotype segregates, clearly showed that the expanding cell death response is a manifestation of the *ARG3* locus effect both in the homozygote resistant and heterozygote states. This phenotypic distinction between the resistance and the
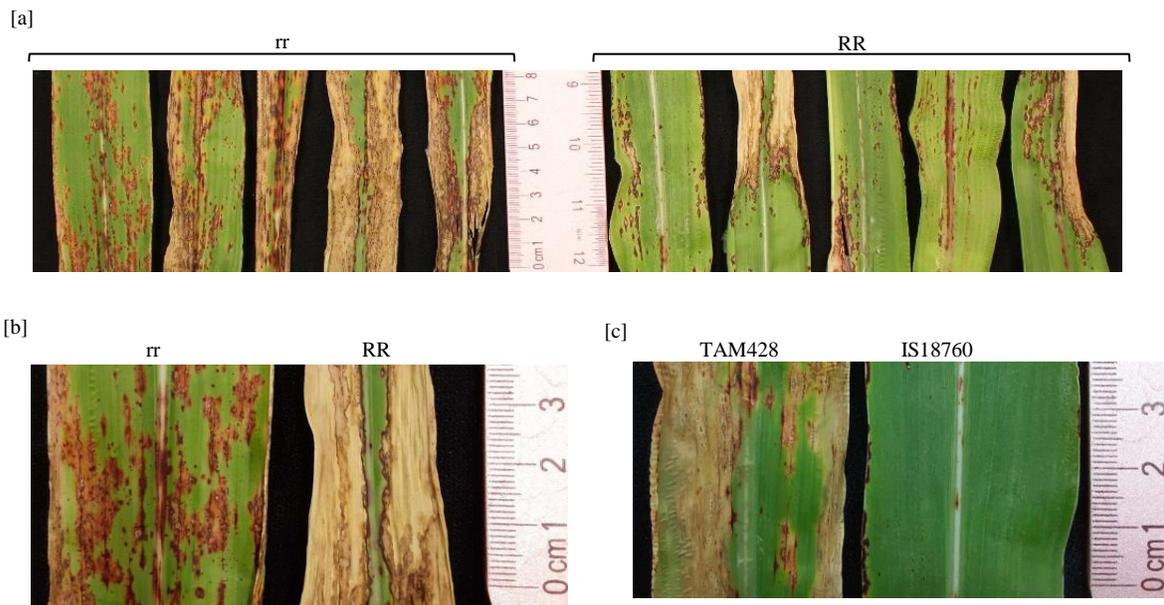


Figure 3-2. Disease phenotype in multiple families of plants that shared a common genomic background show the deregulated ARG3 resistance response.
[a] Disease phenotype in five true-to-type susceptible and five true-to-type resistant F7 families of plants that descend from a single heterozygote ARG3F5 plant. [b] Magnified view of the typical sorghum anthracnose symptom on the mutant *arg3* group [rr genotype] and the deregulated *ARG3* resistance response [RR genotype]. [c] Parental disease phenotype in the same disease assay. Each F7 family is represented by one leaf. The images were taken at 15 dpi.

susceptible *ARG3* genotypes is shown in F7 true-to-type families that descend from a single heterozygote F5 progeny (Figure 3-2 a and b).

### 3.3.2 BSA-Seq to map *ARG3*

The unusual disease phenotype that is associated with the resistance, as discussed above, necessitated the evaluation and advancement of more resistant F2 plants up to the F5 generation. In contrast, the number of clearly symptomatic/susceptible progenies was sufficient to proceed with BSA-Seq at F2. BSA-Seq analysis was conducted on progenies of F5 plants that showed the susceptible and the resistant disease phenotypes (Figure 3-1e).

BSA-Seq revealed a single genomic region that is significantly associated (*p<0.01*) to the disease phenotype between the resistant and the susceptible pooled-progenies of the *ARG3* mapping population. *ARG3* was located towards the distal end of chromosome 4 (Figure 3-3a and 3-4a), and the BSA-Seq results based on the two Mb and the four Mb sliding windows were similar. Thus, I preferred to describe the BSA-Seq statistics based on the smaller (two Mb) sliding window. The chromosomal coordinates of the significant ΔSNP-index estimates in the BTx623 reference genome span 52.50–61.85 Mb at the 95% confidence interval (CI) and 53.50–60.20 Mb at the 99% CI. The significant ΔSNP-index estimates range from 0.31–0.64, with the lowest threshold estimate of 0.30 at 95% CI and 0.39 at 99% CI. In the wider significant *ARG3* genomic region (the 95% CI), the smaller depth and the SNP count range from 32–40 and 1171–2904 with mean values of 37 and 1922, respectively. Thus, the *ARG3* genomic region may span about 9.6 Mb around the peak. The BTx623 reference genome was preferred for *ARG3* mapping for the clarity of the mapped raw-reads in the *ARG3* genomic region and showed sequence similarity to TAM428.

The BSA-Seq analytical pipeline leverages either of the parental genomes as consensus sequence to generate the allelic frequencies in the pooled-progenies genomic sequences. Subsequently, the difference in allelic frequencies between the resistance and the susceptible poled-progeny samples (ΔSNP-index) is used to show peak genomic regions that are associated to the disease phenotype. When BSA-Seq result in a marginally significant peak with respect to both parental lines, as a consensus sequence for the estimation of allelic frequencies (Figure 20 b and c), I referred to this peak genomic region as a predictable smaller effect candidate locus. The peak on chromosome 4 is the only highly significant genomic region for the disease phenotype between the pooled-progeny samples. Nevertheless, regardless of the statistical significance, a consistent peak genomic region was also shown on chromosome 10 which might be influenced by the sample plant identification that is discussed latter (Figure 3-3c). This second peak harbors a BTB domain

gene that plays role in HR suppression and carries a likely high impact indel between the parental lines.

[a]

[b]



Figure 3-3. BSA-Seq plot of ARG3 mapping.
[a] BSA-Seq plot on the ten chromosomes as depicted in the *S. bicolor* reference genome. [b and c] Magnified view of the map on chromosome 04 with respect to the resistant parent[upper] and the susceptible parent [lower]. The x-axis refers to the genomic coordinates along the chromosome, and the y-axis refers to the ΔSNP-index estimates. The vertical positions of ΔSNP-indices with respect to the genomic window were represented by the blue dots, most of them shown as overlapped. The red lines along the chromosome represent the regression plot of the ΔSNP-index estimates. The green lines show the significance threshold [*p=0.05*] of ΔSNP-index estimates along the genomic windows and the pink at p=0.01. The vertical rectangular bar [red] shows peak genomic regions that are associated to *ARG3* resistance.

[c]

[c]

### 3.3.3   Recombination analysis

I thought undertaking recombination analysis in the very large (>9 Mb) significant *ARG3* genomic region may not be effective. As compared to *ARG2* mapping, we assumed that a satisfactory recombination analyses can be made in a smaller genomic interval by exploiting the higher chance of recombination up to the F5 generation in a larger number of progenies. Accordingly, about 3.8 Mb *ARG3* genomic region was used around the maximum ΔSNP-index

estimates to establish the framework of *ARG3* fine-map. The choice of this 3.8 Mb interval was also influenced by the availability of evenly distributed DNA-size markers.

A series of recombination analyses was carried out in 100 F5 progenies using four rounds of marker design, a total of 16 DNA size-markers and other few presence/absence markers. The markers in the first round of recombination analysis were dispersed throughout the 3.8 Mb region and the adjacent markers were several hundred kill-base apart from each other. The markers in the subsequent analyses were fewer in number and progressively closer and closer to each other as per the likely direction of *ARG3* locus that the earlier analyses suggested. The F5 progenies showed many recombination events around the outer marker positions (Figure 3-4b). An exhaustive analysis including presence/absence markers, not included in figure 2, were carried out in the later rounds of the recombination analyses. Nonetheless, no recombinant progeny was found in the middle 0.9 Mb genomic region (54.19 Mb to 55.09 Mb).

On the right side of this 0.9 Mb genomic region, two recombinant progenies showed interesting genotypic feature within a 31 Kb interval (55.09 Mb to 55.12 Mb) (Figures 3-4 b – e). One of these two recombinants (recombinant 1) was homozygote whereas the other (recombinant 2) was heterozygote. The disease phenotype in the subsequent generations revealed that recombinant 1 was true-to-type resistant whereas recombinant 2 did carry both alleles that condition both the resistance and the susceptible phenotypes. The two recombinants carry similar genotypes at all other marker positions (Figures 3-4 c and d), and hence the two markers that are adjacent to the no-recombination genomic region were used to illustrate the final recombination-map based *ARG3* locus (Figure 3-4d). The true-to-type resistant recombinant 1 that show both rr and RR genotypes at 55.09 Mb suggest the position of *ARG3* locus is to the right of this marker whereas both the susceptible and the resistant progenies of recombinant 2 carry the RR genotype at 55.12 Mb that suggests the position of *ARG3* locus to the left of this marker. Thus, the genomic interval between the two markers carries *ARG3* locus.

The entire marker analyses resulted in a bidirectional gradient of the parental allele frequency with respect to the disease phenotypes that pointed to the *ARG3* locus. The subsequent specific linkage analysis in recombinant 1 and recombinant 2 plants showed that *ARG3* is found within the 31 Kb genomic region (Figures 3-4d and e). This genomic region was flanked by the markers at 55.09 Mb to 55.12 Mb coordinates on chromosome 4 in the BTx623 reference genome.

### 3.3.4   Structural variation in *ARG3* locus

Genome analysis showed large structural differences in the 31 Kb *ARG3* locus (Figure 3-4e, Supplemental file 3). The IGV view of the BAM-files of the susceptible pooled-progeny and TAM428 showed the presence of aligned raw-reads throughout *ARG3* locus. However, in the resistant pooled-progeny and the IS18760 samples, the right side of the *ARG3* locus showed the absence of aligned raw-reads. We wanted to generate more evidence about this apparently big indel, that may be difficult to determine using a PCR size-marker. *De novo* genome assembly of both the susceptible parent and other susceptible lines including all the *S. bicolor* reference genomes resulted in two nodes (separately assembled DNA sequences) that carry the conserved sequences from the opposite sides of the big indel. Neither of these two nodes carry the majority of the indel sequence, rather only small proportions of the sequence at the indel junctions were assembled to one or the other node. In contrast, the resistant pooled-progeny and IS18760 genomes result in single, and by far longer, nodes (94,174 bp and 93,774 bp, respectively) that span the flanking sequences of *ARG3* locus (Supplemental file 3). The summary statistics of the assembly of IS18760 genome includes N50 = 12,941 bp and median depth = 16.1x.

Interestingly, the genome assembly provided additional information about the absence of the big indel in the resistant group that was observed on IGV providing a strong evidence to the absence of the big indel in IS18760. This big indel carries a 12,403 bp highly repetitive sequence. Thus, the *ARG3* locus in the resistant parent spans about 18.6 Kb genomic region. However, the attempt to amplify DNA segment from the resistant parent and progenies that flanks the 12,403 bp indel, which is very big to amplify from the susceptible genotypes was not successful. Dense and widespread bands (gel electrophoresis bands) appeared in both the resistant and susceptible groups which complicated even the search for a presence-absence band. The genome analyses showed the big indel that carries the 12,403 bp repetitive sequence is present in the susceptible group including TAM428 and the BTx623, BTx642 and RTx430 reference genomes. The fourth reference genome SbRio show the big indel in two tandem repeats plus about 40 Kb unknown sequence (Figure 3-4e). The presence of this big indel, like in TAM428, was seen in many other sorghum variants, and I did not find a sorghum variant other than IS18760 that does not carry a large indel.
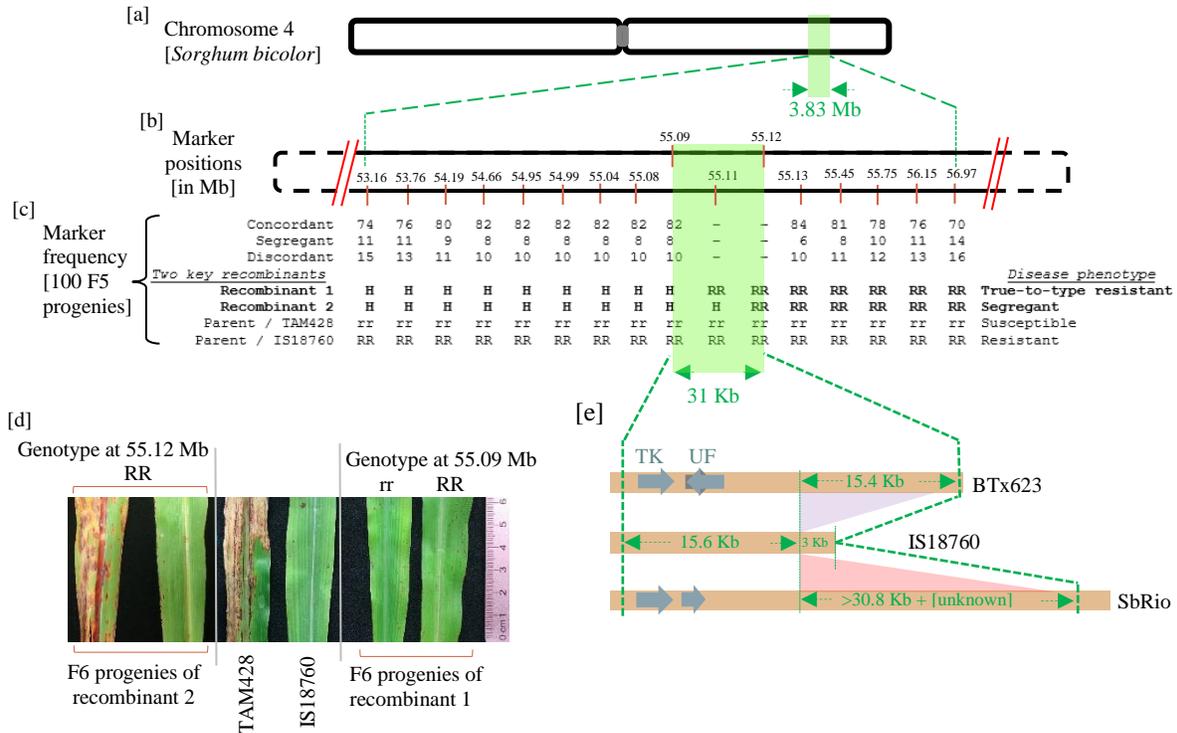
Figure 3-4. A diagrammatic summary of ARG3 fine-mapping.
[a] The chromosomal position of *ARG3* locus as identified by the BSA-Seq. [b] A magnified view of 3.83 Mb around the *ARG3* BSA-Seq peak. Recombination analysis was carried out in this 3.83 Mb region using DNA size markers and 100 F5 progenies. The row of numbers in this panel shows position of the size-markers around the *ARG3* genomic region. [c] The frequency of the concordant and the discordant true-to-type markers, and the heterozygote markers are placed beneath the respective markers. The green shaded 31 Kb segment that is located between the markers 55.09 and 55.12 was the genomic region of *ARG3* that was determined through the recombination analysis. [d] Disease phenotype of *ARG3* parental lines and F6 progenies of recombinant 1 and 2, the picture was taken at 10 dpi. TAM428 is the susceptible parent that carries the recessive allele at all the markers whereas IS18760 is the resistant parent that carries the resistant allele at all the markers. The true-to-type resistant progenies of recombinant 1 [d1 and d2] showed both rr and RR genotypes at 55.09 Mb suggesting the position of *ARG3* locus to the right of this marker. Both susceptible and resistant progenies of recombinant 2 [d4 and d5] carry the RR genotype at 55.12 Mb suggesting the position of *ARG3* locus to the left of this marker. [e] The recombination analysis strongly suggested that the 31 Kb interval in TAM428 and the BTx623 reference genome harbors *ARG3* locus. The interval carries three predicted genes: namely Tyrosine kinase [TK] and two overlapping predicted genes of unknown function [UF]. One of these two overlapping genes was not predicted in two of the four *S. bicolor* reference genomes. A 12,403 bp repetitive sequence of the 31 Kb *ARG3* locus is absent in IS18760 as determined by genome assembly. The 31 Kb *ARG3* genomic region was reduced to 18.6 Kb [15.6 kb + 3 Kb] in the resistant line IS18760. The SbRio reference genome shows two tandem repeats of the indel plus about 40 Kb unknown sequence. The light purple and the light pink shaded triangular projections show the stretches of the repetitive sequence from the indel position in IS18760.

### 3.3.5   Identification of *ARG3*

Sequence comparison and transcript and functional domain analyses between the resistant and the susceptible parental lines does not any predicted gene to be a strong candidate *ARG3*. The18.6 Kb *ARG3* locus in IS18760 carries three predicted genes in the BTx623 reference genome (Figure 3-5b), which carries identical sequence to TAM428 throughout the *ARG3* locus. One of these three predicted genes, Sobic.004G200400, contains a putative tyrosine kinase domain that carries only

two low impact SNPs in one of the exons that is outside of tyrosine kinase domain. The SNPs in TAM428 changed the threonine codon ACG to the methionine codon and the proline codon CCT to a serine codon TCT. Nevertheless, the susceptible lines RTx430 and SC328C (the *ARG2* parental line) carry identical sequence to IS18760 throughout the tyrosine predicted gene and the upstream potential regulatory sequences that reduces the likelihood of Sobic.004G200400 as a candidate *ARG3*. *ARG2* does not confer resistance to the Cs-29 strain to which *ARG3* confers resistance. The small exons of the other two predicted genes, Sobic.004G200501 and Sobic.004G200600, show neither a difference in sequence between the parental lines nor carry a known functional domain. In addition, no transcript was detected from these two predicted exon sequences before and after inoculation (Figure 3-5a).

Sequence analysis based on primer walking provided an interesting clue to the likely *ARG3* gene. We amplified and sequenced a segment of a transcript in the 3.2 Kb region that does not carry a predicted gene in the reference genomes (Figure 3-5). This region carries several high impact SNPs and indels between the parental lines (Figure 3-5f and supplemental file 4), and the amplified transcript segment has 567 bp ORF in IS18760 whereas the reading frame in TAM428 carries multiple stop codons (Figures 3-5F). The sorghum lines of all the four *S. bicolor* reference genomes are susceptible. The candidate genomic region in the BTx623, BTx642 and SbRio reference genomes carry a similar DNA sequence to TAM428 whereas the RTx430 reference genome shows a similar ORF to IS18760. Nevertheless, several high impact SNPs and structural differences exist upstream and downstream of the amplified ORF in RTx430 that suggest the possibility of loss of *ARG3* function (Figure 3-5 d and g). Thus, the candidate *ARG3* in RTx430 has the possibility to be an independent mutant allele. Database searches and sequence comparisons with deduced protein sequence from the amplified cDNA shows functional motifs like ligand receptor, toxin and antimicrobe with multiple transmembrane domains.
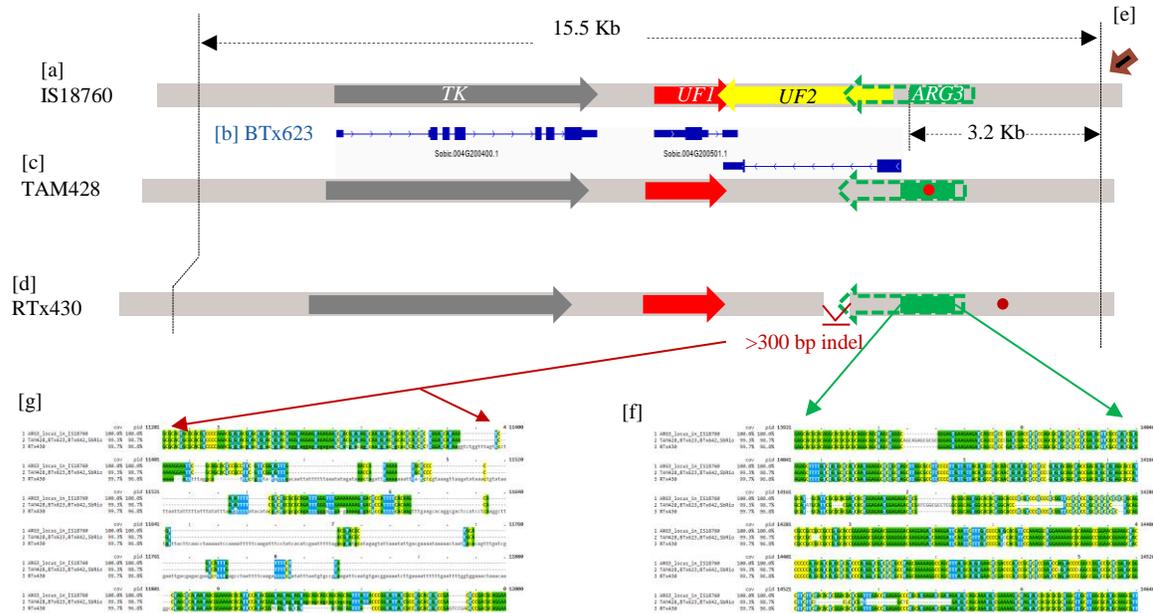
Figure 3-5. The proposed wild-type ARG3 locus and variant mutant alleles.
[a] Genomic organization of *ARG3* locus in IS18760. UF1 and UF2 [the red and the yellow predicted genes] may not exist. [b] Below the upper panel is a snapshot of predicted transcripts in the BTx623 reference genome: the lines with arrows show intron segments and the arrows show orientation of genes; the thicker rectangular segments refer to exon and the thinner ones show UTR. [c] TAM428 and the three reference genomes [BTx623, BTx642 and SbRio] that correspond to *ARG3* locus in IS18760 and [d] the RTx430 reference genome. [e] *ARG3* DNA sequence comparison in IS1876, TAM428, and RTx430. The short arrow in the right side shows the left margin of the big indel in TAM428 and all the reference genomes. The gene with tyrosine kinase domain [TK] does not carry high impact SNPs between the parental lines. RTx430, SbRio and the resistant IS18760 share identical coding sequences and transcript splice junctions. UF1 shows no known functional domain but contains a 312 bp single exon. UF2 that partially overlaps with UF1 carries two exons [457 bp and 35 bp] and no predicted 5'UTR. UF1 and UF2 do not carry sequence difference between the parental lines. The green arrow shows the proposed *ARG3* gene. [f] The *ARG3* transcript [green bar] was found in the 3.2 Kb region as confirmed by PCR and sequencing. The transcript carries 567 bp ORF in IS18760 whereas the susceptible parent, BTx623, BTx642 and SbRio carry high impact indels. [g] RTx430 carried high impact sequence rearrangement downstream of the amplified ARG3 transcript that may disrupt the downstream sequence of *ARG3* transcript. The sequence in RTx430 upstream of *ARG3* transcript also carries multiple indel [deep-red dot]. All lines except IS18760 are susceptible to the strain Cs-29 to which IS18760 confers resistance. Thus, only IS18760 carries the wild-type *ARG3* allele. [c] TAM428 and the three reference genomes [BTx623, BTx642 and SbRio] carry mutant *ARG3* alleles as shown by the red dot [stop codon] in the amplified ARG3 transcript [d] RTx430 carries similar sequence to IS18760 in the amplified *ARG3* transcript although high potential independent premature stop codons are available at both sides [including the position of the light blue dot].

The comparative analysis was expanded beyond the *S. bicolor* species. CoGe and BLASTp analysis of the candidate *ARG3* genomic region did not show any homologous gene in the plant kingdom. Nevertheless, a strong synteny of genes around *ARG3* locus was found in the rice genome (Figure 3-6). The synteny has a row of seven predicted genes that spans the *ARG3* genomic region: six genes to the left and one gene to the right. The sequence in *O. sativa* that corresponds to the *ARG3* locus carries widespread repetitive sequences that does not show any predicted genes.
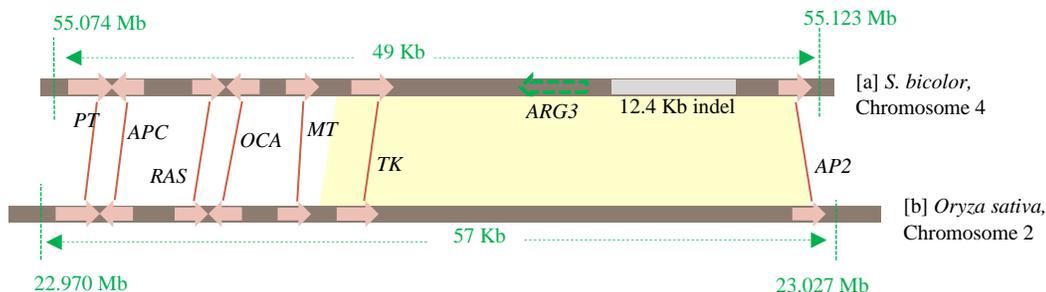
Figure 3-6. Sorghum and rice (Oryza sativa) share a big and highly syntenic genomic region around ARG3 locus.

The synteny has a row of seven predicted-genes: Phosphate transporter family protein [PT], anaphase promoting complex subunit 5 [APC], Leucine rich repeat containing RAS suppressor protein [RAS], Ovarian carcinoma antigen related [OCA] that contains PB1 domain, zinc figure – zz type and lg-like domain from next to BRCA1 gene, 25S rRNA (adenine(645)-N(1))-methyltransferase / 25S rRNA m(1)A(645) methyltransferase [MT], Protein kinase ATN1 - Tyrosine kinase domain [TK] and AP2 domain [AP2] – family not named. The arrows show the position and orientation of the genes and the lines [blue] connect homologous genes. The pink shaded region spans the ARG3 genomic region at which some predicted gene regions, signatures of transcripts and the major repetitive elements are found in the S. bicolor genome. The region in O. sativa that corresponds to the ARG3 genomic region carries a widespread repetitive element and no predicted genes are available.

## 3.4    Discussions

I identified an anthracnose resistance gene, designated as *ARG3,* that likely encodes a non-canonical resistance protein. The *ARG3* locus undergone a large structural rearrangement with patches of repetitive sequence including a big transposon-like indel that is absent in the resistant parent. *ARG3* confers race-specific major effect resistance to the Cs strain Cs-29 that is virulent to the widely known broad-spectrum SA resistant sorghum line SC748-5 (186) (Table 2-1). The resistance is inherited as a dominant allele that shows a deregulated cell death response in some resistant progenies. ARG3 does not have an identified homolog in the plant kingdom. The recently cloned unique gene in wheat, FHB7, that confers resistance to *Fusarium* head blight is thought to be due to a horizontal gene transfer from fungus (296).

The robust resistance response in IS18760 seems a hypersensitivity response that initiates at the site of attempted infection. However, in a large proportion of the resistant progenies, this cell death response expands and covers a large area of the leaves. There was no proliferous pathogen growth in plants that showed this deregulated resistance response. Thus, *ARG3* is required to restrict the growth of this aggressive pathogen strain suggesting a major gene effect. The inconsistent disease phenotype among the progenies of the *ARG3* mapping population to Cs-

27 and Cs-29 may suggest the presence of another locus that conditions the resistance to Cs-27. Similarly, a previous study on IS18760 for SA resistance showed multiple resistance QTL (297).

The nature of the disease phenotype and the BSA-Seq results, along with the ambiguous genetic data, suggest the presence of another locus that contributes to the robust resistant response as in IS18760 and some resistant progenies. Two likely scenarios may happen in case of a digenic unlinked resistance loci to the disease phenotypes in the *ARG3* mapping population. If a hypersensitive response is not triggered, as in the susceptible progenies, penetrance of the contribution of the second locus to the resistance phenotype may become very low. Alternatively, if the role of the second locus is in the regulation of the hypersensitive response, the phenotypic contribution of this locus in the susceptible progenies is expected to be nullified. Thus, the alleles at the second proposed locus tend to segregate randomly especially in the susceptible pooled-progeny sample and drag down the ΔSNP-index estimates (Figure 3-3c). The glimpse of a related phenotypic effect was observed during the progeny evaluation of the resistant group, as some of the F2 and F3 plants that showed the robust resistance like IS18760 were compromised in the subsequent generations. I propose a second genomic region (Figure 3-3c) that contributes to the disease phenotype.

The bidirectional gradient of recombinant allele frequencies with respect to the disease phenotype (Figure 3-4c) that was independent of the BSA-Seq confirmed the *ARG3* genomic region. Interestingly, the recombination analysis resulted in perhaps one of the most successful results of this kind. The >9 Mb significant *ARG3* genomic region was drastically reduced to about 18.6 kb, which harbors a few predicted genes (Figures 3-4 and 3-5).

### 3.4.1   Variants of *ARG3* locus

The *ARG3* locus that carries the big indel seems abundant among the sorghum germplasm whereas I did not find another line like to IS18760 that does not carry the big indel. The big indel that was predicted to carry transposable elements may be associated beyond physical linkage to the loss of the resistance function (256). Structural variation is known to drive the evolution of plant genome and stress related genes (201, 203, 256). The complexity of this repetitive sequence may be partially explained by the absence of any *de novo* assembled node that carried a larger proportion of the big indel sequence (298). The widespread interspersed repetitive sequences around the big indel may have complicated the PCR amplification of a specific genomic sequence

in IS18760, which flanks the big indel in TAM428 that is very long and difficult to amplify. On the other hand, the unusually large sequence rearrangement in the amplified candidate *ARG3* transcript sequence, between TAM428 and IS18760, suggests a rare or newly born gene.

### 3.4.2 *ARG3* reveals a co-evolution of resistance to local pathogen races

The *ARG3* resistant line IS18760 traces back to Sudan (186), and the Cs pathogen strain Cs-29 to which *ARG3* confers resistance was collected from Western Ethiopia. It is worth noting here that IS18760, which carries the resistant *ARG3* allele is susceptible to at least three Cs strains from the US, and vice versa, to the pathogen strains from Ethiopia to many sorghum lines in US (Table 2-1). This region has been known as the center of origin of *S. bicolor*. Analysis of the sporadic SNP data in the Ethiopian sorghum core collection, which was generated using GBS (genotyping by sequencing) showed many accessions that share SNPs with IS18760 in the amplified *ARG3* transcript segment. An in-depth assessment of the DNA sequence throughout the proposed *ARG3* gene region and potential regulatory elements was possible only in re-sequenced lines, and assessment in >20 re-sequenced sorghum lines showed that only IS18760 carries the 18.6 Kb *ARG3* genomic region.

The absence of a homologous gene in related species also supports the case that *ARG3* is either a rare variant or a newly born gene. Studies on the Ethiopian core germplasm collections showed the abundance of rare variants, in the range of 46 - 60% SNP (295, 299). Consistent with the proposed transcript signature-sequence of *ARG3*, studies show the presence of many small ORF transcripts that exhibit translation signature and become a source of new proteins, some of them may confer selective advantage and be maintained over time (300).

The current study established a complete map of *ARG3* to clone characterize the gene and the resistance responses. The nature of the gene network that mediates the resistance in IS18760 may provide insight to further study the biochemistry and metabolome profile. The current progress provides an insight to trace back and fix the ambiguities to empirically determine the genetics of the disease phenotypes in the *ARG3* mapping population. Determining the major cis-regulatory elements of *ARG3*, the protein properties and the mechanism of pathogen recognition may bring novel insight to the existing knowledge of plant immunity. The nature of a widescale sequence and functional variation in *ARG3* locus, and the global geocode of the variation may also

contribute to the practical side of ARG3-mediated disease resistance. Finally, the striking synteny in the wider *ARG3* locus to *O. sativa* may be of interest to evolutionary biologists.

# REFERENCES

1. Carvajal-Yepes, Monica, Cardwell, K, Nelson, A, Garrett, Karen A, Giovani, B, Saunders, DGO, Kamoun, S, Legg, JP, Verdier, Valérie and Lessel, J M. Carvajal-Yepes *et al.*, A global surveillance system for crop diseases. *Science* **364**, 1237-1239 (2019).
2. F. a. A. O. o. t. U. Nations (FAO, 2019) New standards to curb the global spread of plant pests and diseases.
3. B. Shiferaw, B. M. Prasanna, J. Hellin, M. Bänziger, Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food security* **3**, 307 (2011).
4. Fisher, Matthew C, Henk, Daniel A, Briggs, Cheryl J, Brownstein, John S, Madoff, Lawrence C, McCraw, Sarah L and Gurr, Sarah J
   M. C. Fisher *et al.*, Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186-194 (2012).
5. N. Spence, L. Hill, J. Morris, How the global threat of pests and diseases impacts plants, people, and the planet. *Plants, People, Planet* **2**, 5-13 (2020).
6. L. A. Boyd, C. Ridout, D. M. O'Sullivan, J. E. Leach, H. Leung, Plant–pathogen interactions: disease resistance in modern agriculture. *Trends in genetics* **29**, 233-240 (2013).
7. Woyann, Leomar Guilherme, Zdziarski, Andrei Daniel, Zanella, Rodrigo, Rosa, Ana Claudia, de Castro, Ricardo Lima, Caierão, Eduardo, Toigo, Marcelo De Carli, Storck, Lindolfo, Wu, Jixiang and Benin, Giovani
   L. G. Woyann *et al.*, Genetic Gain Over 30 Years of Spring Wheat Breeding in Brazil. *Crop Science* **59**, 2036-2045 (2019).
8. Cooper, Mark, Tang, Tom, Gho, Carla, Hart, Tim, Hammer, Graeme and Messina, Carlos
   M. Cooper *et al.*, Integrating genetic gain and gap analysis to predict improvements in crop productivity. *Crop Science* **60**, 582-604 (2020).
9. E. Donmez, R. Sears, J. Shroyer, G. Paulsen, Genetic gain in yield attributes of winter wheat in the Great Plains. *Crop Science* **41**, 1412-1419 (2001).
10. M. Pastor-Corrales, R. Frederiksen, Sorghum anthracnose. in *Sorghum diseases: a world review. Proceedings of International Workshop on Sorghum Diseases.[Google Scholar]*, pp 289-291 (1978).
11. C. R. Little, R. Perumal, The Biology and Control of Sorghum Diseases. *Sorghum: A State of the Art and Future Perspetives* **58**, 297-346 (2019).
12. R. Thakur, Genetic resistance to foliar anthracnose in sorghum and pathogenic variability in Colletotrichum graminicola. *Indian Phytopathology* **60**, 13-23 (2007).
13. C. Sherriff, M. Whelan, G. Arnold, J. Bailey, rDNA sequence analysis confirms the distinction between Colletotrichum graminicola and C. sublineolum. *Mycological Research* **99**, 475-478 (1995).
14. F. LeBeau, O. Coleman, 2490561. The inheritance of resistance in sorghum to leaf anthracnose. *Agronomy journal* **42**, 33-34 (1950).
15. Burrell, A Millie, Sharma, Arun, Patil, Nikhil Y, Collins, S Delroy, Anderson, William F, Rooney, William L and Klein, Patricia E A. M. Burrell *et al.*, Sequencing of an anthracnose-resistant sorghum genotype and mapping of a major QTL reveal strong candidate genes for anthracnose resistance. *Crop Science* **55**, 790-799 (2015).
16. E. Ahn, G. Odvody, L. K. Prom, C. Magill, Late Growth Stages of Johnsongrass Can Act as an Alternate Host of Colletotrichum sublineola. *Plant Health Progress* **21**, 60-62 (2020).
17. E. Ahn, L. K. Prom, G. Odvody, C. Magill, Defense responses against the sorghum anthracnose pathogen in leaf blade and midrib tissue of johnsongrass and sorghum. *Physiological and Molecular Plant Pathology* **106**, 81-86 (2019).

18. K. Xavier, T. Pfeiffer, D. Parreira, S. Chopra, L. Vaillancourt, Aggressiveness of Colletotrichum sublineola strains from Sorghum bicolor and S. halepense to sweet sorghum variety Sugar Drip, and their impact on yield. *Plant disease* **101**, 1578-1587 (2017).

19. K. Xavier, E. Mizubuti, M. Queiroz, S. Chopra, L. Vaillancourt, Genotypic and Pathogenic Diversity of Colletotrichum sublineola Isolates from Sorghum (Sorghum bicolor) and Johnsongrass (S. halepense) in the Southeastern United States. *Plant disease* **102**, 2341-2351 (2018).

20. R. Bowden, M. Kardin, J. Percich, L. Nickelson, Anthracnose of wild rice. *Plant disease* **68**, 68-69 (1984).

21. Prom, Louis K, Perumal, Ramasamy, Isakeit, Thomas, Radwan, Ghada, Rooney, William L and Magill, Clint L. K. Prom *et al.*, The impact of weather conditions on response of Sorghum genotypes to anthracnose (Colletotrichum sublineola) infection. *Journal of Experimental Agriculture International*, 242-250 (2015).

22. R. Baroncelli, J. M. Sanz-Martín, G. E. Rech, S. A. Sukno, M. R. Thon, Draft genome sequence of Colletotrichum sublineola, a destructive pathogen of cultivated sorghum. *Genome Announc.* **2**, e00540-00514 (2014).

23. M. Ali, H. Warren, Anthracnose of sorghum. *Sorghum and millet diseases: a second world review. In: Milliano WAJ, Frederiksen RA, Bengston GD (Eds.) International Crop Research Institute for the Semi-Arid Tropics. Patancheru. India*, 203-208 (1992).

24. R. Thakur, B. V. Reddy, K. Mathur, *Screening Techniques for Sorghum Diseases. Information: Bulletin No. 76*, International Crops Research Institute for the Semi-Arid Tropics, (2007).

25. G. Johal, S. Briggs, Reductase activity encoded by the HM1 disease resistance gene in maize. *Science* **258**, 985-987 (1992).

26. K. Laluk, T. Mengiste, Necrotroph attacks on plants: wanton destruction or covert extortion? *The Arabidopsis Book/American Society of Plant Biologists* **8** (2010).

27. P. D. Spanu, R. Panstruga, Biotrophic plant-microbe interactions. *Frontiers in plant science* **8**, 192 (2017).

28. Göhre, Vera, Spallek, Thomas, Häweker, Heidrun, Mersmann, Sophia, Mentzel, Tobias, Boller, Thomas, de Torres, Marta, Mansfield, John W and Robatzek, Silke V. Göhre *et al.*, Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. *Current biology* **18**, 1824-1832 (2008).

29. Baroncelli, Riccardo, Talhinhas, Pedro, Pensec, Flora, Sukno, Serenella A, Le Floch, Gaetan and Thon, Michael R R. Baroncelli *et al.*, The Colletotrichum acutatum species complex as a model system to study evolution and host specialization in plant pathogens. *Frontiers in Microbiology* **8**, 2001 (2017).

30. D. S. Manamgoda, D. Udayanga, L. Cai, E. Chukeatirote, K. D. Hyde, Endophytic Colletotrichum from tropical grasses with a new species C. endophytica. *Fungal Diversity* **61**, 107-115 (2013).

31. P. Jeffries, J. Dodd, M. Jeger, R. Plumbley, The biology and control of Colletotrichum species on tropical fruit crops. *Plant pathology* **39**, 343-366 (1990).

32. R. T. Villafana, S. N. Rampersad, Diversity, structure, and synteny of the cutinase gene of Colletotrichum species. *Ecology and Evolution* **10**, 1425-1443 (2020).

33. R. L. Winter *et al.*, Subcutaneous infection of a cat by Colletotrichum species. *Journal of feline medicine and surgery* **12**, 828-830 (2010).

34. J. Cano, J. Guarro, J. Gené, Molecular and morphological identification of Colletotrichum species of clinical interest. *Journal of clinical microbiology* **42**, 2450-2454 (2004).

35. M. Dowling, G. Schnabel, M. Williamson, R. Sekhon, K. Zielinski, First Report of Sorghum Anthracnose Caused by Colletotrichum sublineola in South Carolina. *Plant Disease* **100**, 1236-1236 (2016).

36. K. Cardwell, P. Hepperly, R. Frederiksen, Pathotypes of Colletotrichum graminicola and seed transmission of sorghum anthracnose. *Plant Disease* **73**, 255-257 (1989).

37.    J. Bailey, R. O'connell, R. Pring, C. Nash, Infection strategies of Colletotrichum species. *Infection strategies of Colletotrichum species.*, 88-120 (1992).
38.    J. A. Crouch, M. Tomaso-Peterson, Anthracnose disease of centipedegrass turf caused by Colletotrichum eremochloae, a new fungal species closely related to Colletotrichum sublineola. *Mycologia* **104**, 1085-1096 (2012).
39.    W. A. dos Santos Vieira *et al.*, Optimal markers for the identification of Colletotrichum species. *Molecular Phylogenetics and Evolution* **143**, 106694 (2020).
40.    T. G. Pillai, Pathogen to Endophytic Transmission in Fungi-A Proteomics Approach (2017).
41.    J. A. Crouch, B. Clarke, B. I. Hillman, Diversity in Populations of Colletotrichum Causing Anthracnose Disease in Grasses (2005).
42.    L. Tiffany, J. C. Gilman, Species of Colletotrichum from legumes. *Mycologia* **46**, 52-75 (1954).
43.    S. Freeman, T. Katan, E. Shabi, Characterization of Colletotrichum species responsible for anthracnose diseases of various fruits. *Plant disease* **82**, 596-605 (1998).
44.    X. Cao, X. Xu, H. Che, J. S. West, D. Luo, Three Colletotrichum species, including a new species, are associated to leaf anthracnose of rubber tree in Hainan, China. *Plant disease* **103**, 117-124 (2019).
45.    G. Sharma, A. K. Pinnaka, B. D. Shenoy, Resolving the Colletotrichum siamense species complex using ApMat marker. *Fungal Diversity* **71**, 247-264 (2015).
46.    U. Damm, R. O'Connell, J. Z. Groenewald, P. W. Crous, The Colletotrichum destructivum species complex–hemibiotrophic pathogens of forage and field crops. *Studies in mycology* **79**, 49-84 (2014).
47.    U. Damm, P. Cannon, J. Woudenberg, P. Crous, The Colletotrichum acutatum species complex. *Studies in mycology* **73**, 37-113 (2012).
48.    V. Squissato *et al.*, Colletotrichum truncatum species complex: Treatment considerations and review of the literature for an unusual pathogen causing fungal keratitis and endophthalmitis. *Medical mycology case reports* **9**, 1-6 (2015).
49.    U. Damm *et al.*, The Colletotrichum orbiculare species complex: Important pathogens of field crops and weeds. *Fungal Diversity* **61**, 29-59 (2013).
50.    U. Damm *et al.*, The Colletotrichum boninense species complex. *Studies in mycology* **73**, 1-36 (2012).
51.    F. Liu, L. Cai, P. Crous, U. Damm, The Colletotrichum gigasporum species complex. *Persoonia: Molecular Phylogeny and Evolution of Fungi* **33**, 83 (2014).
52.    B. Weir, P. Johnston, U. Damm, The Colletotrichum gloeosporioides species complex. *Studies in mycology* **73**, 115-180 (2012).
53.    U. Damm, T. Sato, A. Alizadeh, J. Groenewald, P. Crous, The Colletotrichum dracaenophilum, C. ámagnum and C. áorchidearum species complexes. *Studies in mycology* **92**, 1-46 (2019).
54.    F. Perrine-Walker, M. Anderson, Colletotrichum graminicola-Pathogen of the Month April 2019. (2019).
55.    J. A. Crouch, Colletotrichum caudatum sl is a species complex. *IMA fungus* **5**, 17-30 (2014).
56.    J. A. Crouch, B. B. Clarke, B. I. Hillman, Unraveling evolutionary relationships among the divergent lineages of Colletotrichum causing anthracnose disease in turfgrass and corn. *Phytopathology* **96**, 46-60 (2006).
57.    L. A. Beirn, B. B. Clarke, J. A. Crouch, Influence of host and geographic locale on the distribution of Colletotrichum cereale lineages. *PloS one* **9** (2014).
58.    P. Cannon, U. Damm, P. Johnston, B. Weir, Colletotrichum–current status and future directions. *Studies in mycology* **73**, 181-213 (2012).
59.    J. A. Crouch, L. A. Beirn, L. M. Cortese, S. A. Bonos, B. B. Clarke, Anthracnose disease of switchgrass caused by the novel fungal species Colletotrichum navitas. *Mycological research* **113**, 1411-1421 (2009).
60.    J. Crouch, L. Beirn, Anthracnose of cereals and grasses. *Fungal Diversity* **39**, 19 (2009).

61.     M. Carbú, J. Moraga, J. M. Cantoral, I. G. Collado, C. Garrido, Recent approaches on the genomic analysis of the phytopathogenic fungus Colletotrichum spp. *Phytochemistry Reviews*, 1-13 (2019).

62.     L. L. da Silva, H. L. A. Moreno, H. L. N. Correia, M. F. Santana, M. V. de Queiroz, Colletotrichum: species complexes, lifestyle, and peculiarities of some sources of genetic variability. *Applied Microbiology and Biotechnology*, 1-14 (2020).

63.     D. D. De Silva, P. W. Crous, P. K. Ades, K. D. Hyde, P. W. Taylor, Life styles of Colletotrichum species and implications for plant biosecurity. *Fungal Biology Reviews* **31**, 155-168 (2017).

64.     T. Takushi, S. Taba, Z. Moromizato, Latent infection on fruits and infection period of mango anthracnose fungi, Colletotrichum gloeosporioides and C. acutatum in a rainout shelter. *Japanese Journal of Phytopathology* **79**, 142-149 (2013).

65.     W. Skoropad, Effect of temperature on the ability of Colletotrichum graminicola to form appressoria and penetrate barley leaves. *Canadian Journal of Plant Science* **47**, 431-434 (1967).

66.     J. Crouch *et al.*, "The genomics of Colletotrichum" in Genomics of plant-associated fungi: monocot pathogens. Springer, pp. 69-102 (2014).

67.     Y. R. Eloy, I. M. Vasconcelos, A. L. Barreto, F. R. Freire-Filho, J. T. Oliveira, H2O2 plays an important role in the lifestyle of Colletotrichum gloeosporioides during interaction with cowpea [Vigna unguiculata (L.) Walp.]. *Fungal biology* **119**, 747-757 (2015).

68.     L. Delaye, G. García-Guzmán, M. Heil, Endophytes versus biotrophic and necrotrophic pathogens—are fungal lifestyles evolutionarily stable traits? *Fungal Diversity* **60**, 125-135 (2013).

69.     P. H. Fesel, A. Zuccaro, Dissecting endophytic lifestyle along the parasitism/mutualism continuum in Arabidopsis. *Current opinion in microbiology* **32**, 103-112 (2016).

70.     A. Latunde-Dada, R. O'connell, C. Nash, J. Lucas, Stomatal penetration of cowpea (Vigna unguiculata) leaves by a Colletotrichum species causing latent anthracnose. *Plant pathology* **48**, 777-785 (1999).

71.     K. Hiruma *et al.*, Root endophyte Colletotrichum tofieldiae confers plant fitness benefits that are phosphate status dependent. *Cell* **165**, 464-474 (2016).

72.     S. Hacquard *et al.*, Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature communications* **7**, 1-13 (2016).

73.     X.-H. Xu *et al.*, The rice endophyte Harpophora oryzae genome reveals evolution from a pathogen to a mutualistic endophyte. *Scientific Reports* **4**, 5783 (2014).

74.     C. B. de Queiroz *et al.*, The repertoire of effector candidates in Colletotrichum lindemuthianum reveals important information about Colletotrichum genus lifestyle. *Applied microbiology and biotechnology* **103**, 2295-2309 (2019).

75.     Z.-l. Yuan, Z.-z. Su, C.-l. Zhang, "11 Understanding the Biodiversity and Functions of Root Fungal Endophytes: The Ascomycete Harpophora oryzae as a Model Case" in Environmental and Microbial Relationships. Springer, pp. 205-214 (2016).

76.     P. R. Hardoim *et al.*, The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* **79**, 293-320 (2015).

77.     J. A. Crouch, Genomic signatures of pathogenicity and endophytism in five species of grass-associated Colletotrichum impacting the health and production of bioenergy feedstocks, agriculture and the environment.  DOE Joint Genome Institute (2011).

78.     P.-L. Plaumann, J. Schmidpeter, M. Dahl, L. Taher, C. Koch, A dispensable chromosome is required for virulence in the hemibiotrophic plant pathogen Colletotrichum higginsianum. *Frontiers in microbiology* **9**, 1005 (2018).

79.     P. Gan *et al.*, Genus-wide comparative genome analyses of Colletotrichum species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome biology and evolution* **8**, 1467-1481 (2016).

80.     E. A. Buiate *et al.*, A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species Colletotrichum graminicola and Colletotrichum sublineola. *BMC genomics* **18**, 67 (2017).
81.     E. Buiate *et al.*, Correction to: A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species Colletotrichum graminicola and Colletotrichum sublineola. *BMC genomics* **19**, 1-1 (2018).
82.     A. Sesma, A. E. Osbourn, The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature* **431**, 582-586 (2004).
83.     S. Breen, S. J. Williams, B. Winterberg, B. Kobe, P. S. Solomon, Wheat PR-1 proteins are targeted by necrotrophic pathogen effector proteins. *The Plant Journal* **88**, 13-25 (2016).
84.     R. Baroncelli *et al.*, Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. *BMC genomics* **17**, 555 (2016).
85.     J. Moraga *et al.*, The current status on secondary metabolites produced by plant pathogenic Colletotrichum species. *Phytochemistry Reviews* **18**, 215-239 (2019).
86.     D. Prusky, J. L. McEvoy, B. Leverentz, W. S. Conway, Local modulation of host pH by Colletotrichum species as a mechanism to increase virulence. *Molecular Plant-Microbe Interactions* **14**, 1105-1113 (2001).
87.     N. E. El Gueddari, U. Rauchhaus, B. M. Moerschbacher, H. B. Deising, Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytologist* **156**, 103-112 (2002).
88.     M. B. G. Rajulu *et al.*, Chitinolytic enzymes from endophytic fungi. *Fungal Diversity* **47**, 43-53 (2011).
89.     J. Fernandez, R. A. Wilson, Characterizing roles for the glutathione reductase, thioredoxin reductase and thioredoxin peroxidase-encoding genes of Magnaporthe oryzae during rice blast disease. *PLoS One* **9** (2014).
90.     F. Liu *et al.*, Molecular and phenotypic characterization of Colletotrichum species associated with anthracnose disease in peppers from Sichuan Province, China. *Scientific reports* **6**, 32761 (2016).
91.     T. Boon-Long, Sorghum diseases in Thailand. *Sorghum and Millets Diseases: A Second World Review. WAJ de Milliano, RA Fredericksen, y GD Bengston,(eds). International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India*, 41-43 (1992).
92.     K. Mathur, R. Thakur, V. Rao, A strain of Colletotrichum gloeosporioides causing anthracnose on yellow sorghum. *International Sorghum and Millets Newsletter* **38**, 69-70 (1997).
93.     G. F. Zanette, G. Nóbrega, L. D. Meirelles, Morphogenetic characterization of Colletotrichum sublineolum strains, causal agent of anthracnose of sorghum. *Tropical Plant Pathology* **34**, 146-151 (2009).
94.     E. Souza-Paccola, L. Fávaro, C. Bomfeti, S. Mesquita, L. Paccola-Meirelles, Cultural characterization and conidial dimorphism in Colletotrichum sublineolum. *Journal of Phytopathology* **151**, 383-388 (2003).
95.     C. R. CASELA, R. Frederiksen, Pathogenic variability in monoconidial isolates of the sorghum anthracnose fungus Colletotrichum graminicola from single lesions and from monoconidial cultures. *Embrapa Milho e Sorgo-Artigo em periódico indexado (ALICE)* (1994).
96.     J. Moore, M. Ditmore, D. TeBeest, Pathotypes of Colletotrichum sublineolum in Arkansas. *Plant disease* **92**, 1415-1420 (2008).
97.     E. Souza-Paccola, L. Fávaro, C. Casela, L. Paccola-Meirelles, Genetic recombination in Colletotrichum sublineolum. *Journal of Phytopathology* **151**, 329-334 (2003).
98.     U. Rosewich, R. E. Pettway, B. A. McDonald, R. Duncan, R. Frederiksen, Genetic structure and temporal dynamics of a Colletotrichum graminicola population in a sorghum disease nursery. *Phytopathology* **88**, 1087-1093 (1998).
99.     J. M. Lenné, R. Sonoda, D. Parbery, Production of conidia by setae of Colletotrichum species. *Mycologia* **76**, 359-362 (1984).

100. R. Williams, R. A. Frederiksen, J.-C. Girard, Sorghum and pearl millet disease identification handbook. International Crops Research Institute for the Semi-Arid Tropics (1978).

101. R. V. d. Costa, C. R. Casela, L. Zambolim, A. S. Ferreira, A antracnose do sorgo. *Fitopatologia Brasileira* **28**, 345-354 (2003).

102. J. Blakeman, D. Hornby, The persistence of Colletotrichum coccodes and Mycosphaerella ligulicola in soil, with special reference to sclerotia and conidia. *Transactions of the British Mycological Society* **49**, 227-IN227 (1966).

103. C. Venard, L. Vaillancourt, Penetration and colonization of unwounded maize tissues by the maize anthracnose pathogen Colletotrichum graminicola and the related nonpathogen C. sublineolum. *Mycologia* **99**, 368-377 (2007).

104. J. L. Dodd, The photosynthetic stress translocation balance concept of sorghum stalk rot. in *Proceedings of International Workshop of Sorghum Diseases*. Texas A&M University/ICRISAT Press Hyderabad, India, pp 300-305 (1980).

105. D. Hess, R. Bandyopadhyay, I. Sissoko, Reactions of Sorghum Genotypes to Leaf, Panicle, and Grain Anthracnose (Colletotrichum graminicola) under Field Conditions in Mali. (2001).

106. O. H. Coleman, I. Stokes, The inheritance of resistance to stalk red rot in sorghum. *Agronomy Journal* **46**, 61-63 (1954).

107. M. A. Mofokeng, H. Shimelis, M. Laing, N. Shargie, Sorghum ['Sorghum bicolor'(L.) Moench] breeding for resistance to leaf and stalk anthracnose, Colletotrichum sublineolum, and improved yield: Progress and prospects. *Australian Journal of Crop Science* **11**, 1078 (2017).

108. J. V. Patil, *Millets and sorghum : biology and genetic improvement*. Chichester, England : Wiley-Blackwell (2017).

109. R. A. Frederiksen, *Compendium of sorghum diseases*, The Disease compendium series. American Phytopathological Society in cooperation with Dept. of Plant Pathology and Microbiology, Texas A&M University, St. Paul, Minn (1986).

110. A. D. S. Ferreira, Evaluation of leaf anthracuose, head blight and red stalk rot of sorghum caused by Colletotrichum graminicola. MS thesis, Purdue University (1979).

111. L. Kelly, Y. Tan, M. Ryley, E. Aitken, Fusarium species associated with stalk rot and head blight of grain sorghum in Queensland and New South Wales, Australia. *Plant Pathology* **66**, 1413-1423 (2017).

112. F. Jamil, R. Nicholson, Susceptibility of corn to isolates of Colletotrichum graminicola pathogenic to other grasses. *Plant disease* **71**, 809-810 (1987).

113. S. A. Sukno, V. M. García, B. D. Shaw, M. R. Thon, Root infection and systemic colonization of maize by Colletotrichum graminicola. *Appl. Environ. Microbiol.* **74**, 823-832 (2008).

114. P. Wharton, A. Julian, A cytological study of compatible and incompatible interactions between Sorghum bicolor and Colletotrichum sublineolum. *New Phytologist* **134**, 25-34 (1996).

115. P. Wharton, A. Julian, R. O'Connell, Ultrastructure of the infection of Sorghum bicolor by Colletotrichum sublineolum. *Phytopathology* **91**, 149-158 (2001).

116. P. Basavaraju, N. P. Shetty, H. Shetty, E. De Neergaard, H. J. L. Jørgensen, Infection biology and defence responses in sorghum against Colletotrichum sublineolum. *Journal of applied microbiology* **107**, 404-415 (2009).

117. R. A. Wilson, N. J. Talbot, Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nature Reviews Microbiology* **7**, 185-195 (2009).

118. D. Gwary, T. Rabo, A. Anaso, The development of anthracnose symptoms on sorghum genotypes in the Nigerian savanna. *Journal of Plant Diseases and Protection* **111**, 96-103 (2004).

119. J. E. Erpelding, Anthracnose Resistance in Sorghum Breeding Lines Developed from Ethiopian Germplasm. *Plant health progress* **11** (2010).

120. R. Biffen, Studies on the inheritance of disease. *The Journal of Agricultural Science* **2**, 109 (1907).

121. E. Stakman, J. H. Parker, F. Piemeisel, Can biologic forms of stem rust on wheat change rapidly enough to interfere with breeding for rust resistance. *Jour. Agr. Res* **14**, 111-123 (1918).

122. H. Flor, "The complementary genic systems in flax and flax rust" in Advances in genetics. Elsevier, vol. 8, pp. 29-54 (1956).

123. E. Lagudah, S. K. Periyannan, Wheat stem rust resistance gene. Google Patents (2018).

124. F. V. Kazantsev, E. S. Skolotneva, V. N. Kelbin, E. A. Salina, S. A. Lashin, MIGREW: database on molecular identification of genes for resistance in wheat. *BMC bioinformatics* **20**, 36 (2019).

125. S. Fukuoka *et al.*, Gene pyramiding enhances durable blast disease resistance in rice. *Scientific reports* **5**, 1-7 (2015).

126. S. Fukuoka *et al.*, Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* **325**, 998-1001 (2009).

127. X. Chen *et al.*, "Genetic and Molecular Analyses of Blast Resistance in a Universal Blast Resistant Variety, Digu" in Advances in Genetics, Genomics and Control of Rice Blast Disease. Springer, pp. 149-159 (2009).

128. S. G. Krattinger *et al.*, A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**, 1360-1363 (2009).

129. S. Ceccarelli, S. Grando, From conventional plant breeding to molecular biology. *International Crop Science I*, 533-537 (1993).

130. R. Singh, J. Huerta-Espino, S. Rajaram, Achieving near-immunity to leaf and stripe rusts in wheat by combining slow rusting resistance genes. *Acta phytopathologica et entomologica hungarica* **35**, 133-139 (2000).

131. S. H. Kim, D. Qi, T. Ashfield, M. Helm, R. W. Innes, Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* **351**, 684-687 (2016).

132. B. Laflamme *et al.*, The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science* **367**, 763-768 (2020).

133. Y. Petit-Houdenot, I. Fudal, Complex interactions between fungal avirulence genes and their corresponding plant resistance genes and consequences for disease resistance management. *Frontiers in plant science* **8**, 1072 (2017).

134. S. Sharma, N. Gupta, Defense signaling in plants against micro-creatures: do or die.

135. J. Zhang *et al.*, Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. *Cell host & microbe* **7**, 290-301 (2010).

136. D. C. Boyes, J. Nam, J. L. Dangl, The Arabidopsis thaliana RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proceedings of the National Academy of Sciences* **95**, 15849-15854 (1998).

137. D. Mackey, Y. Belkhadir, J. M. Alonso, J. R. Ecker, J. L. Dangl, Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* **112**, 379-389 (2003).

138. R. B. Meeley, G. S. Johal, S. P. Briggs, J. D. Walton, A biochemical phenotype for a disease resistance gene of maize. *The Plant Cell* **4**, 71-77 (1992).

139. F. Tugizimana, A. T. Djami-Tchatchou, P. A. Steenkamp, L. A. Piater, I. A. Dubery, Metabolomic analysis of defense-related reprogramming in Sorghum bicolor in response to Colletotrichum sublineolum infection reveals a functional metabolic web of phenylpropanoid and flavonoid pathways. *Frontiers in plant science* **9**, 1840 (2019).

140. S.-C. C. Lo, J. D. Hipskind, R. L. Nicholson, cDNA cloning of a sorghum pathogenesis-related protein (PR-10) and differential expression of defense-related genes following inoculation with Cochliobolus heterostrophus or Colletotrichum sublineolum. *Molecular plant-microbe interactions* **12**, 479-489 (1999).

141. D. Godfrey, J. P. Rathjen, Recognition and Response in Plant PAMP-Triggered Immunity. *e LS* (2001).

142. T. Nürnberger, V. Lipka, Non-host resistance in plants: new insights into an old phenomenon. *Molecular plant pathology* **6**, 335-345 (2005).

143. R. Strugala, R. Delventhal, U. Schaffrath, An organ-specific view on non-host resistance. *Frontiers in Plant Science* **6**, 526 (2015).

144. J. Zhao *et al.*, Characterization and genetic analysis of rice mutant crr1 exhibiting compromised non-host resistance to Puccinia striiformis f. sp. tritici (Pst). *Frontiers in plant science* **7**, 1822 (2016).

145. Anonymous, (!!! INVALID CITATION !!! ).

146. L. Király, A. Künstler, R. Bacsó, Y. Hafez, Z. Király, Similarities and differences in plant and animal immune systems—what is inhibiting pathogens? *Acta Phytopathologica et Entomologica Hungarica* **48**, 187-205 (2013).

147. P. N. Dodds, J. P. Rathjen, Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* **11**, 539-548 (2010).

148. L. E. Rose *et al.*, The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in Arabidopsis thaliana. *Genetics* **166**, 1517-1527 (2004).

149. J. Lee *et al.*, Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. *The Plant Journal* **49**, 79-90 (2007).

150. H. G. Mang *et al.*, The Arabidopsis RESURRECTION1 gene regulates a novel antagonistic interaction in plant defense to biotrophs and necrotrophs. *Plant physiology* **151**, 290-305 (2009).

151. J. Monaghan, C. Zipfel, Plant pattern recognition receptor complexes at the plasma membrane. *Current opinion in plant biology* **15**, 349-357 (2012).

152. D. Chinchilla *et al.*, A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497-500 (2007).

153. L. Li *et al.*, The FLS2-associated kinase BIK1 directly phosphorylates the NADPH oxidase RbohD to control plant immunity. *Cell host & microbe* **15**, 329-338 (2014).

154. M. Benedetti *et al.*, Plant immunity triggered by engineered in vivo release of oligogalacturonides, damage-associated molecular patterns. *Proceedings of the National Academy of Sciences* **112**, 5533-5538 (2015).

155. Y. Li *et al.*, RESISTANCE TO POWDERY MILDEW 8.1 boosts pattern-triggered immunity against multiple pathogens in Arabidopsis and rice. *Plant biotechnology journal* **16**, 428-441 (2018).

156. J. Wang *et al.*, Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science* **364** (2019).

157. R. T. Leister, F. M. Ausubel, F. Katagiri, Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the Arabidopsis genes RPS2 and RPM1. *Proceedings of the National Academy of Sciences* **93**, 15497-15502 (1996).

158. P. Veronese *et al.*, The membrane-anchored BOTRYTIS-INDUCED KINASE1 plays distinct roles in Arabidopsis resistance to necrotrophic and biotrophic pathogens. *The Plant Cell* **18**, 257-273 (2006).

159. X. Tang *et al.*, Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. *Science* **274**, 2060-2063 (1996).

160. K. A. Parsons, F. G. Chumley, B. Valent, Genetic transformation of the fungal pathogen responsible for rice blast disease. *Proceedings of the National Academy of Sciences* **84**, 4161-4165 (1987).

161. B. Valent, Rice blast as a model system for plant pathology. *Phytopathology* **80**, 33-36 (1990).

162. M. G. Kim, X. Geng, S. Y. Lee, D. Mackey, The Pseudomonas syringae type III effector AvrRpm1 induces significant defenses by activating the Arabidopsis nucleotide-binding leucine-rich repeat protein RPS2. *The Plant Journal* **57**, 645-653 (2009).

163. W. Gassmann, S. Bhattacharjee, Effector-triggered immunity signaling: from gene-for-gene pathways to protein-protein interaction networks. *Molecular Plant-Microbe Interactions* **25**, 862-868 (2012).

164. M. J. Axtell, B. J. Staskawicz, Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**, 369-377 (2003).

165. D. Mackey, B. F. Holt III, A. Wiig, J. L. Dangl, RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell* **108**, 743-754 (2002).
166. B. Day, D. Dahlbeck, B. J. Staskawicz, NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. *The Plant Cell* **18**, 2782-2791 (2006).
167. H. Irieda *et al.*, Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proceedings of the National Academy of Sciences* **116**, 496-505 (2019).
168. H. K. Chen, Anthracnose of grain sorghum caused by Colletotrichum lineola corda. (1934).
169. S. Pande, R. Thakur, R. Karunakar, R. Bandyopadhyay, B. V. Reddy, Development of screening methods and identification of stable resistance to anthracnose in sorghum. *Field Crops Research* **38**, 157-166 (1994).
170. C. Schlötterer, The evolution of molecular markers—just a matter of fashion? *Nature reviews genetics* **5**, 63-69 (2004).
171. P. J. Mehta *et al.*, Classification and inheritance of genetic resistance to anthracnose in sorghum. *Field Crops Research* **93**, 1-9 (2005).
172. N. Y. Patil *et al.*, Quantitative trait loci associated with anthracnose resistance in sorghum. *Crop Science* **57**, 877-890 (2017).
173. T. Isakeit, S. Collins, W. Rooney, L. Prom, Reaction of sorghum hybrids to anthracnose, grain mold and grain weathering in Burleson County, Texas, 2007. *Plant Disease Management Reports* **2** (2008).
174. J. E. Erpelding, L. K. Prom, Evaluation of Malian sorghum germplasm for resistance against anthracnose. *Plant Pathol. J* **3**, 65-71 (2004).
175. J. E. Erpelding, L. K. Prom, Variation for anthracnose resistance within the sorghum germplasm collection from Mozambique, Africa. *Plant Pathol J* **5**, 28-34 (2006).
176. J. E. Erpelding, Sorghum germplasm resistance to anthracnose. *Am J Plant Sci Biotechnol* **2**, 42-46 (2008).
177. R. Sharma, H. Upadhyaya, S. Manjunatha, V. Rao, R. Thakur, Resistance to foliar diseases in a mini-core collection of sorghum germplasm. *Plant disease* **96**, 1629-1633 (2012).
178. H. E. Cuevas, L. K. Prom, E. A. Cooper, J. E. Knoll, X. Ni, Genome-wide association mapping of anthracnose (Colletotrichum sublineolum) resistance in the US sorghum association panel. *The plant genome* **11** (2018).
179. H. E. Cuevas, L. K. Prom, C. M. Cruet-Burgos, Genome-wide association mapping of anthracnose (Colletotrichum sublineolum) resistance in NPGS Ethiopian sorghum germplasm. *G3: Genes, Genomes, Genetics* **9**, 2879-2885 (2019).
180. H. E. Cuevas, L. K. Prom, Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS Sudan Sorghum Core collection. *BMC genomics* **21**, 88 (2020).
181. R. Nelson, T. Wiesner-Hanks, R. Wisser, P. Balint-Kurti, Navigating complexity to breed disease-resistant crops. *Nature Reviews Genetics* **19**, 21 (2018).
182. S.-C. C. LO, K. DE VERDIER, R. L. NICHOLSON, Accumulation of 3-deoxyanthocyanidin phytoalexins and resistance to Colletotrichum sublineolum in sorghum. *Physiological and Molecular Plant Pathology* **55**, 263-273 (1999).
183. G. Mengistu, H. Shimelis, M. Laing, D. Lule, J. Mashilo, Combining ability and heterosis among sorghum (Sorghum bicolor [L.] Moench) lines for yield, yield-related traits, and anthracnose resistance in western Ethiopia. *Euphytica* **216**, 33 (2020).
184. P. Klein, W. L. Rooney, A. M. Burrell, D. Collins, Methods and compositions for producing sorghum plants with anthracnose resistance. Google Patents (2017).
185. R. Da Costa *et al.*, Genetic control of sorghum resistance to leaf anthracnose. *Plant pathology* **60**, 1162-1168 (2011).

186.     L. Prom *et al.*, Genetic diversity and pathotype determination of Colletotrichum sublineolum isolates causing anthracnose in sorghum. *European journal of plant pathology* **133**, 671-685 (2012).
187.     J. E. Erpelding, Inheritance of Anthracnose Resistance for the Sorghum Cultivar Redi an. *Plant Pathol. J* **6**, 187-190 (2007).
188.     A. caused by Colletotrichum, Resistance of sorghum to Colletotrichum graminicola. *Plant Disease* **66**, 773-775 (1982).
189.     A. Tenkouano, F. Miller, A single locus with multiple alleles as the genetic basis of anthracnose resistance in sorghum. *Sorghum Newsletter* **34** (1993).
190.     K. S. Boora, R. Frederiksen, C. Magill, DNA-based markers for a recessive gene conferring anthracnose resistance in sorghum. *Crop Science* **38**, 1708-1709 (1998).
191.     M. Singh, K. Chaudhary, H. Singal, C. Magill, K. Boora, Identification and characterization of RAPD and SCAR markers linked to anthracnose resistance gene in sorghum [Sorghum bicolor (L.) Moench]. *Euphytica* **149**, 179-187 (2006).
192.     P. Ramasamy *et al.*, Molecular mapping of Cg1, a gene for resistance to anthracnose (Colletotrichum sublineolum) in sorghum. *Euphytica* **165**, 597 (2009).
193.     H. E. Cuevas, L. K. Prom, J. E. Erpelding, Inheritance and molecular mapping of anthracnose resistance genes present in sorghum line SC112-14. *Molecular breeding* **34**, 1943-1953 (2014).
194.     C. M. Cruet-Burgos *et al.*, Genomic Dissection of Anthracnose (Colletotrichum sublineolum) Resistance Response in Sorghum Differential Line SC112-14. *G3: Genes, Genomes, Genetics* **10**, 1403-1412 (2020).
195.     H. D. Upadhyaya, Y.-H. Wang, R. Sharma, S. Sharma, Identification of genetic markers linked to anthracnose resistance in sorghum using association analysis. *Theoretical and applied genetics* **126**, 1649-1657 (2013).
196.     T. J. Felderhoff, L. M. McIntyre, A. Saballos, W. Vermerris, Using genotyping by sequencing to map two novel anthracnose resistance loci in Sorghum bicolor. *G3: Genes, Genomes, Genetics* **6**, 1935-1946 (2016).
197.     S. M. Mohan *et al.*, Identification of quantitative trait loci associated with resistance to foliar diseases in sorghum [Sorghum bicolor (L.) Moench]. *Euphytica* **176**, 199-211 (2010).
198.     M. Biruma, T. Martin, I. Fridborg, P. Okori, C. Dixelius, Two loci in sorghum with NB-LRR encoding genes confer resistance to Colletotrichum sublineolum. *Theoretical and applied genetics* **124**, 1005-1015 (2012).
199.     E. Ahn, L. Prom, G. Odvody, C. Magill, Responses of Johnsongrass against Sorghum Anthracnose isolates. *Journal of Plant Pathology & Microbiology* **9**, 442-448 (2018).
200.     C. Mérot, R. A. Oomen, A. Tigano, M. Wellenreuther, A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution* (2020).
201.     R. R. Fuentes *et al.*, Structural variants in 3000 rice genomes. *Genome research* **29**, 870-880 (2019).
202.     S. Kim *et al.*, New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome biology* **18**, 210 (2017).
203.     S. Pinosio *et al.*, Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Molecular biology and evolution* **33**, 2706-2719 (2016).
204.     A. Dolatabadian *et al.*, Characterization of disease resistance genes in the Brassica napus pangenome reveals significant structural variation. *Plant biotechnology journal* **18**, 969-982 (2020).
205.     R. F. McCormick *et al.*, The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal* **93**, 338-354 (2018).
206.     J. P. Bezerra-Neto *et al.*, "NBS-LRR genes—Plant health sentinels: Structure, roles, evolution and biotechnological applications" in Applied Plant Biotechnology for Improving Resistance to Biotic Stress. Elsevier, pp. 63-120 (2020).

207.  A.-L. Van de Weyer *et al.*, The Arabidopsis thaliana pan-NLRome. *BioRxiv*, 537001 (2019).
208.  X. Yang, J. Wang, Genome-wide analysis of NBS-LRR genes in sorghum genome revealed several events contributing to NBS-LRR gene evolution in grass species. *Evolutionary Bioinformatics* **12**, EBO. S36433 (2016).
209.  E. Mace *et al.*, The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. *BMC Plant Biology* **14** (2014).
210.  A.-L. Van de Weyer *et al.*, A species-wide inventory of NLR genes and alleles in Arabidopsis thaliana. *Cell* **178**, 1260-1272. e1214 (2019).
211.  E. Stella, E. Maud, E. Philippe, E. Thomas, N. D. Peter, A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. *Frontiers in Plant Science* **5** (2014).
212.  X. Cheng *et al.*, A genomic analysis of disease-resistance genes encoding nucleotide binding sites in Sorghum bicolor. *Genetics and Molecular Biology* **33**, 292-297 (2010).
213.  H. Mizuno *et al.*, Evolutionary dynamics and impacts of chromosome regions carrying R-gene clusters in rice. *Scientific reports* **10**, 1-9 (2020).
214.  S. Singh, S. Chand, N. Singh, T. R. Sharma, Genome-wide distribution, organisation and functional characterization of disease resistance and defence response genes across rice species. *Plos one* **10** (2015).
215.  B.-H. Wang, D. J. Ebbole, Z.-H. Wang, The arms race between Magnaporthe oryzae and rice: Diversity and interaction of Avr and R genes. *Journal of Integrative Agriculture* **16**, 2746-2760 (2017).
216.  N. J. Talbot, A. J. Foster, Genetics and genomics of the rice blast fungus Magnaporthe grisea: developing an experimental model for understanding fungal diseases of cereals. (2001).
217.  S. Ashkani *et al.*, Molecular progress on the mapping and cloning of functional genes for blast disease in rice (Oryza sativa L.): current status and future considerations. *Critical reviews in biotechnology* **36**, 353-367 (2016).
218.  P. K. Singh *et al.*, Co-evolutionary interactions between host resistance and pathogen avirulence genes in rice-Magnaporthe oryzae pathosystem. *Fungal Genetics and Biology* **115**, 9-19 (2018).
219.  W. Li, M. Chern, J. Yin, J. Wang, X. Chen, Recent advances in broad-spectrum resistance to the rice blast disease. *Current opinion in plant biology* **50**, 114-120 (2019).
220.  W. Schnippenkoetter *et al.*, The wheat Lr34 multipathogen resistance gene confers resistance to anthracnose and rust in sorghum. *Plant biotechnology journal* **15**, 1387-1396 (2017).
221.  S. G. Krattinger *et al.*, The wheat durable, multipathogen resistance gene Lr34 confers partial blast resistance in rice. *Plant biotechnology journal* **14**, 1261-1268 (2016).
222.  L. Li, F. Zhu, H. Liu, A. Chu, C. Lo, Isolation and expression analysis of defense-related genes in sorghum–Colletotrichum sublineolum interaction. *Physiological and molecular plant pathology* **84**, 123-130 (2013).
223.  A. C. Velho, M. J. Stadnik, Non-host resistance of arabidopsis and apple is associated with callose accumulation and changes in preinfective structures of Colletotrichum species. *Physiological and Molecular Plant Pathology* **110**, 101463 (2020).
224.  A. Tenkouano, F. R. Miller, G. E. Hart, R. A. Frederiksen, R. L. Nicholson, Phytoalexin assay in juvenile sorghum: an aid to breeding for anthracnose resistance. *Crop science* **33**, 243-248 (1993).
225.  H. Cui, K. Tsuda, J. E. Parker, Effector-triggered immunity: from pathogen perception to robust defense. *Annual review of plant biology* **66**, 487-511 (2015).
226.  A. J. Anderson, Isolation from three species of Colletotrichum of glucan-containing polysaccharides that elicit browning and phytoalexin production in bean. *Phytopathology* **68**, 189-194 (1978).
227.  D. Muth, P. Kachlicki, P. Krajewski, M. Przystalski, M. Stobiecki, Differential metabolic response of narrow leafed lupine (Lupinus angustifolius) leaves to infection with Colletotrichum lupini. *Metabolomics* **5**, 354-362 (2009).

228. S. G. Krattinger *et al.*, Abscisic acid is a substrate of the ABC transporter encoded by the durable wheat disease resistance gene Lr34. *New Phytologist* **223**, 853-866 (2019).
229. Y. Deng *et al.*, Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* **355**, 962-965 (2017).
230. X. Liu, H. Inoue, N. Hayashi, C.-J. Jiang, H. Takatsuji, CC-NBS-LRR-type R proteins for rice blast commonly interact with specific WRKY transcription factors. *Plant molecular biology reporter* **34**, 533-537 (2016).
231. K. Hyde *et al.*, Colletotrichum—names in current use. *Fungal Diversity* **39**, 147-182 (2009).
232. M. Rai, G. Agarkar, D. Rathod, Multiple applications of endophytic Colletotrichum species occurring in medicinal plants. *Novel plant bioresources: applications in food, medicine and Cosmetics*, 227-236 (2014).
233. K. Lamsal *et al.*, Application of silver nanoparticles for the control of Colletotrichum species in vitro and pepper anthracnose disease in field. *Mycobiology* **39**, 194-199 (2011).
234. H. Ishii, F. Zhen, M. Hu, X. Li, G. Schnabel, Efficacy of SDHI fungicides, including benzovindiflupyr, against Colletotrichum species. *Pest management science* **72**, 1844-1853 (2016).
235. H. Dinler, S. Benlioğlu, The possibility to control diseases caused by colletotrichum species in strawberries. *International Journal of Agriculture Forestry and Life Sciences* **3**, 362-370 (2019).
236. N. Landero Valenzuela *et al.*, Antifungal potential of extracts from four vegetables species on the growth of Colletotrichum gloeosporioides in postharvest papaya (Carica papaya). *Revista Venezolana de Ciencia y Tecnología de Alimentos* **4**, 47-62 (2013).
237. R. S. Resende, C. A. Milagres, D. Rezende, C. E. Aucique-Perez, F. Á. Rodrigues, Bioprospecting of saprobe fungi from the Semi-Arid North-East of Brazil for the control of anthracnose on sorghum. *Journal of Phytopathology* **163**, 787-794 (2015).
238. M. M. Vasanthakumari, M. B. Shivanna, Biological control of sorghum anthracnose with rhizosphere and rhizoplane fungal isolates from perennial grasses of the Western Ghats of India. *European journal of plant pathology* **139**, 721-733 (2014).
239. R. S. Resende, F. Á. Rodrigues, R. V. Costa, D. D. Silva, Silicon and fungicide effects on anthracnose in moderately resistant and susceptible sorghum lines. *Journal of Phytopathology* **161**, 11-17 (2013).
240. D. Gwary, N. Bwatanglang, B. Bdliya, Integrated management of sorghum anthracnose through the use of fungicides, crop varieties and manipulation of sowing dates in Sudan Savanna of Nigeria. *International Journal of Agriculture and Biology* **10**, 661-664 (2008).
241. K. K. Romana, G. Chander, S. Deshpande, R. Gupta, "Genomic-Assisted Enhancement in Stress Tolerance for Productivity Improvement in Sorghum" in Biotechnologies of Crop Improvement, Volume 3. Springer, pp. 265-288 (2018).
242. D. A. St. Clair, Quantitative disease resistance and quantitative resistance loci in breeding. *Annual review of phytopathology* **48**, 247-268 (2010).
243. W. Vermerris, H. Cuevas, L. Prom, J. Knoll, Genomic dissection of anthracnose resistance response in sorghum [*Sorghum bicolor* (L.) Moench]. Univ. of Florida, Gainesville, FL .United States (2018).
244. S. Kaur, S. Jindal, M. Kaur, P. Chhuneja, "Utilization of wild species for wheat improvement using genomic approaches" in Biotechnologies of Crop Improvement, Volume 3. Springer, pp. 105-150 (2018).
245. L. Sun, J. Qin, K. Wang, J. Zhang, Expansion of pathogen recognition specificity in plants using pattern recognition receptors and artificially designed decoys. *Science China Life Sciences* **60**, 797-805 (2017).
246. H. Takagi *et al.*, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal* **74**, 174-183 (2013).
247. L. Lambright, "Hybrid Sorghum Product Development and Production" in Sorghum. Springer, pp. 3-9 (2019).

248.    s. Laxman, Plastic Bag Emasculation Technique in Sorghum *Sorghum bicolor* L. (Moench) (1997).

249.    A. Redkar, E. Jaeger, G. Doehlemann, Visualization of Growth and Morphology of Fungal Hyphae in planta Using WGA-AF488 and Propidium Iodide Co-staining (2018).

250.    F. Davila-Gomez, C. Chuck-Hernandez, E. Perez-Carrillo, W. Rooney, S. Serna-Saldivar, Evaluation of bioethanol production from five different varieties of sweet and forage sorghums (*Sorghum bicolor* (L) Moench). *Industrial Crops and Products* **33**, 611-616 (2011).

251.    T. Maekawa *et al.*, Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell host & microbe* **9**, 187-199 (2011).

252.    M. R. Grant *et al.*, Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* **269**, 843-846 (1995).

253.    R. Mishra *et al.*, Differential expression of defense-related genes in chilli pepper infected with anthracnose pathogen Colletotrichum truncatum. *Physiological and Molecular Plant Pathology* **97**, 1-10 (2017).

254.    R. E. Atkins, Registration of IAP2B(M)C3 Sorghum Germplasm1 (Reg. No. GP131). *Crop Science* **22**, 1275-1275 (1982).

255.    M. Vyska, N. Cunniffe, C. Gilligan, Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective. *Journal of The Royal Society Interface* **13**, 20160451 (2016).

256.    S. Pinosio *et al.*, Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. *Molecular Biology and Evolution* **33**, 2706-2719 (2016).

257.    L.-Y. Zheng *et al.*, Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor). *Genome biology* **12**, R114 (2011).

258.    M. C. Schatz *et al.*, New whole genome de novo assemblies of three divergent strains of rice (O. sativa) documents novel gene space of aus and indica. *bioRxiv*, 003764 (2014).

259.    W. Wang, B. Feng, J. M. Zhou, D. Tang, Plant immune signaling: Advancing on two frontiers. *Journal of Integrative Plant Biology* (2019).

260.    T. M. Burch-Smith *et al.*, A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biol* **5**, e68 (2007).

261.    N. Panchy, M. Lehti-Shiu, S.-H. Shiu, Evolution of gene duplication in plants. *Plant physiology* **171**, 2294-2316 (2016).

262.    T. Ashfield, L. E. Ong, K. Nobuta, C. M. Schneider, R. W. Innes, Convergent evolution of disease resistance gene specificity in two flowering plant families. *The Plant Cell* **16**, 309-318 (2004).

263.    J. M. McDowell, Convergent evolution of disease resistance genes. *Trends in plant science* **9**, 315-317 (2004).

264.    A. A. Agrawal, Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *The American Naturalist* **190**, S1-S12 (2017).

265.    C. W. Smith, R. A. Frederiksen, *Sorghum: Origin, history, technology, and production*. John Wiley & Sons (2000).

266.    V. Geffroy *et al.*, Identification of an ancestral resistance gene cluster involved in the coevolution process between Phaseolus vulgaris and its fungal pathogen *Colletotrichum lindemuthianum*. Molecular plant-microbe interactions **12**, 774-784 (1999).

267.    B. Yuan *et al.*, The Pik-p resistance to Magnaporthe oryzae in rice is mediated by a pair of closely linked CC-NBS-LRR genes. *Theoretical and applied genetics* **122**, 1017-1028 (2011).

268.    C. H. Khang *et al.*, Translocation of Magnaporthe oryzae effectors into rice cells and their subsequent cell-to-cell movement. *The Plant Cell* **22**, 1388-1403 (2010).

269.    P. Kankanala, K. Czymmek, B. Valent, Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *The Plant Cell* **19**, 706-724 (2007).

270. M. F. Torres *et al.*, A Colletotrichum graminicola mutant deficient in the establishment of biotrophy reveals early transcriptional events in the maize anthracnose disease interaction. *BMC genomics* **17**, 202 (2016).
271. Y. Lai, T. Eulgem, Transcript-level expression control of plant NLR genes. *Molecular plant pathology* **19**, 1267-1281 (2018).
272. B. Y. Chung, C. Simons, A. E. Firth, C. M. Brown, R. P. Hellens, Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC genomics* **7**, 120 (2006).
273. M. Laxa *et al.*, The 5′ UTR intron of Arabidopsis GGT1 aminotransferase enhances promoter activity by recruiting RNA polymerase II. *Plant physiology* **172**, 313-327 (2016).
274. Y. Kurihara *et al.*, Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proceedings of the National Academy of Sciences* **115**, 7831-7836 (2018).
275. H. Zhang *et al.*, Genome editing of upstream open reading frames enables translational control in plants. *Nature Biotechnology* **36**, 894+ (2018).
276. G. Xu *et al.*, uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* **545**, 491-494 (2017).
277. R. Niu *et al.*, uORFlight: a vehicle towards uORF-mediated translational regulation mechanisms in eukaryotes. *bioRxiv*, 713321 (2019).
278. D. A. Halterman, R. P. Wise, Upstream open reading frames of the barley Mla13 powdery mildew resistance gene function co-operatively to down-regulate translation. *Molecular plant pathology* **7**, 167-176 (2006).
279. J. K. Brown, A cost of disease resistance: paradigm or peculiarity? *Trends in Genetics* **19**, 667-671 (2003).
280. K. Christine *et al.*, A stilbene synthase gene (SbSTS1) is involved in host and nonhost defense responses in sorghum. *Plant physiology* **138**, 393-401 (2005).
281. C. R. Mareya *et al.*, Untargeted metabolomics reveal defensome-related metabolic reprogramming in *Sorghum bicolor* against infection by *Burkholderia andropogonis*. *Metabolites* **9**, 8 (2019).
282. F. Bao *et al.*, *Arabidopsis* HSP90 protein modulates RPP4-mediated temperature-dependent cell death and defense responses. *New Phytologist* **202**, 1320-1334 (2014).
283. L. Li, A. Habring, K. Wang, D. Weigel, Atypical resistance protein RPW8/HR triggers oligomerization of the NLR immune receptor RPP7 and autoimmunity. *Cell Host & Microbe* (2020).
284. R. Büschges *et al.*, The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell* **88**, 695-705 (1997).
285. E. J. Andersen *et al.*, Wheat Disease Resistance Genes and Their Diversification Through Integrated Domain Fusions. *Frontiers in Genetics* **11**, 898 (2020).
286. C. Zipfel, Plant pattern-recognition receptors. *Trends in immunology* **35**, 345-351 (2014).
287. G. S. Johal, S. P. Briggs, Reductase activity encoded by the HM1 disease resistance gene in maize. *Science* **258**, 985-987 (1992).
288. P. Römer *et al.*, Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645-648 (2007).
289. R. Brueggeman *et al.*, The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences* **99**, 9328-9333 (2002).
290. M. M. Richard *et al.*, Fine mapping of Co-x, an anthracnose resistance gene to a highly virulent strain of Colletotrichum lindemuthianum in common bean. *Theoretical and applied genetics* **127**, 1653-1666 (2014).
291. V. Ntoukakis *et al.*, The tomato Prf complex is a molecular trap for bacterial effectors based on Pto transphosphorylation. *PLoS Pathog* **9**, e1003123 (2013).

292. Z. Zhang, B. P. Thomma, Structure-function aspects of extracellular leucine-rich repeat-containing cell surface receptors in plants. *Journal of integrative plant biology* **55**, 1212-1223 (2013).
293. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* **70**, e102 (2020).
294. R. R. Wick, M. B. Schultz, J. Zobel, K. E. Holt, Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350-3352 (2015).
295. G. Girma *et al.*, A large-scale genome-wide association analyses of ethiopian sorghum landrace collection reveal loci associated with important traits. *Frontiers in plant science* **10**, 691 (2019).
296. H. Wang *et al.*, Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science* **368** (2020).
297. C. M. Cruet-Burgos, Genome mapping of anthracnose resistance in sorghum germplasm (2018).
298. C. Aluome *et al.*, De novo construction of a "Gene-space" for diploid plant genome rich in repetitive sequences by an iterative Process of Extraction and Assembly of NGS reads (iPEA protocol) with limited computing resources. *BMC research notes* **9**, 81 (2016).
299. H. E. Cuevas, G. Rosa-Valentin, C. M. Hayes, W. L. Rooney, L. Hoffmann, Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC genomics* **18**, 108 (2017).
300. J. Ruiz-Orera, J. L. Villanueva-Cañas, M. M. Albà, Evolution of new proteins from translated sORFs in long non-coding RNAs. *Experimental Cell Research*, 111940 (2020).