Genetic analysis of sorghum (*Sorghum bicolor* (L.) Moench ssp.) for resistance to anthracnose (*Colletotrichum sublineolum*) and grain mold diseases

by

Diriba Hika Chere

B.S., Haramaya University, 2002
M.S., Haramaya University, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

Fungal diseases, such as anthracnose and grain mold pose significant challenges to the production and productivity of sorghum. Favored by warm and humid conditions, sorghum Anthracnose *(Colletotrichum sublineolum*) is among the most devastating diseases of sorghum that can cause yield losses of up to 50% in susceptible cultivars. It can attack all above-ground parts including foliage, stalk, panicle, and grain. The severity of infection depends on the genotype, the pathotype, and the environment. Grain mold is another major sorghum disease that is prevalent under warm and wet conditions that occur during grain development. This disease is caused by one or more genera of fungi of which *Aspergillus* and *Fusarium* are the most common. Breeding efforts have focused on the development of resistant varieties that can reduce the impact of these diseases. Conventional breeding supported by genomic tools may enhance this effort and accelerate the release of new resistant cultivars. To contribute to this endeavor three experiments were initiated and executed. This dissertation thus presents the results of a comprehensive investigation into enhancing resistance to sorghum anthracnose and grain mold diseases through genomic selection, genome-wide association studies (GWAS), and combining ability analysis.

The first chapter explores the utilization of genomic selection techniques to improve anthracnose and grain mold resistance in sorghum through prediction using genomic estimated breeding values (GEBVs). Through the integration of advanced genomic tools and phenotypic data, we demonstrated the potential for accurately predicting resistance levels in sorghum lines, that may facilitate the development of more resilient cultivars. During the 2015 and 2016 cropping seasons, a study was conducted in Bako, Ethiopia, involving 946 sorghum accessions obtained from the gene bank. These accessions were examined for resistance to sorghum leaf anthracnose (*Colletotrichum sublineolum*) and grain mold. Genotyping of these accessions was performed

using the genotype-by-sequencing (GBS) method, resulting in the generation of 410,717 high-quality SNP markers. To investigate the impact of SNP density on prediction accuracy, the markers were pruned using TASSEL software, maintaining marker distances of 1,000 bp, 50,000 bp, 75,000 bp, and 100,000 bp between them. The analysis involved varying the size of the training population from 10% to 90% at intervals of 10%. Principal component analysis was carried out to assess the genetic relatedness patterns among the sorghum accessions. For anthracnose resistance, we obtained a prediction accuracy of 0.38 and 0.45 for the 10% and 90% training population sizes respectively for the 2016 data. A prediction accuracy of 0.38 and 0.41 were recorded for grain mold at 10% and 80% training population sizes.

In the second experiment, a GWAS study was conducted to identify key genomic regions and candidate genes associated with sorghum anthracnose resistance. Our analysis unveiled the presence of Single Nucleotide Polymorphisms (SNPs) on several chromosomes, specifically chromosomes 1, 2, 3, 4, 6, 8, and 10. Notably, chromosomes 8 and 1 exhibited a higher abundance of significant SNPs. Chromosome 8 displayed the most pronounced genetic associations, with nine specific loci (S8_16230498, S8_16389715, S8_37466666, S8_39664335, S8_42468301, S8_42983043, S8_43776246, S8_7908407, and S8_570182) having significant associations with anthracnose disease reaction. Similarly, we identified multiple loci on chromosome 1 (S1_20098885, S1_49271328, S1_49271328, S1_49454848, S1_50111275, S1_50294000, S1_50533418, S1_30455882, and S1_50721682) that also have significant association with the disease. Chromosome 2 revealed two significant loci (S2_55407842 and S2_69275221), while only one locus on chromosome 3 (S3_53970498) showed significant association. Two loci each on chromosomes 4 (S4_62789639, S4_7538383) and chromosome 6, (S6_25755784, and S6_55783758) were shown to have significant association with the disease. By leveraging high-

throughput genotyping data and extensive phenotypic evaluations, this research provides valuable insights into the genetics of anthracnose resistance, offering a foundation for marker-assisted breeding efforts.

The third experiment explored the combining ability of sorghum lines for anthracnose resistance. By assessing the genetic interactions among different sorghum parental lines, this study elucidated the potential for breeding strategies that harness the complementary genetic contributions of parental lines to enhance resistance traits in the progeny. We conducted the experiment across five different environments and collected data on various agronomic traits, including flowering time (DTF), plant height (PTH), panicle weight (PKW), grain yield per panicle (GYP), hundred-grain weight (HGW), and resistance to anthracnose (ANT). Our findings indicated that there were significant variations in trait reactions across different replicates and environmental conditions, particularly for PTH, GYP, and ANT due to notable environmental effects. The entry effect, which represents the contributions of different parental lines and hybrid combinations, was highly significant for all traits we examined. Additionally, the interaction between male and female parents (M vs. F) showed a high level of significance specifically for ANT. A negative significant correlation was observed between panicle weight and anthracnose and grain yield per panicle and anthracnose. Through our combining ability analysis, we were able to identify the genetic contributions of various genotypes. Notably, certain parents such as R245, R604, R633, A539, and A553 exhibited high general combining ability (GCA) for anthracnose resistance. Furthermore, some hybrid combinations like A539 × R669, A527 × R672, A531 × R672, A539 × R588 demonstrated highly significant specific combining ability (SCA) effects for anthracnose resistance. These findings provide valuable insights for breeders, enabling them to make well-informed decisions when designing breeding strategies.

Collectively, this thesis represents a multidisciplinary approach to addressing the pressing issue of anthracnose and grain mold resistance in sorghum. The findings contribute to the development of more resilient sorghum cultivars, thus supporting global food security and sustainable agriculture practices.

Genetic analysis of sorghum (*Sorghum bicolor* (L.) Moench ssp.) for resistance to anthracnose (*Colletotrichum sublineolum*) and grain mold diseases

by

Diriba Hika Chere

B.S., Haramaya University, 2002
M.S., Haramaya University, 2009

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Tesfaye Tesso

Copyright

Abstract

Fungal diseases, such as anthracnose and grain mold pose significant challenges to the production and productivity of sorghum. Favored by warm and humid conditions, sorghum Anthracnose *(Colletotrichum sublineolum*) is among the most devastating diseases of sorghum that can cause yield losses of up to 50% in susceptible cultivars. It can attack all above-ground parts including foliage, stalk, panicle, and grain. The severity of infection depends on the genotype, the pathotype, and the environment. Grain mold is another major sorghum disease that is prevalent under warm and wet conditions that occur during grain development. This disease is caused by one or more genera of fungi of which *Aspergillus* and *Fusarium* are the most common. Breeding efforts have focused on the development of resistant varieties that can reduce the impact of these diseases. Conventional breeding supported by genomic tools may enhance this effort and accelerate the release of new resistant cultivars. To contribute to this endeavor three experiments were initiated and executed. This dissertation thus presents the results of a comprehensive investigation into enhancing resistance to sorghum anthracnose and grain mold diseases through genomic selection, genome-wide association studies (GWAS), and combining ability analysis.

The first chapter explores the utilization of genomic selection techniques to improve anthracnose and grain mold resistance in sorghum through prediction using genomic estimated breeding values (GEBVs). Through the integration of advanced genomic tools and phenotypic data, we demonstrated the potential for accurately predicting resistance levels in sorghum lines, that may facilitate the development of more resilient cultivars. During the 2015 and 2016 cropping seasons, a study was conducted in Bako, Ethiopia, involving 946 sorghum accessions obtained from the gene bank. These accessions were examined for resistance to sorghum leaf anthracnose (*Colletotrichum sublineolum*) and grain mold. Genotyping of these accessions was performed

using the genotype-by-sequencing (GBS) method, resulting in the generation of 410,717 high-quality SNP markers. To investigate the impact of SNP density on prediction accuracy, the markers were pruned using TASSEL software, maintaining marker distances of 1,000 bp, 50,000 bp, 75,000 bp, and 100,000 bp between them. The analysis involved varying the size of the training population from 10% to 90% at intervals of 10%. Principal component analysis was carried out to assess the genetic relatedness patterns among the sorghum accessions. For anthracnose resistance, we obtained a prediction accuracy of 0.38 and 0.45 for the 10% and 90% training population sizes respectively for the 2016 data. A prediction accuracy of 0.38 and 0.41 were recorded for grain mold at 10% and 80% training population sizes.

In the second experiment, a GWAS study was conducted to identify key genomic regions and candidate genes associated with sorghum anthracnose resistance. Our analysis unveiled the presence of Single Nucleotide Polymorphisms (SNPs) on several chromosomes, specifically chromosomes 1, 2, 3, 4, 6, 8, and 10. Notably, chromosomes 8 and 1 exhibited a higher abundance of significant SNPs. Chromosome 8 displayed the most pronounced genetic associations, with nine specific loci (S8_16230498, S8_16389715, S8_37466666, S8_39664335, S8_42468301, S8_42983043, S8_43776246, S8_7908407, and S8_570182) having significant associations with anthracnose disease reaction. Similarly, we identified multiple loci on chromosome 1 (S1_20098885, S1_49271328, S1_49271328, S1_49454848, S1_50111275, S1_50294000, S1_50533418, S1_30455882, and S1_50721682) that also have significant association with the disease. Chromosome 2 revealed two significant loci (S2_55407842 and S2_69275221), while only one locus on chromosome 3 (S3_53970498) showed significant association. Two loci each on chromosomes 4 (S4_62789639, S4_7538383) and chromosome 6, (S6_25755784, and S6_55783758) were shown to have significant association with the disease. By leveraging high-

throughput genotyping data and extensive phenotypic evaluations, this research provides valuable insights into the genetics of anthracnose resistance, offering a foundation for marker-assisted breeding efforts.

The third experiment explored the combining ability of sorghum lines for anthracnose resistance. By assessing the genetic interactions among different sorghum parental lines, this study elucidated the potential for breeding strategies that harness the complementary genetic contributions of parental lines to enhance resistance traits in the progeny. We conducted the experiment across five different environments and collected data on various agronomic traits, including flowering time (DTF), plant height (PTH), panicle weight (PKW), grain yield per panicle (GYP), hundred-grain weight (HGW), and resistance to anthracnose (ANT). Our findings indicated that there were significant variations in trait reactions across different replicates and environmental conditions, particularly for PTH, GYP, and ANT due to notable environmental effects. The entry effect, which represents the contributions of different parental lines and hybrid combinations, was highly significant for all traits we examined. Additionally, the interaction between male and female parents (M vs. F) showed a high level of significance specifically for ANT. A negative significant correlation was observed between panicle weight and anthracnose and grain yield per panicle and anthracnose. Through our combining ability analysis, we were able to identify the genetic contributions of various genotypes. Notably, certain parents such as R245, R604, R633, A539, and A553 exhibited high general combining ability (GCA) for anthracnose resistance. Furthermore, some hybrid combinations like A539 × R669, A527 × R672, A531 × R672, A539 × R588 demonstrated highly significant specific combining ability (SCA) effects for anthracnose resistance. These findings provide valuable insights for breeders, enabling them to make well-informed decisions when designing breeding strategies.

Collectively, this thesis represents a multidisciplinary approach to addressing the pressing issue of anthracnose and grain mold resistance in sorghum. The findings contribute to the development of more resilient sorghum cultivars, thus supporting global food security and sustainable agriculture practices.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am humbled to start by expressing my gratitude to the Almighty God, the source of all blessings and guidance. He has been with me in every endeavor and every moment of my life.

To my supervisor, Dr. Tesfaye Tesso, I extend my sincere thanks for your invaluable guidance and expertise. I am grateful for the mentorship and for the opportunities you have provided for my academic growth. I am thankful for the contributions of my committee members, Dr. Ramasamy Perumal, Dr. Christopher Little, and Dr. Sanzhen Liu.

I am thankful to my colleagues Dr. Kinde Ketema, Dr. Yemane Girma Dr. Jebril Jebril, and Mr. Daniel Hopper for their camaraderie and insightful contributions. As one! I am also thankful to Dr. Kebede Muleta for the support and encouragement. Dr. Gezahegn Girma and Dr. Habte Nida thank you for being available to provide information used in the research.

I appreciate the support from Yela farm, Puerto Rico, Mr. Hugo Cuevas (USDA, Tropical Agriculture Research Station, Puerto Rico), Dr. Joseph E. Knoll (USDA ARS-Tifton, GA), Researchers from Ethiopian Institute of Agricultural Research (EIAR), Oromia Agricultural Research Institute (OARI), Bako, and all the participants and individuals who generously contributed to this research project.

Mom, though you never went to a formal education you always wanted your kids to pursue their education to the highest level by investing your life in us. I have tried it and I believe you are looking down on me with a smile.

To my wife, you encouraged me to start this journey and have always been my courage. Taking care of us all hasn't been an easy task. I am grateful for the support, love, and encouragement that have been my pillars of strength. Thank you and love you Giiftii!

To our kids (Ketyeron, Naflet, Keol and Hawinan), I am grateful to have you in my life. Every time I leave home at night and on weekends, you always asked me when I would get time to spend with you. I think I will have some time now. Thank you for the love and the patience you have shown me, much love!

Dedication

Dedicated to my mom and all the Oromo heroes.

Chapter 1 - Literature Review

Sorghum is a C4 plant known to efficiently utilize solar energy and hence adapted to warm tropical environments (Sage and Zhu, 2011). It belongs to the grass family *Poaceae*, tribe *Andropogoneae*, subtribe *Sorghinae*, and genus *Sorghum* Moench (Clayton and Renvoize, 1986). The name *Sorghum bicolor* (L.) Moench was first proposed by Clayton in 1961 and remains the preferred name for cultivated sorghum.

Globally, sorghum [*Sorghum bicolor* (L.) Moench] ranks fifth in total production after maize, wheat, rice, and barley (Figure 1.1). It is a staple food for over 500 million people worldwide (Lindsay, 2010). Almost 89 million tons of sorghum was produced on about 40 million ha in 2020, the average yield being 1.45 t ha^{-1} (FAO, 2020).

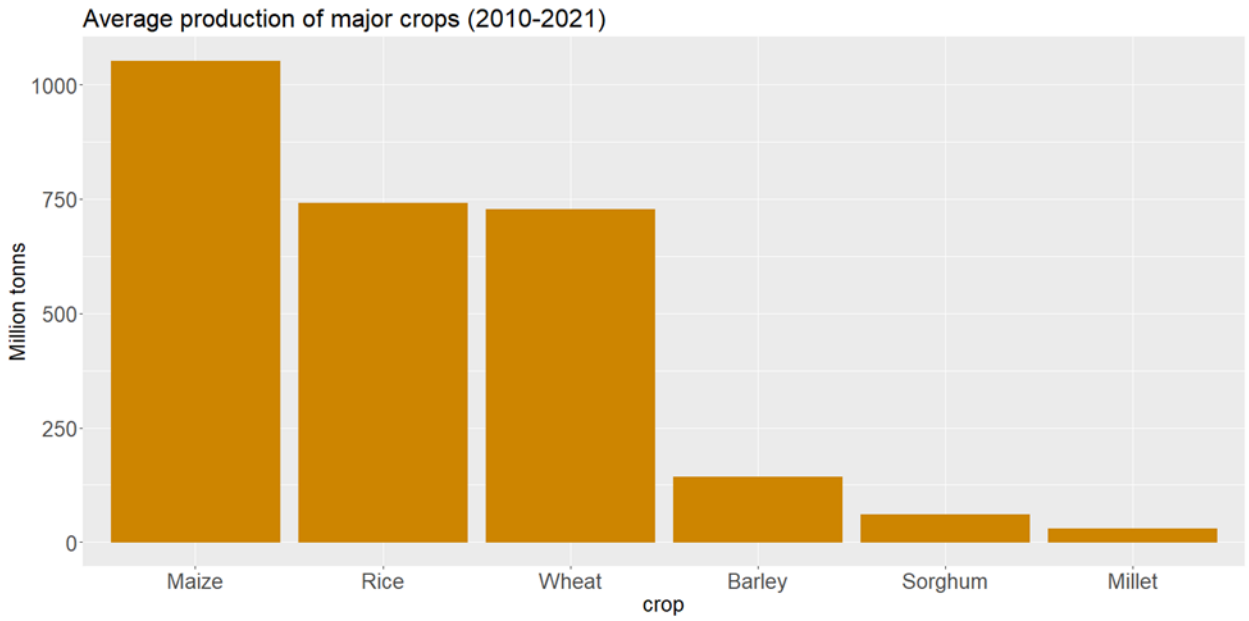


Figure 1.1. Average production of major cereal crop for the year 2010-2021.

The United States of America is the leading producer of sorghum with 9.47 million tons produced in 2020. Sorghum ranks third in Africa, after maize and rice (FAO, 2020). Ethiopia and Nigeria are two major producers in Africa and the second and third largest producers in the world with a production of 6.36 and 5.05 million tons, respectively (Figure 1.2).

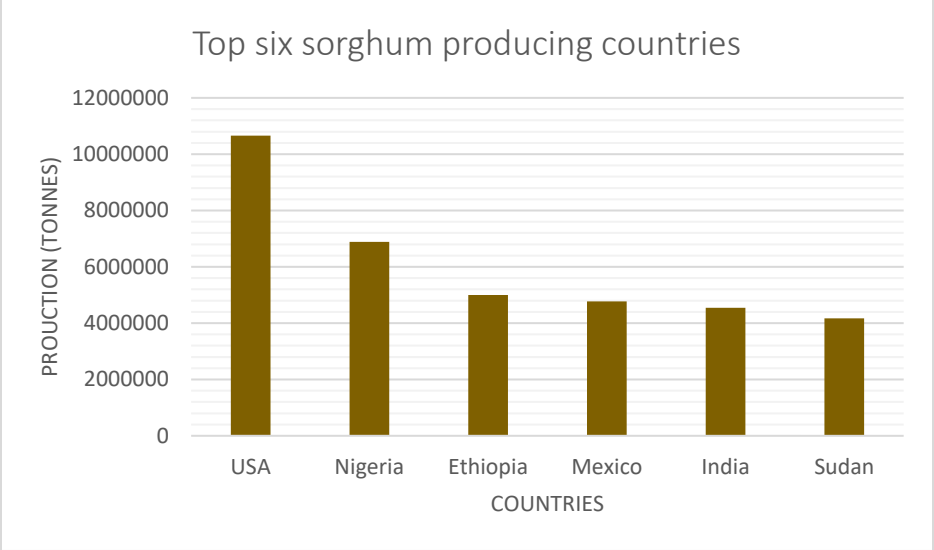


Figure 1.2. Average production of sorghum by the top ten producers from 2015-2020 (Source: FAOSTAT, 2020).

1.1 Origin, domestication, and diversity of Sorghum

Sorghum is believed to have been domesticated in Africa some 7000 years ago (Doggett and Prasada Rao, 1995), and hence most of the diversity is found there (Reddy et al., 2006). This variability can be observed both in cultivated and wild species (Gebrekidan, 1982). The northeast Africa region is recognized as the center of origin and diversity for sorghum and the region still harbors novel accessions that have immense potential for improvement of the crop. The varied climate, geography, and impact of human and natural selections led to the evolution and adaptation of diverse types of the crop in the region (Doggett, 1988).

Cultivated sorghum is grouped into five basic races: Bicolor, Kafir, Guinea, Caudatum, and Durra (Harlan and de Wet, 1972). Each of these has specific adaptation regions and distinct features that help in identification. Bicolor is the most primitive form widely distributed in Africa and Asia and is characterized by its open inflorescences with long clasping glumes that cover the grain at maturity. Kafir, grown in subequatorial Africa, is characterized by its symmetrical grains that are spherical, and the glumes, unlike that of the bicolor race, do not enclose the grain. Another race, Guinea, which is characterized by long obliquely twisted glumes, is predominant in West Africa. Caudatum originated recently in central Africa and has an asymmetrical grain. Durra races are predominant in Near East, India, and Pakistan. In Africa, it is distributed in the mid-altitude areas of Ethiopia, the Nile valley of Egypt, and Sudan. It exists in a belt between 10-15° north latitude extending from Ethiopia in the east to Mauritania to the west (Mann et al., 1983). These races are interfertile and thus led to ten intermediate races that exhibit the characteristics of their parents.

Sorghum is one of the major crops represented by many germplasm collected across the world. The centralized collection to conserve and make the germplasm resources available to users started in Ethiopia between 1958 and 1960 (Rosenow and Dahlberg, 2000). This effort continued in the early 1970s by the Ethiopian sorghum improvement project. Among the Ethiopian collection zera zera, durra, and durra-bicolor are distinct forms. Zera zera is an important group that has been used extensively in the U.S. and India to develop a disease resistant, high yielding and good grain quality varieties (Rosenow and Dahlberg, 2000).

1.2 Sorghum in Ethiopia

Sorghum is one of the staple foods in the most food-insecure dryland area of Ethiopia where other cereals totally fail or poorly perform. Out of a total of about 10 million hectares under

cereals in 2020/21 crop season, sorghum was planted to 1.79 million hectares representing over 17% of the acreage under cereals (CSA, 2021). The average yield of the crop is 2.8 tons ha^{-1}. In 2021, a total of 5.1million tons of grain were harvested which represents 18% of total cereal grain output (CSA, 2021). Over the last two decades, area under sorghum has increased by 78% from about 1 million ha in 2000 to 1.79 million ha in 2020 (FAO, 2020).

Sorghum is mainly grown by subsistence farmers who process the grain into a fermented bread called '*Budena*' in Afaan Oromo, and '*Injera*' in Amharic. The crop is also used to make porridge and traditional homemade beverages. Sorghum tolerates harsh environmental conditions, and this ability makes it an important crop in these environments. Sorghum has served as a primary source of food for people living in the semi-arid tropics of Africa and Asia. In Ethiopia, sorghum is grown under diverse environmental conditions from highland, mid-altitude to lowlands, and different races tend to occur in different agroecologies although there is evidence of occurrence of multiple races in each area. For example, race bicolor dominates the west banks of the Rift Valley while durra sorghums are grown in the eastern high and lowlands of the country (Stemler et al., 1977).

1.3 Sorghum production challenges

Among cereals, sorghum is known to suffer the most from biotic and abiotic stresses such as parasitic weed *Striga*, fungi, bacteria, viruses, nematodes, armyworms, shoot fly, local and migratory birds, and drought and nutrient deficiency. Although sorghum seem to fair well with regards to diseases, several fungal pathogens cause serious diseases to the crop including grain mold, anthracnose, leaf blight, downy mildew, charcoal rot, rust, ergot, and smut (Thakur et al., 2007). This study is focused on two major fungal diseases (anthracnose and grain mold) that are most destructive in the humid western part of Ethiopia.

1.3.1 Sorghum foliar anthracnose

Anthracnose diseases in cereals and grasses are caused by four distinct Colletotrichum species. Sorghum foliar anthracnose (*Colletotrichum sublineolum*), which was previously referred to as *Colletotrichum graminicola*, and believed to cause anthracnose both in maize and sorghum, is one of the major diseases. Even though *Colletotrichum graminicola* and *Colletotrichum sublineolum* have morphological similarities and close evolutionary relationship, studies on appressorial morphology, rDNA sequence analysis, mating, and restriction fragment length polymorphism (RFLP) have confirmed that they are reproductively isolated and belonged to two separate species (Vaillancourt and Hanau, 1992; Sherriff et al., 1995; Wharton and Julian, 1996). *C. graminicola* causes anthracnose disease in maize while *C. sublineolum* (Crouch et al., 2006) is responsible for sorghum anthracnose disease (Vaillancourt and Hanau, 1992; Sherriff et al. 1995). *C. sublineolum* is favored by warm and humid climatic conditions and is known to have high variability resulting in the presence of several pathotypes (Rosewich et al., 1998; Marley et al., 2001; Valèrio et al., 2005). More than 40 races or pathotypes of the pathogen were reported in different environments. The major diseases caused by these pathogens are seedling blight, stalk rot, leaf blight, head blight, and grain mold. These diseases limit both grain and forage production causing yield losses as high as 50%, with susceptible varieties even suffering more (Harris et al., 1964; Cota et al., 2017a). Affected leaves experience a reduction in photosynthetic area, while stem infection reduces transportation. When grain is infected, it will have a direct effect on yield and quality.

Sorghum foliar anthracnose is typified by a small circular lesion with red, purple, tan, or black discoloration depending on the host pigmentation (Frederiksen, 1986; TeBeest et al., 2004; and Mathur, 2000). The lesions may enlarge under conditions that favor the pathogen and the

plants eventually become defoliated because of the pathogen. *C. sublineola* is a hemibiotrophic pathogen; it begins infection as biotrophic and later transitions to a necrotrophic lifestyle (Münch et al., 2008; Wharton et al., 2001; Wharton and Julian, 1996). Soil, seeds, and decomposing plant residues all serve as potential media on which the pathogen can overwinter in the form of mycelium, acervuli and sclerotia (Casela and Frederiksen, 1993). It is one of the most economically devastating diseases of sorghum, causing yield losses of up to 50% mainly by reducing kernel size and kernel numbers per head (Ali and Warren, 1992; Thakur and Mathur, 2000).

1.3.2 Sorghum grain mold

Most sorghum growing regions of the tropics and subtropics are hot and dry thus have less disease problems. However, there are also major sorghum growing areas such as in western Ethiopia that are hot and humid that provides ideal condition for disease development. Favored by high humidity and warm temperature, grain mold is the major yield and quality limiting disease of sorghum. The disease can cause yield losses of 30-100% depending on the cultivar and the prevailing environmental conditions (Singh and Bandyopadhyay, 2000). Grain mold pathogens exhibit various trophic behaviors including necrotrophic, saprophytic, and hemibiotrophic. Hemibiotrophic pathogens start as biotrophic (parasitism on living hosts) and shift to the necrotrophic phase later. Necrotrophic fungi kill the host tissue and thrive on dead plant tissues while saprophytic fungi feed on decaying organic matter (Cooke and Rayner, 1984). The disease is caused by a multitude of fungi of various species including *Fusarium* spp., *Alternaria* spp., *Aspergillus* spp., *Curvularia* spp., *Colletotrichum* spp., *Phoma* spp., *Bipolaris* spp., *Aspergillus* and *Exserohilum* spp. (Forbes, 1986; Reddy et al., 2000; Thakur and Indira, 2006; Little et al., 2012). These fungal complexes act at various points during the vegetative growth stage of the

sorghum plant as well as during the grain filling. The infection starts at anthesis and progresses until physiological maturity. The disease results in reduction of caryopsis formation, arrest of kernel development, and decrease in grain mass and grain density may occur when the infection starts early (Little and Magill, 2009). The disease doesn't cease to cause damage at maturity; it continues to deteriorate grain quality at which stage it is referred to as grain weathering (Das et al., 2020). Grain weathering is the discoloration of grains and damage of tissues caused by fungal colonization under wet conditions (Thakur and Indira, 2006). Different fungal species act under different conditions and growth stages. *Fusarium verticillioides* predominantly occur under field condition. It exhibits both biotrophic and necrotrophic lifestyles, and it infects the developing panicles through areal spores (Williams and Rao, 1981). *Aspergillus spp.* is a dominant post-harvest pathogen (Cuevas et al., 2019). It is soil-borne and exhibits a saprophytic lifestyle and infect the developing caryopsis through wounds inflicted by insects (Pfliegler et al., 2020). *Curvularia lunata* is common in drier areas and is known to cause damage during germination (Prom et al., 2016). *F. moniliforme* sensu lato and *C. lunata* infections interfere with carbohydrate translocation and cause reduction in seed size. *Colletotrichum sublineolum,* which also causes anthracnose disease in sorghum, is another grain mold pathogen prevalent in humid and warmer environments.

Different fungal species develop different mold colors: *Fusarium moniliforme* produces mold colors ranging from pinkish white to orange-white. On the other hand, *Curvularia* develops shiny, velvety black, fluffy growth, and *Phoma* spp. produces black discoloration (Bandyopadhyay, 1986).

1.4 Foliar anthracnose and sorghum grain mold diseases in Ethiopia

Several studies in Ethiopia reported high incidence and severity of anthracnose in the southern and western parts of the country where higher humidity and warmer temperatures prevail (Chala et al., 2011; Tsedaley et al., 2016). A survey on anthracnose incidence in the southern and Southwestern parts of the country reported severe occurrence of the diseases in all surveyed areas (Chala et al., 2010b). A separate study conducted in Southern Tigray reported anthracnose infection in about 98% of the fields with 53% severity (Teklay, 2015). Another study reported rainfall as the significant factor aggravating anthracnose incidence while temperature had less significance (Chala et al., 2010a) . However, a greenhouse study conducted in India showed a high anthracnose severity at 25°C in both susceptible and resistant lines (Pande et al., 1994).

Sorghum grain mold is another economically important disease in Ethiopia. Terefe, (2002) reported *Fusarium proliferatum* as the most frequently isolated species in Ethiopia while *Alternaria alternata* had the highest incidence in almost all sorghum growing regions.

1.5 Resistance mechanisms to foliar anthracnose and grain mold diseases

Several factors contribute to resistance to grain mold. Some of these are morphological characters, distribution and quantity of secondary metabolites, and presence, type, and quantity of antifungal proteins in the seed endosperm (Ambekar et al., 2011). Plant morphological and biochemical characteristics such as panicle compactness, grain pigmentation, presence of a pigmented testa, glume coverage, grain hardness, pericarp thickness, and endosperm texture have all been reported to contribute to grain mold resistance (Brown et al., 2006; Menkir et al., 1996; Williams and Rao, 1981). A more recent study attributed resistance to grain mold disease mainly to panicle compactness and grain pigmentation (Sharma et al., 2010). Sorghums with open panicle structures tend to be more resistant to grain mold than those with compact panicle structures. This

appears to be resulted from better air circulation and less moisture retention in open panicles. However, not all cultivars with open panicles are resistant to grain mold or *vice versa*. Hence, resistance based on panicle structure is elicited more from the morphological attribute than the resistance genetics. Another important trait for grain mold resistance is grain pigmentation. It is reported that sorghums with pigmented grains are less affected by grain mold (Sharma et al., 2010). This type of resistance is governed by dominance and epistatic interaction (Audilakshmi et al., 2000). On the other hand, polygenes with a significant additive by additive interaction and genotype by environment (G x E) interaction effects have been reported to be associated with grain mold resistance in the white-grain sorghums (Rodriguez-Herrera et al., 2000). The use of grain pigmentation is prone to bias as grain mold can be easily recognized on white seeds as opposed to pigmented pericarps. Therefore, disease scoring of multi-colored test genotypes needs a great deal of care and attention (Ackerman et al., 2021).

Grain mold resistance is polygenic in nature, and it involves up to 10 major and minor genes with additive and epistatic effects. The effect of G × E interaction is also highly significant (Audilakshmi et al., 2000; Rodriguez-Herrera et al., 2000).

Studies on foliar anthracnose disease resistant also reported both dominant and recessive mechanisms. Earlier study conducted by Coleman and Stokes, (1954) reported anthracnose resistance to be controlled by the action of two closely linked loci with dominant effects. More recent studies on sorghum lines SC112-14, BTx378, and SC748-5 reported anthracnose disease response as conditioned by dominant genes (Mehta et al., 2005; Erpelding, 2007; Cuevas et al., 2014). On the other hand, a recessive response has also been reported in the sorghum line SC326-6 (Boora et al., 1998). Preliminary studies conducted in our program also show that different

resistance mechanisms come in to play with resistance in different backgrounds seeming to be under the control of dominant, partially dominant, and recessive genes.

1.6 Screening techniques for anthracnose and grain mold resistance

Screening for anthracnose and grain mold can be conducted in the field under natural infection condition or through artificial inoculation. For grain mold, the occurrence of frequent rainfall during grain development stage is important (Forbes et al., 1992; Williams and Rao, 1981). To obtain best results, the screening needs to be carried out at physiological maturity as this can help capture late colonizers (Forbes et al., 1992). Grain mold disease can be scored either by qualitative or quantitative methods. The qualitative method simply categorizes the genotypes as immune, resistant, moderately resistant, susceptible, and highly susceptible. On the other hand, the quantitative method utilizes a scale of 1-5 or 1-9 to score disease reaction where a scale of '1' means resistant and '5' or '9' means highly susceptible depending on the scale employed (Bandyopadhyay and Mughogho, 1988).

Like grain mold, anthracnose disease is also favored by warm and humid environments. Disease symptoms may start to appear 30-40 days after emergence and well extend through grain filling stage (Thakur et al., 2007). The inoculums from crop residue are the primary source of infection for seedlings.

Sorghum anthracnose screening can also be conducted both in the field and under greenhouse conditions. For field screening, the test materials can be planted in an environment that is a hotspot for the disease. The screening can also be done by artificial inoculation. In this case, a known susceptible line that serves as a spreader is planted in combination with test genotypes. The susceptible genotypes are planted on designated rows (spreader rows) at a pre-determined interval, often every nine rows of the test genotypes. Two to three sorghum grains inoculated with the

desired pathogen strain is placed in the whorl of the spreader plants to facilitate infection (Thakur et al., 2007; Prom et al., 2009). Five to ten plants of the test genotypes are then tagged, and the disease data recorded on them. Qualitative disease scores are rated as highly resistant (HR), resistant (R), moderately resistant (MR), susceptible (S), and highly susceptible (HS) depending on the symptoms observed (Thakur et al., 2007; Prom et al., 2009). Anthracnose severity can also be scored based on a scale of 1-9 or 1-5. A scale of '1' represents highly resistant genotypes in both scale categories. A rating of '2' or '3' is assigned to indicate resistance in the 1-9 scale while '4' and '5' are given to moderately resistant genotypes. Susceptible entries will be scored as '6' or '7' depending on the severity. Those which are highly susceptible will be given a score of '8' or '9'. On the 1-5 scale, the moderately resistant genotypes are scored '3' while those susceptible are assigned a score of '4'. Those found to be highly susceptible are given a score of '5' (Thakur et al., 2007; Prom et al., 2009).

1.7 Genomic selection

Marker-assisted selection has a valuable use for improving traits controlled by one or few genes (qualitative traits) (Ragot et al., 1994). But its application is limited when it comes to traits controlled by several genes that have small effects (quantitative genes). Polygenic traits affected by several genes of minor effect are better targeted through genomic selection (GS). Genomic selection is a marker-based method used to determine associations between markers across the genome and a trait phenotype (Meuwissen et al., 2001). Genomic selection involves two sets of populations: the training population, a set that has both phenotypic and genotypic information, and the testing population, one that has genotypic information but no phenotypic data (Meuwissen et al., 2001). GS is a technique that utilizes associations between several markers across the whole

genome and the trait values from the training set for calculating genomic estimated breeding values (GEBVs) of untested genotypes.

The accuracy of genomic prediction model depends on the size of the training and testing (breeding) population. In addition, the heritability of the traits, the effect of the environment (G x E interaction), genetic diversity and marker density, and the relationship between the training and the testing set are crucial factors affecting the accuracy of the genomic prediction (De Roos et al., 2009; Howard et al., 2014; Wang et al., 2018; Zhang et al., 2019).

The accuracy of genomic selection also heavily depends on the choice of models. Several prediction models have been made available and among them are, Ridge Regression–Best Linear Unbiased Prediction (RR-BLUP), Genomic Best linear unbiased prediction (GBLUB) (VanRaden, 2008), BayesA, BayesB, BayesC, and least absolute shrinkage selection operator (LASSO) (Tibshiranit, 1996; Kukreja et al., 2006) with BLUP being the most widely used.

The model used for genomic selection as outlined by Heslot et al., 2012 is:

$Y = \mu + X\beta + \varepsilon$

Where Y represents the trait value, μ is the population mean, X represents the marker design matrix, β is the vector of marker effect, and ε is the error term, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

Prediction accuracy is calculated as the Pearson's correlation between the GEBV and the true breeding value as:

$r_A = \sqrt{\frac{h^2}{h^2 + \frac{M_e}{N_p}}}$ where r_A denotes the prediction accuracy, h^2 is the narrow sense heritability, M_e is the number of independent chromosome segments and N_p is the number of individuals in a training population (Daetwyler et al., 2008, 2010).

Higher prediction accuracies were reported with higher marker densities with the GBLUP (Zhang et al., 2019). For traits with high or medium heritability, BayesR model was found to be better than the GBLUP in predicting accuracy (Zhang et al., 2019).

Since its introduction in 2001 (Meuwissen et al., 2001), genomic selection has been successfully applied in animal breeding and is slowly gaining momentum in crop research through works on a few crops such as maize (Bernardo and Yu, 2007; Beyene et al., 2021; Riedelsheimer et al., 2012; Shikha et al., 2017; Zhao et al., 2012), rice (Spindel et al., 2015; Cui et al., 2020; Xu et al., 2021), barley and sorghum (Habyarimana and Lopez-Cruz, 2019; Hunt et al., 2018a; Muleta et al., 2019; Puglisi et al., 2021; Velazco et al., 2020; Zhong et al., 2009).

In a study conducted on a doubled haploid (DH) maize, genomic selection improved response over marker-assisted recurrent selection (MARS) by 18-43% (Bernardo and Yu, 2007). In a separate study, prediction accuracy of 0.19 to 0.31 was obtained for grain yield under different conditions using 3,068 tropical maize DH lines (Beyene et al., 2021). In the same study, a prediction value of 0.53-0.71 was reported by including breeding data from the previous year and by including 30% of the testing population in the training set (Beyene et al., 2021). In rice, a predictive value of 0.31 was reported for grain yield in a panel of 363 elite breeding lines. For plant height and flowering time, the values were 0.34 and 0.63, respectively (Spindel et al., 2015). Genomic prediction studies in sorghum mainly focused on bioenergy sorghums and on genomic prediction model comparisons (Table 1.1).

Table 1.1. Previous works on genomic selection in sorghum

Population	SNPs	Statistical model	Trait	predictive value	Reference
Multi-parental random mating population	146,306 SNPs	GBLUP	Grain yield Plant height Al tolerance	0.28 0.53 0.45	(Bernardino et al., 2021)
869 biomass sorghum lines	100,435 SNP markers	Bayesian network, Pleiotropic Bayesian network, Dynamic Bayesian network, multi-trait GBLUP (MTr-GBLUP), and multi-time GBLUP	plant height dry biomass yield	0.47-0.75 0.46-0.49	(Dos Santos et al., 2020)
Pre-breeding population of biomass sorghum	59,264 SNPs	GBLUP	biomass yield	0.40	(Fernandes et al., 2018)
Sorghum diversity panel and landraces and *S. bicolor* × *S. halepense* RILs	61,976 SNPs	GBLUP, BRR, Bayesian LASSO and BayesB	polyphenol, flavonoids, condensed tannins, total antioxidant concentrations	0.49- 0.58	(Habyarimana and Lopez-Cruz, 2019)
180 Sorghum bicolor landraces and lines and 189 *S. bicolor* × *S. halepense* advanced recombinant	61,976 SNPs	BLUP	aboveground biomass yield, plant height, and dry mass fraction	0.36-0.59	(Habyarimana et al., 2020)

Population	SNPs	Statistical model	Trait	predictive value	Reference
inbred lines					
Preliminary yield Trials sets	581 DArT markers		Grain yield	0.12-0.59	(Hunt et al., 2018b)
400 hybrids	6,619 markers	GBLUP, GBLUP-AD, GK,	Brix, culm length, and fresh weight		(Ishimori et al., 2020)
238 Chibas training set and 2,736 other individuals			height, brix, juice weight, leaf weight, earliness, stem weight, and grain yield	0.57–0.73	(Jensen et al., 2020)
200 sorghum landraces and breeding lines	258,220 markers	rrBLUP	days to flowering, plant height, fresh and dry matter yield, and fiber, cellulose, hemicellulose, and lignin proportion	0.66-0.85	(de Oliveira et al., 2018)
281 maize inbred Lines and 320 sorghum inbred lines	51,742 SNP Maize, 90,441 SNPs sorghum	rrBLUP			(Rice and Lipka, 2019)
389 diverse sorghum accessions	224,007 SNPs	GBLUP	panicle branch length, days to anthesis, flag leaf height,	0.52–0.69	(Sapkota et al., 2020)

Population	SNPs	Statistical model	Trait	predictive value	Reference
			grain number per primary panicle, thousand grain weight, grain yield per primary panicle, plant height, panicle length		
2645 testcross hybrids and their female parents	4781 SNPs	GBLUP	grain yield, stay-green, plant height, and flowering time	0.35-0.48	(Velazco et al., 2019)
Testcross evaluations of 603 female parental lines		KBLUP			(Velazco et al., 2020)

1.8 Genome-Wide Association Study (GWAS)

Marker-trait associations in disease resistance study are detected by examining the co-segregation of disease phenotype and alleles at marker loci. Nevertheless, this method has its own limitations in addressing complex diseases which are influenced by genetic variations at several loci in the genome. Hence, studying diseases at population level by linkage disequilibrium (LD) mapping helps in detecting and locating quantitative trait loci (QTL) based on the strength of the correlation between a marker and a trait (Mackay and Powell, 2007). The concept of linkage disequilibrium is based on studies by (Geiringer, 1944; Lewontin and Kojima, 1960) and refers to

non-random association between alleles at different loci. Linkage disequilibrium occurs when alleles at separate loci on the same chromosome are inherited together more frequently than anticipated by chance.

LD mapping, also known as association mapping, exploits historical recombination events at population level for dissecting complex traits (Lewontin and Kojima, 1960; Nordborg and Tavaré, 2002). Variants or single-nucleotide polymorphisms (SNPs) that are associated with these complex traits can be detected with genome-wide association analysis.

The following linear regression model is used in GWAS analysis (Yu et al., 2006):

$$y = X\beta + Zg + S\tau + \varepsilon$$

where y is a vector of phenotypic observation; $X\beta$ is fixed effects; g models the genetic background of each line as a random effect; τ is a vector of additive SNP effects as a fixed effect; ε a vector of residual effects. X , Z and S are incidence matrices of 1s and 0s relating y to β , g and τ , respectively.

Linkage disequilibrium undergoes gradual deterioration over successive generations due to recombination events. Recombination occurs as alleles from different parental chromosomes are rearranged during the formation of gametes. The associations between alleles disrupt when recombination occurs between loci in proximity leading to linkage disequilibrium decay. Factors that influence linkage disequilibrium decay are population type, recombination hotspots, physical distance, mutation rate and natural selection. In maize, LD decay was reported within 1 kb in landraces, in approximately 2 kb in diverse inbred lines, and up to 100 kb in commercial elite inbred lines (Remington et al., 2001; Tenaillon et al., 2001; Ching et al., 2002).

In sorghum, LD over short distances was reported to be more extensive than in maize (Hamblin and Mitchell, 2014). This may favor LD mapping of genes responsible for complex traits.

1.9 Combining ability analysis

The use of combining ability as a measure of potential of an individual as breeding parent has taken root in breeding programs. The concept was first introduced by Sprague and Tatum, (1942). The relative performance of a line when crossed with several other lines is referred to as its general combining ability (GCA), while the specific combining ability (SCA) refers to the performance of a hybrid relative to the performance of its parents in other crosses (Sprague and Tatum, 1942). Various mating designs and analysis methods have been developed to allow estimation of GCA, SCA or both. Breeders utilize one or more of these designs and methods that best suit the specific crops they work with. Because GCA measures the average performance of an individual based on the performance of its progenies, it directly reflects the impacts of alleles shared by the individual regardless of the impact of alleles coming from the other parent. Since only additive genes are transferred from parents to offspring, then GCA can be ascribed as the measure of additive genetic effect. On the other hand, SCA is a measure of the performance of a given cross relative to the parents' performance in other crosses. The cross can be superior to either of the parents, hence the performance is not only because of the combined effects of genes from the parents but also because of interaction between genes contributed by the parents. This interaction occurs at each locus affecting a trait which in genetics terms is called dominance. Hence SCA is simply a measure of dominance genetic effect. Both additive and dominance genetic effects can contribute to a given trait and one may have bigger contribution than the other. Hence, the gene action responsible for GCA is additive and additive x additive epistasis and is considered

fixable while that of SCA is due to non-additive gene action, dominance, and epistasis, or both and is non-fixable (Falconer, 1989). Likewise, variance estimates due to GCA can be considered as additive genetic variance while those due to SCA as dominance variance. Thus, to run a successful breeding program, information regarding GCA of parental genotypes and SCA of a cross from certain parental genotypes is essential. In hybrid breeding programs, it is part of a routine practice to determine the GCA of new breeding lines as soon as they are developed. Even in self-pollinated crops, knowledge of the GCA of potential parents is useful.

A higher variance of SCA than GCA were reported in rice for yield and yield-related traits except for plant height stipulating preponderance of non-additive gene action over that of additive action (Bagheri and Jelodar, 2010). In a study on 15 sorghum restorer lines (male) and 5 male-sterile A-lines, Kenga et al. (2004) reported highly significant GCA effects for males in all the traits considered: grain yield, days to anthesis, plant height, inflorescence length, threshing percentage, and seed mass. All traits except inflorescence length also showed a significant SCA. This result is in line with what (Mengistu et al., 2010) reported on Ethiopian sorghum genotypes for agronomic traits. Inheritance of grain mold resistance in sorghum was reported to be majorly due to additive gene action (Rao et al., 2016).

References

Ackerman, A., Wenndt, A., & Boyles, R. (2021). The sorghum grain mold disease complex: Pathogens, host responses, and the bioactive metabolites at play. *Frontiers in Plant Science*, 12, 660171. Frontiers Media S.A. https://doi.org/10.3389/fpls.2021.660171

Ali, M. E. K., & Warren, H. L. (1992). Anthracnose of sorghum. In Sorghum and millets diseases: a second world review. (Ed.), in: *Milliano, WAJ, Frederiksen, R.A., Bengston, G.D (*Eds.*)* (pp. 203–208). International Crops Research Institute for the Semi-Arid Tropics.

Ambekar, S. S., Kamatar, M. Y., Ganesamurthy, K., Ghorade, R. B., Saxena, U., Chand, P., Jadav, B. D., Das, I. K., Nageshwararao, T. G., Audilakshmi, S., & Seetharama, N. (2011). Genetic enhancement of Sorghum (Sorghum bicolor (L) Moench) for grain mold resistance: II. Breeding for grain mold resistance. *Crop Protection*, *30*(7), 759–764. https://doi.org/10.1016/j.cropro.2010.06.024

Audilakshmi, S., Stenhouse, J. W., & Reddy, T. P. (2000). Genetic analysis of grain mold resistance in colored sorghum genotypes. *Euphytica* (Vol. 116).

Bagheri, N., & Jelodar, B. (2010). Heterosis and Combining Ability Analysis for Yield and Related-Yield Traits in Hybrid Rice. *International Journal of Biology*, *2*(2). www.ccsenet.org/ijb

Bandyopadhyay, R. (1986). Grain mold. In: Frederiksen, R. A., (Ed.), Compendium of sorghum diseases, St. Paul MN, USA, *The American Phytopathological Society*, 36–38.

Bandyopadhyay, R., & Mughogho, K. (1988). Evaluation of field screening techniques for resistance of sorghum grain molds. *Plant Disease*, *72*, 500–503.

Bernardino, K. C., de Menezes, C. B., de Sousa, S. M., Guimarães, C. T., Carneiro, P. C. S., Schaffert, R. E., Kochian, L. V., Hufnagel, B., Pastina, M. M., & Magalhaes, J. V. (2021). Association mapping and genomic selection for sorghum adaptation to tropical soils of Brazil in a sorghum multiparental random mating population. *Theoretical and Applied Genetics*, *134*(1), 295–312. https://doi.org/10.1007/s00122-020-03697-8

Bernardo, R., & Yu, J. (2007). Prospects for genome-wide selection for quantitative traits in maize. *Crop Science*, *47*(3), 1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Beyene, Y., Gowda, M., Pérez-Rodríguez, P., Olsen, M., Robbins, K. R., Burgueño, J., Prasanna, B. M., & Crossa, J. (2021). Application of genomic selection at the early stage of breeding pipeline in tropical maize. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.685488

Boora, K. S., Frederiksen, R., & Magill, C. (1998). DNA-based markers for a recessive gene conferring anthracnose resistance in sorghum. *Crop Science*, *38*(6), 1708–1709. https://doi.org/10.2135/cropsci1998.0011183X003800060048x

Brown, P. J., Klein, P. E., Bortiri, E., Acharya, C. B., Rooney, W. L., & Kresovich, S. (2006). Inheritance of inflorescence architecture in sorghum. *Theoretical and Applied Genetics*, *113*(5), 931–942. https://doi.org/10.1007/s00122-006-0352-9

Casela, C. R., & Frederiksen, R. A. (1993). Survival of Colletotrichum graminicola sclerotia in sorghum stalk residues. *Plant Disease*, *77*(8), 825-827.

Chala, A., Alemu, T., Prom, K., & Tronsmo, M. (2010). Effect of host genotypes and weather variables on the severity and temporal dynamics of sorghum anthracnose in Ethiopia. *Plant Pathology Journal*, *9*(1), 39–46. https://doi.org/10.3923/ppj.2010.39.46

Chala, A., Brurberg, M., & Tronsmo, A. (2010). Incidence and severity of sorghum anthracnose in Ethiopia. *Plant Pathology Journal*, *9*(1), 23–30. https://doi.org/10.3923/ppj.2010.23.30

Ching, A. D. A., Caldwell, K. S., Jung, M., Dolan, M., S (Howie) Smith, O., Tingey, S., ... & Rafalski, A. J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC genetics*, *3*, 1-14. http://www.arabidopsis.org/cereon/].

Clayton, W. D., & Renvoize, S. A. (1986). Genera Graminum: *Grasses of the World*. In *Kew Bulletin Addition Series, XIII, Royal Botanic Gardens, Kew, London*,338–345.

Coleman, O. H., & Stokes, I. E. (1954). The Inheritance of Resistance to Stalk Red Rot in Sorghum. *Agronomy Journal*, *46*(2), 61–63.

Cooke, R. C., & Rayner, A. D. (1984). *Ecology of saprotrophic fungi*. Longman.

Cota, L. V., Souza, A. G. C., Costa, R. V., Silva, D. D., Lanza, F. E., Aguiar, F. M., & Figueiredo, J. E. F. (2017). Quantification of yield losses caused by leaf anthracnose on sorghum in Brazil. *Journal of Phytopathology*, *165*(7–8). https://doi.org/10.1111/jph.12582

Crouch, J. A., Clarke, B. B., & Hillman, B. I. (2006). Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. *Phytopathology*, *96*(1), 46–60. https://doi.org/10.1094/PHYTO-96-0046

CSA. (2021). The Federal Democratic Republic of Ethiopia Central Statistical Agency, Agricultural Sample Survey. *Area and production of major crops*. *Statistical Bulletin*, Addis Ababa.

Cuevas, H. E., Fermin-Pérez, R. A., Prom, L. K., Cooper, E. A., Bean, S., & Rooney, W. L. (2019). Genome-wide association mapping of grain mold resistance in the US sorghum association panel. *The Plant Genome*, *12*(2), 180070. https://doi.org/10.3835/plantgenome2018.09.0070

Cuevas, H. E., Prom, L. K., & Erpelding, J. E. (2014). Inheritance and molecular mapping of anthracnose resistance genes present in sorghum line SC112-14. *Molecular Breeding*, *34*(4), 1943–1953. https://doi.org/10.1007/s11032-014-0151-y

Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Zhang, Q., Ali, J., Li, Z., & Xu, S. (2020). Hybrid breeding of rice via genomic selection. *Plant Biotechnology Journal*, *18*(1), 57–67. https://doi.org/10.1111/pbi.13170

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, *185*(3), 1021–1031. https://doi.org/10.1534/genetics.110.116855

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, *3*(10). https://doi.org/10.1371/journal.pone.0003395

Das, I. K., Aruna, C., & Tonapi, V. A. (2020). Sorghum grain mold. *ICAR-Indian Institute of Millets Research, Hyderabad, India*, 1-86.

de Oliveira, A. A., Pastina, M. M., de Souza, V. F., da Costa Parrella, R. A., Noda, R. W., Simeone, M. L. F., Schaffert, R. E., de Magalhães, J. V., Damasceno, C. M. B., & Margarido, G. R. A. (2018). Genomic prediction applied to high-biomass sorghum for bioenergy production. *Molecular Breeding*, *38*(4). https://doi.org/10.1007/s11032-018-0802-5

De Roos, A. P. W., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, *183*(4), 1545–1553. https://doi.org/10.1534/genetics.109.104935

Doggett, H., & Prasada Rao, K. E. (1995). Sorghum (*Sorghum bicolor*). In J. Smartt & N. Simmonds (Eds.), *Evolution of crop plants*.

Doggett, Hugh. (1988). *Sorghum* (2nd ed.). Longman Scientific & Technical.

Dos Santos, J. P. R., Fernandes, S. B., McCoy, S., Lozano, R., Brown, P. J., Leakey, A. D. B., Buckler, E. S., Garcia, A. A. F., & Gore, M. A. (2020). Novel bayesian networks for genomic prediction of developmental traits in biomass sorghum. *G3: Genes, Genomes, Genetics*, *10*(2), 769–781. https://doi.org/10.1534/g3.119.400759

Erpelding, J. e. (2007). Inheritance of Anthracnose Resistance for the Sorghum Cultivar Redlan. *Plant Pathology*, *6*(2), 187–190.

Falconer, D. S. (1989). *Introduction to quantitative genetics*. Pearson Education India.

FAO (2018) Food and Agriculture Organization of the United Nations Database of agricultural production. FAO Statistical Databases. http://faostat.fao.org/site/339/default.aspx. Accessed 20 Oct 2023.

Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., & Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics*, *131*(3), 747–755. https://doi.org/10.1007/s00122-017-3033-y

Forbes, G. A. (1986). *Characterization of grain mold resistance in sorghum (Sorghum bicolor (L.) Moench) (histology, fluorescence microscopy)*. Texas A&M University.

Forbes, G. A., Bandyopadhyay, R., & Garcia, G. (1992). A Review of Sorghum Grain mold. In *A review of sorghum grain mold*.

Frederiksen, R. A. (1986). *Compendium of sorghum diseases*. American Phytopathological Society in cooperation with Dept. of Plant Pathology and Microbiology.

Gebrekidan, B. (1982). *Sorghum improvement in Eastern Africa: Proceedings of the Regional Workshop on Sorghum Improvement in Eastern Africa*. http://publication.eiar.gov.et:8080/xmlui/bitstream/handle/123456789/3182/014%20Sorghum%20Improvement%20in%20Eastern%20Africa.pdf?sequence=1

Geiringer, H. (1944). On the Probability Theory of Linkage in Mendelian Heredity. In *Source: The Annals of Mathematical Statistics*,15(1).

Habyarimana, E., & Lopez-Cruz, M. (2019). Genomic selection for antioxidant production in a panel of sorghum bicolor and s. bicolor × s. halepense lines. *Genes*, *10*(11). https://doi.org/10.3390/genes10110841

Habyarimana, E., Lopez-Cruz, M., & Baloch, F. S. (2020). Genomic selection for optimum index with dry biomass yield, dry mass fraction of fresh material, and plant height in biomass sorghum. *Genes*, *11*(1). https://doi.org/10.3390/genes11010061

Hamblin, M. T., Mitchell, S. E., White, G. M., Gallego, J., Kukatla, R., Wing, R. A., ... & Kresovich, S. (2004). Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of Sorghum bicolor. *Genetics*, *167*(1), 471-483.https://www.researchgate.net/publication/8540645

Harlan, J. R., & de Wet, J. M. J. (1972). A simplified classification of cultivated sorghum. *Crop Science*, *12*(2), 172–176.

Harris, H. B., B. J. Johnson, J., Dobson Jr., J. W., & Luttrell, E. S. (1964). Evaluation of anthracnose on grain sorghum. *Crop Science*, *4*(5), 460–462.

Heslot, N., Yang, H. P., Sorrells, M. E., & Jannink, J. L. (2012). Genomic selection in plant breeding: A comparison of models. *Crop Science*, *52*(1), 146–160. https://doi.org/10.2135/cropsci2011.06.0297

Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomes, Genetics*, *4*(6), 1027–1046. https://doi.org/10.1534/g3.114.010298

Hunt, C. H., van Eeuwijk, F. A., Mace, E. S., Hayes, B. J., & Jordan, D. R. (2018a). Development of genomic prediction in sorghum. *Crop Science*, *58*(2), 690–700. https://doi.org/10.2135/cropsci2017.08.0469

Hunt, C. H., van Eeuwijk, F. A., Mace, E. S., Hayes, B. J., & Jordan, D. R. (2018b). Development of genomic prediction in sorghum. *Crop Science*, *58*(2), 690–700. https://doi.org/10.2135/cropsci2017.08.0469

Ishimori, M., Hattori, T., Yamazaki, K., Takanashi, H., Fujimoto, M., Kajiya-Kanegae, H., Yoneda, J., Tokunaga, T., Fujiwara, T., Tsutsumi, N., & Iwata, H. (2020). Impacts of dominance effects on genomic prediction of sorghum hybrid performance. *Breeding Science*, *70*(5), 605–616. https://doi.org/10.1270/jsbbs.20042

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., Gore, M. A., Gupta, R., Ilut, D. C., Johnson, L., Lozano, R., Miller, Z., Ramu, P., Rathore, A., Romay, M. C., Upadhyaya, H. D., Varshney, R. K., Morris, G. P., Pressoir, G., … Ramstein, G. P. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *Plant Genome*, *13*(1). https://doi.org/10.1002/tpg2.20009

Kenga, R., Alabi, S. O., & Gupta, S. C. (2004). Combining ability studies in tropical sorghum (Sorghum bicolor (L.) Moench). *Field Crops Research*, *88*(2–3), 251–260. https://doi.org/10.1016/j.fcr.2004.01.002

Kukreja, S. L., Löfberg, J., & Brenner, M. J. (2006). A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC Proceedings Volumes*, *39*(1), 814–819. https://doi.org/10.3182/20060329-3-au-2901.00128

Lewontin, R. C., & Kojima, K. I. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution*, 458-472.

Lindsay, J. (2010). Sorghum: An ancient, healthy and nutritious old-world cereal. *United Sorghum Checkoff Program*. https://digitalcommons.unl.edu/intsormilpubs/7

Little, C. R., & Magill, C. W. (2009). The grain mold pathogen, fusarium thapsinum, reduces caryopsis formation in sorghum bicolor. *Journal of Phytopathology*, *157*(7–8), 518–519. https://doi.org/10.1111/j.1439-0434.2008.01530.x

Little, C. R., Perumal, R., Tesso, T., Prom, L. K., Odvody, G. N., & Magill, C. W. (2012). *Sorghum Pathology and Biotechnology - A fungal disease perspective: Part I. grain mold, head smut, and ergot*.

Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, 12(2), 57–63. https://doi.org/10.1016/j.tplants.2006.12.001

Mann, J. A., Kimber, C. T., & Miller, F. R. (1983). The Origin and Early Cultivation of Sorghums in Africa. *Texas Farmer Collection*.

Marley, P. S., Thakur, R. P., & Ajayi, O. (2001). Variation among foliar isolates of Colletotrichum sublineolum of sorghum in Nigeria. *Field Crops Research*, *69*(2). https://doi.org/10.1016/S0378-4290(00)00128-3

Mehta, P. J., Wiltse, C. C., Rooney, W. L., Collins, S. D., Frederiksen, R. A., Hess, D. E., Chisi, M., & Tebeest, D. O. (2005). Classification and inheritance of genetic resistance to anthracnose in sorghum. *Field Crops Research*, *93*(1), 1–9. https://doi.org/10.1016/j.fcr.2004.09.001

Mengistu, G., Ayana, A., & Belete, K. (2010). Combining ability for yield and its components in Ethiopian sorghum (*Sorghum bicolor* (L.) Moench) landraces. *East African Journal of Sciences*, *4*(1), 34–40.

Menkir, A., Ejeta, G., Butler, L., & Melakeberhan, A. (1996). Physical and chemical kernel properties associated with resistance to grain mold in sorghum. *Cereal Chemistry*, *73*(5), 613-617.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, *157*(4), 1819-1829. https://academic.oup.com/genetics/article/157/4/1819/6048353

Muleta, K. T., Pressoir, G., & Morris, G. P. (2019). Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. *G3: Genes, Genomes, Genetics*, *9*(2), 391–401. https://doi.org/10.1534/g3.118.200932

Münch, S., Lingner, U., Floss, D. S., Ludwig, N., Sauer, N., & Deising, H. B. (2008). The hemibiotrophic lifestyle of Colletotrichum species. *Journal of Plant Physiology*, *165*(1), 41–51. https://doi.org/10.1016/j.jplph.2007.06.008

Nordborg, M., & Tavaré, S. (2002). Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, 18(2). http://tig.trends.com

Pande, S., Thakur, R. P., Karunakar, R. I., Bandyopadhyay, R., & Reddy, B. V. (1994). Development of screening methods and identification of stable resistance to anthracnose in sorghum. *Field Crops Research*, *38*(3), 157-166.

Pfliegler, W. P., Pócsi, I., Győri, Z., & Pusztahelyi, T. (2020). The Aspergilli and their mycotoxins: Metabolic interactions with plants and the soil biota. *Frontiers in microbiology*, *10*, 2921. Frontiers Media S.A. https://doi.org/10.3389/fmicb.2019.02921

Prom, L. K., Perumal, R., Erpelding, J., Isakeit, T., Montes-Garcia, N., & Magill, C. W. (2009). A Pictorial Technique for Mass Screening of Sorghum Germplasm for Anthracnose (Colletotrichum sublineolum) Resistance. *The Open Agriculture Journal*, *3*(1), 20–25. https://doi.org/10.2174/1874331500903010020

Prom, L. K., Radwan, G., Perumal, R., Cuevas, H., O. Katile, S., Isakeit, T., & Magill, C. (2016). Grain Biodeterioration of Sorghum Converted Lines Inoculated with a Mixture of Fusarium thapsinum and Curvularia lunata. *Plant Pathology Journal*, *16*(1), 19–24. https://doi.org/10.3923/ppj.2017.19.24

Puglisi, D., Delbono, S., Visioni, A., Ozkan, H., Kara, I., Casas, A. M., ... & Fricano, A. (2021). Genomic prediction of grain yield in a barley MAGIC population modeling genotype per environment interaction. *Frontiers in plant science*, *12*, 664148. https://doi.org/10.3389/fpls.2021.664148

Ragot, M., Dell'Orco, A., Malgarini, L., Thevenin, P., Vernoy, J., Vivant, J., Zimmermann, R., & Gay, G. (1994). Marker-assisted backcrossing: a practical example. *Techniques et Utilizations Des Marqueurs Moleculaires. Montpellier, France*, 45–56.

Thirumala Rao, V., Sanjana Reddy, P., & Reddy, B. V. S. (2016). Combining ability for grain mold resistance in sorghum (Sorghum bicolor (L.) Moench). *SABRAO Journal of Breeding and Genetics*, *48*(03), 240-246.

Reddy, B. V., Bandyopadhyay, R., Ramaiah, B., & Ortiz, R. (2000). Breeding grain mold resistant sorghum cultivars. In *Technical and institutional options for sorghum grain mold management: proceedings of an international consultation* (pp. 195-224). ICRISAT, Patancheru, India.

Reddy, B. V. S., Ramesh, S., & Reddy, P. S. (2006). Sorghum genetic resources, cytogenetics, and improvement. *Genetic Resources, Chromosome Engineering, and Crop Improvement*, *2*, 309–363. https://doi.org/10.1201/9780203489260

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., ... & Buckler IV, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the national academy of sciences*, *98*(20), 11479-11484. www.pnas.org.

Rice, B., & Lipka, A. E. (2019). Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum. *The Plant Genome*, *12*(1), 180052. https://doi.org/10.3835/plantgenome2018.07.0052

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L., & Melchinger, A. E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics*, *44*(2), 217–220. https://doi.org/10.1038/ng.1033

Rodriguez-Herrera, R., Rooney, W. L., Rosenow, D. T., & Frederiksen, R. A. (2000). Inheritance of grain mold resistance in grain sorghum without a pigmented testa. *Crop Science*, *40*(6), 1573-1578.

Rosenow, D. T., & Dahlberg, J. A. (2000). Collection, conversion, and utilization of sorghum. *Sorghum: origin, history, technology and production. Frederiksen ISBN 0-471-24237-3 John Wiley and Son, Inc*.

Rosewich, U. L., Pettway, R. E., McDonald, B. A., Duncan, R. R., & Frederiksen, R. A. (1998). Genetic structure and temporal dynamics of a *Colletotrichum graminicola* population in a sorghum disease nursery. *Phytopathology*, *88*(10), 1087-1093.

Sage, F., & Zhu, G. (2011). Exploiting the engine of C4 photosynthesis. *Journal of Experimental Botany*, *62*(9), 2989–3000. https://doi.org/10.1093/jxb/err179

Sapkota, S., Boyles, R., Cooper, E., Brenton, Z., Myers, M., & Kresovich, S. (2020). Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. *Crop Science*, *60*(1), 132–148. https://doi.org/10.1002/csc2.20060

Sharma, R., Rao, V. P., Upadhyaya, H. D., Reddy, V. G., & Thakur, R. P. (2010). Resistance to grain mold and downy mildew in a mini-core collection of sorghum germplasm. *Plant Disease*, *94*(4), 439–444. https://doi.org/10.1094/PDIS-94-4-0439

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., & Nepolean, T. (2017). Genomic selection for drought tolerance using genome-wide SNPs in Maize. *Frontiers in Plant Science*, *8, 550*. https://doi.org/10.3389/fpls.2017.00550

Singh, S. D., & Bandyopadhyay, R. (2000). Grain mold. In *Compendium of Sorghum diseases*

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J. L., & McCouch, S. R. (2015). Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS Genetics*, *11*(2), 1–25. https://doi.org/10.1371/journal.pgen.1004982

Sprague, G., & Tatum, L. (1942). General vs specific combining ability in single crosses of corn. *Journal of the American Society of Agronomy.*, *34*, 923–932.

Stemler, A. B. L., Harlan, J. R., & De Wet, J. M. J. (1977). The sorghums of Ethiopia. *Economic Botany*, *31*, 446-460.

TeBeest, D., Kirkpatrick, T., & Cartwright, R. (2004). Common and Important Diseases of Grain Sorghum. *Grain Sorghum Production Handbook*, 37–46. https://uaex.edu/publications/pdf/mp297/6_diseases.pdf

Teklay, A. (2015). Prevalence and intensity of economically important fungal diseases of sorghum in south Tigray, Ethiopia. *Journal of Plant Sciences (Science Publishing Group)*, *3*(2), 92. https://doi.org/10.11648/j.jps.20150302.18

Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F., & Gaut, B. S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). *Proceedings of the National Academy of Sciences, 98(16)*, *98*, 9161–9166. www.pnas.orgcgidoi10.1073pnas.151244298

Terefe, T. G. (2002). *Epidemiology of grain mold of sorghum in South Africa and Ethiopia* (Doctoral dissertation, University of the Free State).

Thakur, R. P., Reddy, B. V., & Mathur, K. (2007). *Screening techniques for sorghum diseases. Information: Bulletin No. 76*. International Crops Research Institute for the Semi-Arid Tropics.

Thakur, R., & Mathur, K. (2000). Sorghum anthracnose. In *Compendium of Sorghum Diseases.*

Thakur, R. P., Reddy, B. V. S., Indira, S., Rao, V. P., Navi, S. S., Yang, X. B., & Ramesh, S. (2006). Sorghum grain mold information Bulletin No. 72.

Tibshiranit, R. (1996). Regression Shrinkage and Selection via the Lasso. In *J. R. Statist. Soc. B*, 58(1).

Vaillancourt, L. J., & Hanau, R. M. (1992). Genetic and morphological comparisons of Glomerella (Colletotrichum) isolates from maize and from sorghum. *Experimental Mycology*, *16*(3), 219–229. https://doi.org/10.1016/0147-5975(92)90030-U

Valèrio, H. M., Rèsende, M. A., Weikert-Oliveira, R. C. B., & Casela, C. R. (2005). Virulence and molecular diversity in Colletotrichum graminicola from Brazil. *Mycopathologia*, *159*(3), 449–459. https://doi.org/10.1007/s11046-005-0373-y

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Velazco, J. G., Jordan, D. R., Hunt, C. H., Mace, E. S., & van Eeuwijk, F. A. (2020). Genomic prediction for broad and specific adaptation in sorghum accommodating differential variances of SNP effects. *Crop Science*, *60*(5), 2328–2342. https://doi.org/10.1002/csc2.20221

Velazco, J. G., Jordan, D. R., Mace, E. S., Hunt, C. H., Malosetti, M., & van Eeuwijk, F. A. (2019). Genomic Prediction of Grain Yield and Drought-Adaptation Capacity in Sorghum Is Enhanced by Multi-Trait Analysis. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00997

Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. In *Crop Journal* (Vol. 6, Issue 4, pp. 330–340). Crop Science Society of China/ Institute of Crop Sciences. https://doi.org/10.1016/j.cj.2018.03.001

Wharton, P. S., & Julian, A. M. (1996). A cytological study of compatible and incompatible interactions between Sorghum bicolor and Colletotrichum sublineolum. *New Phytologist*, *134*(1), 25-34.

Wharton, P. S., Julian, A. M., & O'Connell, R. J. (2001). Ultrastructure of the infection of Sorghum bicolor by Colletotrichum sublineolum. *Phytopathology*, *91*(2), 149–158. https://doi.org/10.1094/PHYTO.2001.91.2.149

Williams, R., & Rao, K. (1981). A review of sorghum grain molds. *International Journal of Pest Management*, *87*(2), 200–211.

Xu, Y., Ma, K., Zhao, Y., Wang, X., Zhou, K., Yu, G., Li, C., Li, P., Yang, Z., Xu, C., & Xu, S. (2021). Genomic selection: A breakthrough technology in rice breeding. *Crop Journal*, 9(3), 669–677. Institute of Crop Sciences. https://doi.org/10.1016/j.cj.2021.03.008

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Zhang, H., Yin, L., Wang, M., Yuan, X., & Liu, X. (2019). Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in genetics*, *10*, 189. https://doi.org/10.3389/fgene.2019.00189

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H. P., Longin, F. H., Ranc, N., & Reif, J. C. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics*, *124*(4), 769–776. https://doi.org/10.1007/s00122-011-1745-y

Zhong, S., Dekkers, J. C., Fernando, R. L., & Jannink, J. L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics*, *182*(1), 355-364. https://doi.org/10.1534/genetics.108.098277

Chapter 2 - Application of genomic selection for the prediction of foliar anthracnose and grain mold disease resistance in sorghum

Abstract

Crop genetic resources held in gene banks are sources of genetic diversity for numerous traits that are essential for crop improvement. The Ethiopian sorghum germplasm collection comprises over 9000 accessions collected from different parts of the country. These genetic resources have been underexploited in current breeding programs mainly due to limitations of both financial and intellectual predicament. However, a rapid decline in genotyping cost and improvement in bioinformatic resources are enabling the implementation of genomic approaches for efficient utilization of these resources. Here, we evaluate the potential of genomic prediction for facilitating the utilization of sorghum germplasm collection in crop improvement. A total of 946 sorghum accessions were sampled from the gene bank and evaluated for Sorghum leaf Anthracnose *(Colletotrichum sublineolum*) and grain mold resistance in Bako, Ethiopia during the 2015 and 2016 cropping seasons. The accessions were genotyped using genotype-by-sequencing (GBS) method that generated a total of 410,717 high quality SNP markers. The markers were thinned in TASSEL software maintaining distances of 1,000bp, 50,000bp, 75,000bp, and 100,000bp between markers to see the effects of SNP density on prediction accuracy. The analysis was conducted by varying the training population size from 10% to 90% at an interval of 10%. Principal component analysis was conducted to assess the pattern of genetic relatedness among the accessions and to optimize kinship between the training and validation population, thereby improving the accuracy of predicting breeding values of the selection population.

2.1 Introduction

Despite its wide adaptability and resilience, sorghum suffers from various stress factors either due to less attention paid to the management of the crop or because it is often grown in environments where these factors are rampant. Whenever encountered, these stresses cause yield losses and grain quality deterioration affecting productivity, market value and nutritional attributes of the crop. The major biotic stressors affecting sorghum are parasitic weed Striga, foliar and panicle diseases caused by an assortment of fungal pathogens, sporadic occurrence of armyworms, shoot fly and migratory birds. Sorghum foliar anthracnose and grain mold are the two major diseases of sorghum and more than 50% of sorghum yield losses that occur in warm and humid environments are attributable to these diseases. Anthracnose affects almost all aboveground parts including leaves, stems, and grains while grain mold affects the sorghum panicle. The yield loss caused by these diseases have been reported to range from 50% to 86% depending on severity (Ali and Warren, 1992; Thakur and Mathur, 2000). Sorghum is endowed with vast genetic variability which can be exploited in breeding programs to enhance adaptation to these stresses. The whole genome of sorghum has been sequenced (Paterson et al., 2009) and available for public use.

Knowledge of the extent of genetic variation, the pattern of inheritance of traits in working populations, and marker-trait associations are of paramount importance for charting the way for genetic improvement. The application of genomics becomes imperative for unlocking the potential of genetic diversity. The declining genotyping cost in recent years is permitting whole genome sequencing of novel germplasm for gene discovery and the application of genomic selection to aid breeding efforts (Jannink et al., 2010).

Genomic selection is a revolutionary tool that involves high-throughput genotyping and phenotyping to predict the genetic potential of individuals for complex traits. This approach

exploits the information embedded in the entire genome, enabling the selection of superior materials without the time-consuming and resource-intensive field phenotyping. By combining genomic data with precise phenotypic information, genomic selection facilitates the identification of genetic markers associated with disease resistance, enabling breeders to efficiently develop resistant sorghum varieties.

Genomic selection has been implemented to predict grain yield and its attributes such as number of grains per panicle, thousand-grain weight, number of productive tillers by utilizing different training population types and models. The type of prediction model, t heritability of the trait, the type of population and training population sizes were found to affect the prediction accuracy. A genomic prediction of 0.28 was reported for complex traits such as the distribution of weight to the individual grains in the panicle of rice (Yabe et al., 2018). On the other hand, a prediction of 0.78 was reported for grain yield in maize (Rio et al., 2019). A prediction accuracy of 0.6 was reported in populations developed using mating design II while only 0.1 was reported from random test cross populations (Fristche-Neto et al., 2018). The size of the training population and genetic relationships between the training and breeding populations were shown to influence the accuracy of genomic selection for grain yield (Lozada et al., 2019, 2020).

Marker-assisted selection (MAS) has been demonstrated to be effective in enhancing breeding efforts for qualitative resistance. However, its efficacy diminishes when applied to diseases with quantitative resistance that is influenced by numerous genes with minor effects. Hence, genomic selection steps in and fills the gap by serving as a vital tool for the improvement of resistance controlled by quantitative genes. The technique has been successfully implemented for disease resistance in wheat, rice, maize, and barley. Genomic prediction accuracies ranging from 0.14 to 0.85 were reported for wheat diseases including Fusarium head blight, rust, powdery

mildew Septoria tritici blotch, Stagonospora nodorum blotch and tan spot. GS has also been applied in rice to identify blast-tolerant lines; and it has been used to select for tolerance to Stenocarpella maydis and Striga in maize. For barley, GS has been used to predict Fusarium head blight severity achieving accuracy of up to 0.72 (Lorenz et al., 2012; Daetwyler et al., 2014; Mirdita et al., 2015; dos Santos et al., 2016; Sallam and Smith, 2016; Juliana et al., 2017; Sarinelli et al., 2019; Badu-Apraku et al., 2019; Huang et al., 2019). However, due to the change in environmental conditions, pathogens are continuously evolving and as a result, new races and biotypes are being reported. Thus, identifying and incorporating resistance genes into germplasms is a continuous process and has to mirror the evolution of new pathogen strains (Váry et al., 2015; Fones et al., 2020).

On the other hand, only little information is available on the prospect of genomic selection to enhance disease resistance in sorghum. Thus, the current work was aimed at determining the potential of genomic prediction for genetic improvement of resistance to sorghum foliar anthracnose and grain mold diseases.

2.2 Materials and methods

2.2.1 Germplasm and Phenotyping

A set of 946 sorghum accessions were utilized in this experiment. These are subset of the 2000 accessions sampled from the over 9000 national sorghum collections maintained by the Ethiopian Biodiversity Institute (EBI). The Ethiopian sorghum collections serve as the genetic resource for local variety development efforts such as by the Ethiopian Institute of Agricultural Research and regional agricultural research institutes. Representing about 17% of the USDA sorghum germplasm collection and of the international gene banks, Ethiopian sorghums have been the primary sources of novel genes in global sorghum improvement. The accessions were collected from diverse agroecological zones of Ethiopia at various elevations ranging from below 500 m to over 3200 meters above sea level as well as from different parts of the country with a range of soil type, moisture condition, and accessions of unique utilization attributes (Figure 2.1 and Figure 2.2).

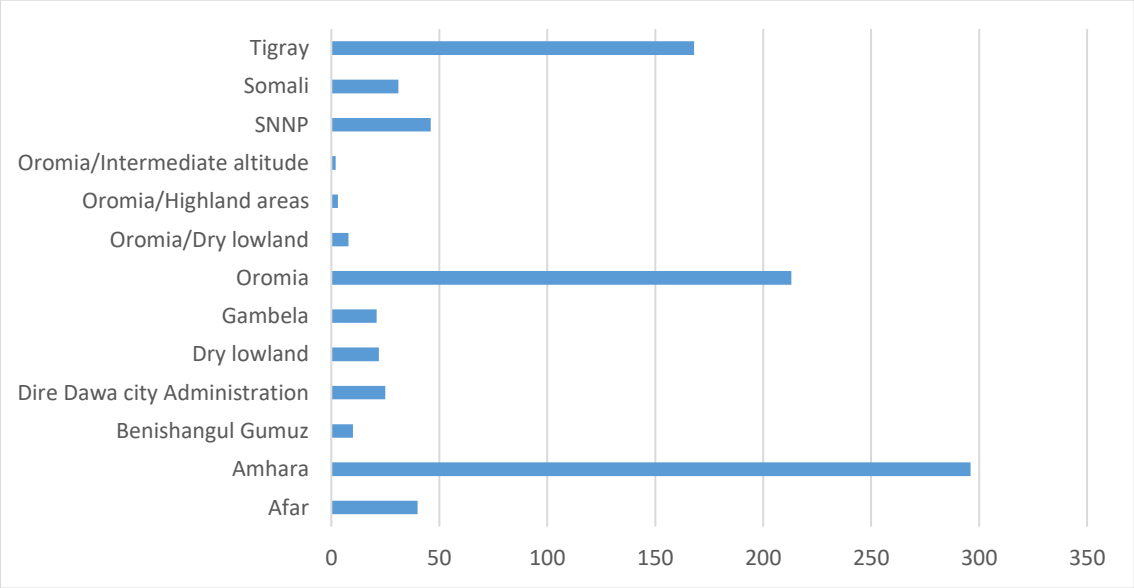


Figure 2.1. Number of accessions by collection area/region.

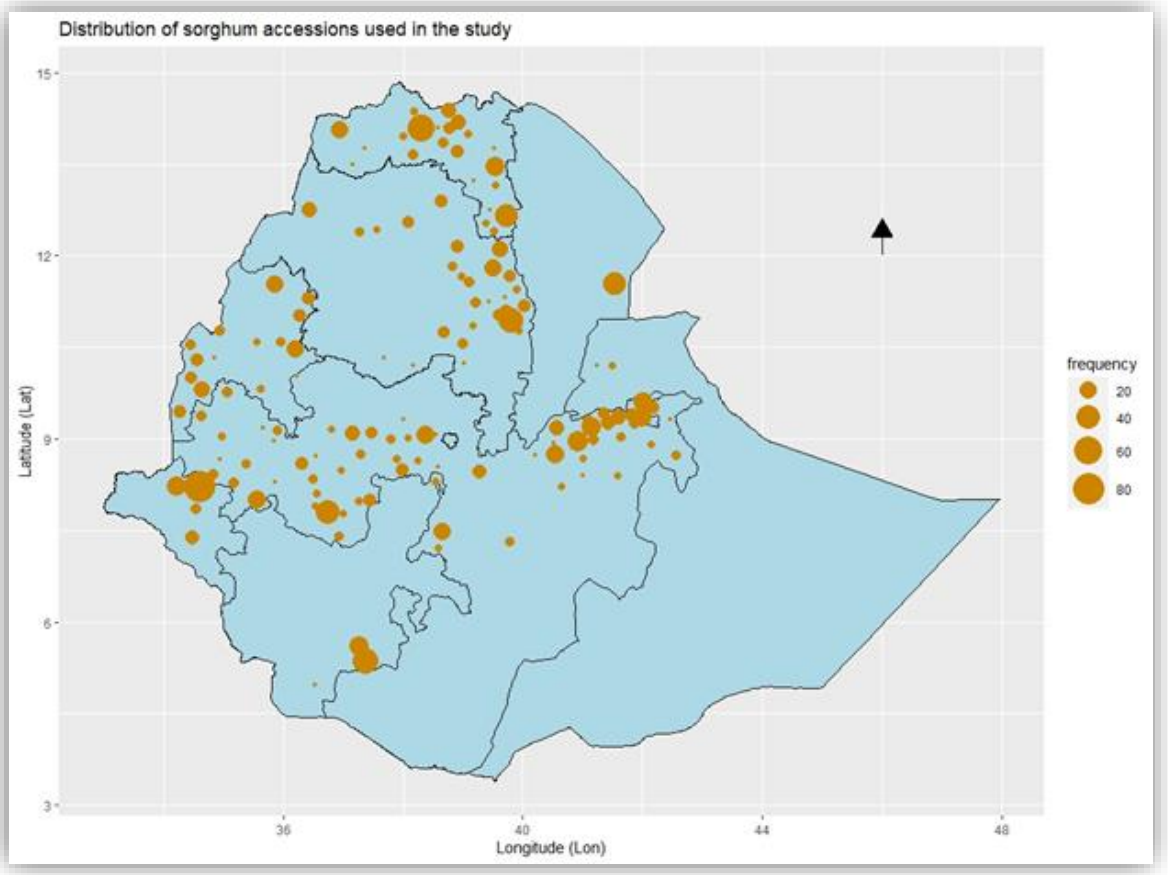


Figure 2.2. Distribution and magnitude of accessions collected from different parts of Ethiopia (dot size represents the relative number of accessions collected).

The accessions were planted at Bako (9°08'N,37°03'E), western Ethiopia in 2015 and 2016 seasons. The station is located at an elevation of 1642 m a.s.l. and has a mean annual temperature of 19.7°C and a mean annual rainfall of 1281 mm. The accessions were grown on 5-meters long single row plots spaced 0.75 meters apart. The spacing between plants was 20 cm. Fertilizer phosphorus was applied in the form of Diammonium phosphate (DAP) at the rate of 100 kg/ha DAP at planting. Nitrogen was applied in the form of Urea at the rate of 50 kg/ha when the plants reached knee height. Other agronomic management practices were employed based on the recommendations for the area. Five plants per row (accession) were selected and tagged for data collection on disease response. Disease reactions were scored using a 1-5 scale for both leaf

anthracnose and grain mold diseases based on descriptions provided by (Thakur et al., (2007 and Prom et al. (, 2009).

For anthracnose, a score of '1' was assigned to accessions with leaves showing no sign of infection, score '2' was assigned to accessions showing hypersensitive reactions with local lesions but no visible acervuli. Accessions with infected bottom leaves with acervuli were assigned score of '3'; and a score of '4' was assigned to accessions where middle to bottom leaves were infected and acervuli were visible, while score '5' was given to accessions where infections reached up to the flag leaf and acervuli were observed (Thakur et al., 2007; Prom et al., 2009).

For grain mold, a score of '1' was assigned for highly resistant accessions showing less than 1% grain mold on a panicle. A rating of '2' was assigned where the percentage of grains molded on a panicle was between 1 and 10%. A score of '3' was assigned to those showing 11% and 25% (moderate resistance), and '4' to those with grain mod incidence of 26-50%. Accessions with >50% molded grains on a panicle were assigned a score of '5' (Thakur et al., 2007).

2.2.2 DNA isolation, library preparation and sequencing

DNA was extracted from freeze-dried tissue collected from week-old seedlings by use of the acetyl trimethylammonium bromide (CTAB) DNA extraction protocol modified for 96 well plates (Mace et al., 2003). Ten libraries were constructed for the 946 accessions, using the restriction enzyme *Ape*KI (recognition cite: 5'...G|CWGC...3') for digestion. The libraries were sent to the University of Wisconsin, Biotechnology Center DNA Sequencing Facility for genotyping. The accessions were genotyped by the genotyping-by-sequencing (GBS) platform (Elshire et al., 2011) on an Illumina HiSeq 2500 lane sequencing machine.

2.2.3 SNP calling, data filtering and missing data imputation.

After obtaining the genotype data, sequence reads were aligned to the most recent BTx623 sorghum reference genome version 3 and SNP calling was conducted using TASSEL-GBS 5.0 (Glaubitz et al., 2014). The genotypes generated were filtered for minor allele frequency (MAFs) of (> 0.01) and percent of missing SNPs (< 20%) resulting in 410,727 SNPs for analysis. Afterward, the remaining missing data were imputed using BEAGLE 4.1 software package (Browning & Browning, 2016). Imputation involves inferring the ungenotyped variants of the target samples based off the available data from the reference genome (Marchini and Howie, 2010). Further thinning of the SNPs was conducted in TASSEL-GBS 5.0 by maintaining minimum distances of 1000 bp, 50,000 bp and 100,000 bp to see its effect on prediction accuracy. Accordingly, 90,074 SNPs, 10,458 SNPs and 5,731 SNPs, respectively, were obtained after thinning. Analysis was conducted based on these three SNP categories.

2.2.4 Genomic-estimated breeding values

The Genomic estimated breeding values (GEBVs) for individuals in the test/validation sets were calculated based on the marker effects estimated from the training population. GEBVs were computed using ridge regression best linear unbiased perdition (rrBLUP) package (Endelman, 2011) in the R programming language (Newell and Jannink, 2014). The values were estimated based on cross-validation tests using two sets of populations. The first set consisted of the training set while the other set was the testing set. For the training set, samples of 10% (95 accessions), 20% (189 accessions), 30% (284 accessions), 40% (378 accessions), 50% (473 accessions), 60% (567 accessions), 70% (662 accessions), and 80% (757 accessions) were considered and the remaining proportions were used as the testing sets in each case. Tests were run for 500 iterations and the means were used for analysis.

Prediction accuracy estimate

The prediction accuracy was calculated using the following formula:

$$\frac{\mathrm{r(GEBV:PEBV)}}{\sqrt{h^2}},$$

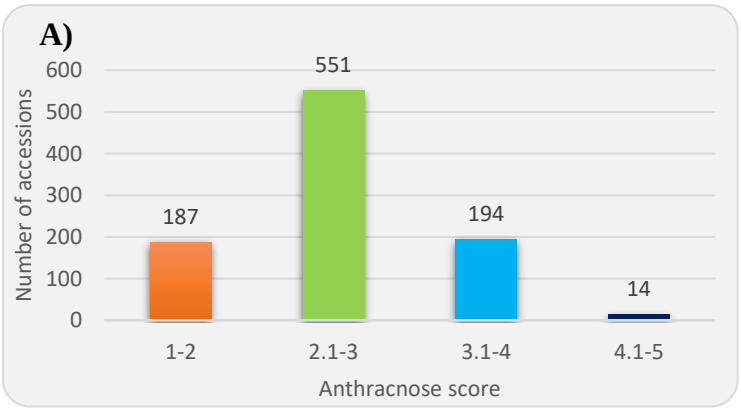
where ***r*** is the Pearson's correlation between the **GEBVs** and the phenotypic estimated breeding values (**PEBVs**) in the test population, $\boldsymbol{h^2}$ is the broad-sense heritability on a line-mean basis.

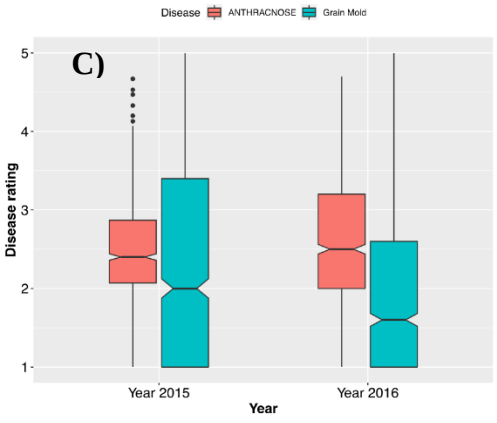
2.3 Results

The analysis of disease data revealed remarkable variation among the accessions both for anthracnose and grain mold resistance. Genomic prediction of disease reaction for the accessions demonstrated prospect for application of genomic selection in genetic improvement of sorghum.

2.3.1 Variation for anthracnose resistance

There was remarkable variation between the accessions for resistance to foliar anthracnose. Taking into account the average of the two years, 187 accessions were found to be resistant or highly resistant (mean disease score of 2 or lower), while 208 were rated as 3 or above and were susceptible to foliar anthracnose (Figure 2.3). The bulk of the entries (551) were shown to have moderate reaction (score 2.1-3) to the disease. Looking at the two years separately, 202 accessions were rated 2 or below, 570 were rated to be between 2 and 3 while 214 were rated above 3 for the year 2015. On the other hand, 264 accessions were rated 2 or below, 417 were rated between 2 and three while 265 were rated above 3 for the year 2016 (Figure 2.3B).





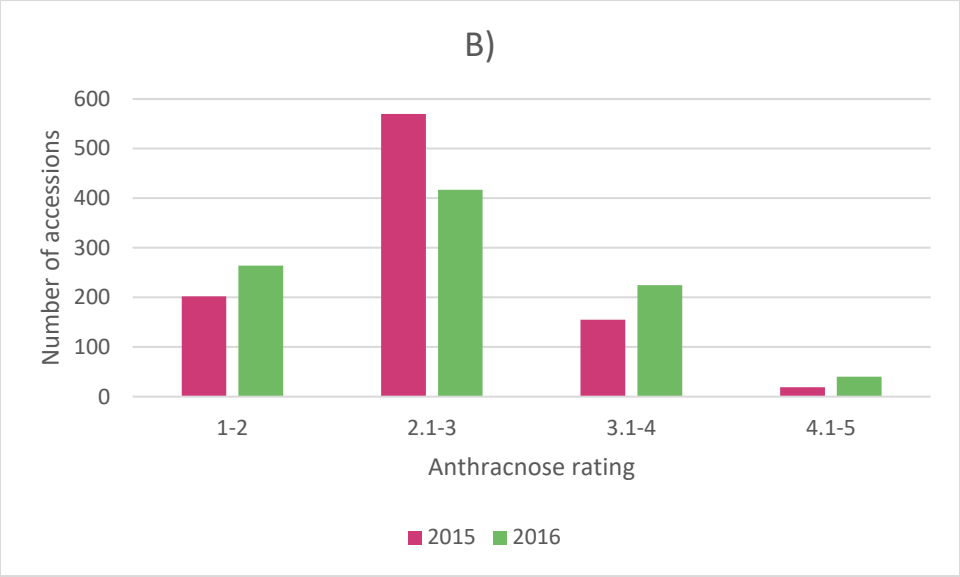


Figure 2.3. Frequency distribution of anthracnose resistance rating (mean of two years) (A), and the two years separately (B). Reactions of sorghum accessions to anthracnose disease and grain mold (C).
Five plants were scored for each accession and the means were used. The scoring scale was based on 1-5 with 1, plants with no sign of infection; 2, plants with hypersensitive reactions, local lesions without acervuli; 3, plants showing infected bottom leaves with acervuli; 4, plants with middle to bottom leaves infected and visible acervuli; and 5, plants with infections reaching flag leaf and acervuli visible).

2.3.2 Variation in grain mold resistance

Unlike the anthracnose, the bulk of the accessions showed resistance reaction to infection by grain mold pathogens. Based on the two-year mean, four hundred forty-three accessions (47%) were shown to be highly resistant with a score of < 2. Further, another 298 accessions fall in resistant category with mean disease score of 2.0 and 2.9, while 186 accessions only moderately resistant/susceptible with a score of 3-3.9. Only 19 accessions had disease score of 4 and above (Figure 2.4A). For 2015, 697 accessions were rated 3 or below and the remaining were rated above 3. On the other hand, 836 accessions were rated below 3 and 110 were rated above 3 for the year 2016 (Figure 2.4B).

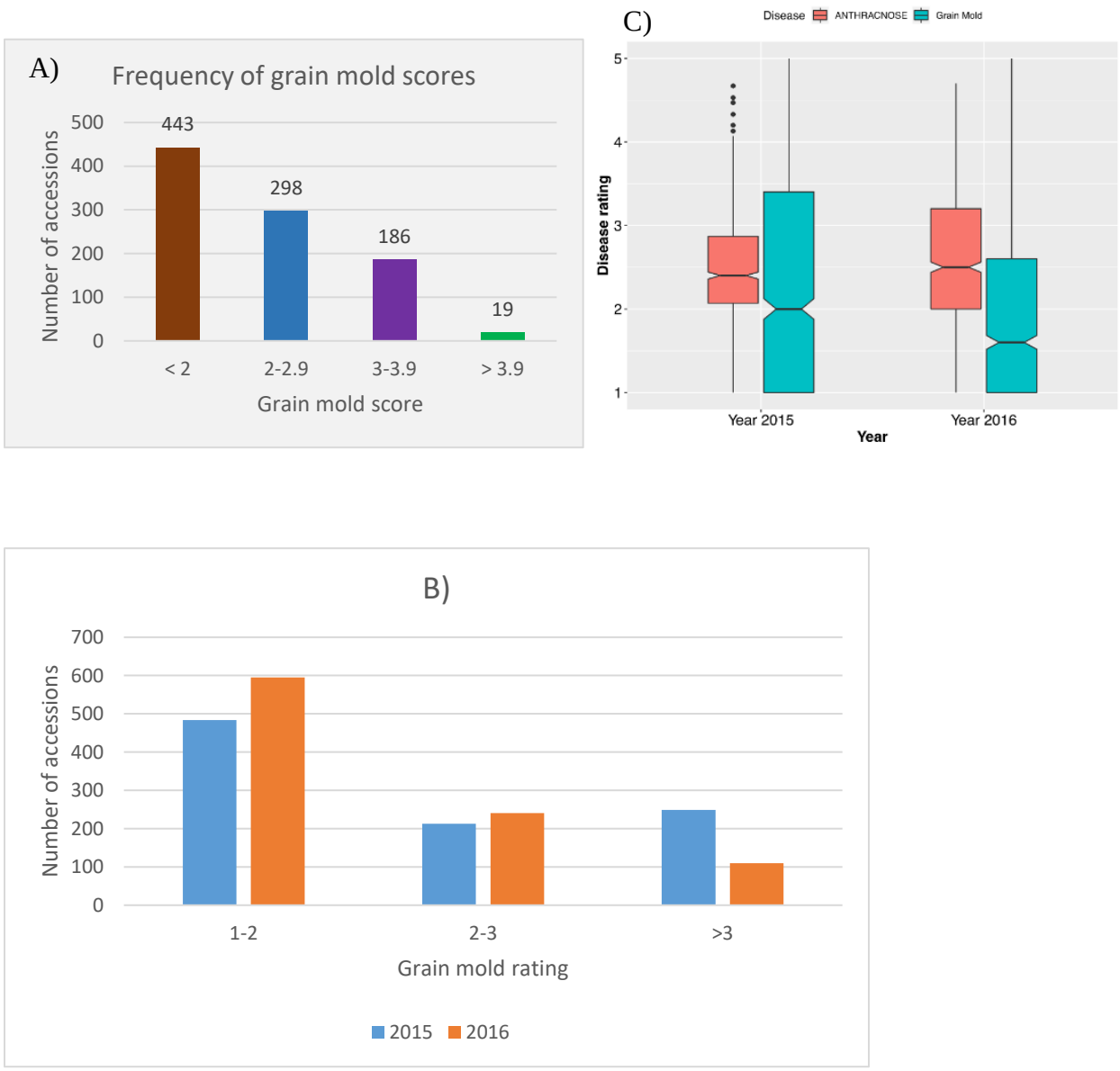


Figure 2.4. Frequency distribution for grain mold resistance for mean of the two years(A) and for 2015 and 2016 (B). Reactions of sorghum accessions to anthracnose disease and grain mold (C). Five plants were scored for each accession and the means were used. The scoring scale was based on 1-5 (1 – highly resistant with less than 1% grain mold on a panicle. 2 – only 1 to 10% of grains on a panicle molded, 3 – moderate resistance (11% and 25%), 4 – susceptible (26%-50% of the grains on a panicle molded), 5 – highly susceptible (>50% grains on a panicle had molds).

2.3.3 Marker distribution and genomic prediction accuracy

The 410,717 SNP markers were uniformly distributed across all the ten sorghum chromosomes. Marker density for the 1 Mb region ranged from 0 to over 1,665 showing a general pattern of denser SNPs towards the telomer and less dense towards the centromere (Figure 2.5).

Prediction accuracy estimates for both anthracnose and grain mold reactions were conducted by considering combinations of different population sizes for training set and different SNP marker densities. These were considered to observe the effect that different population sizes have on the prediction accuracy and how the number of SNPs affect prediction accuracy.

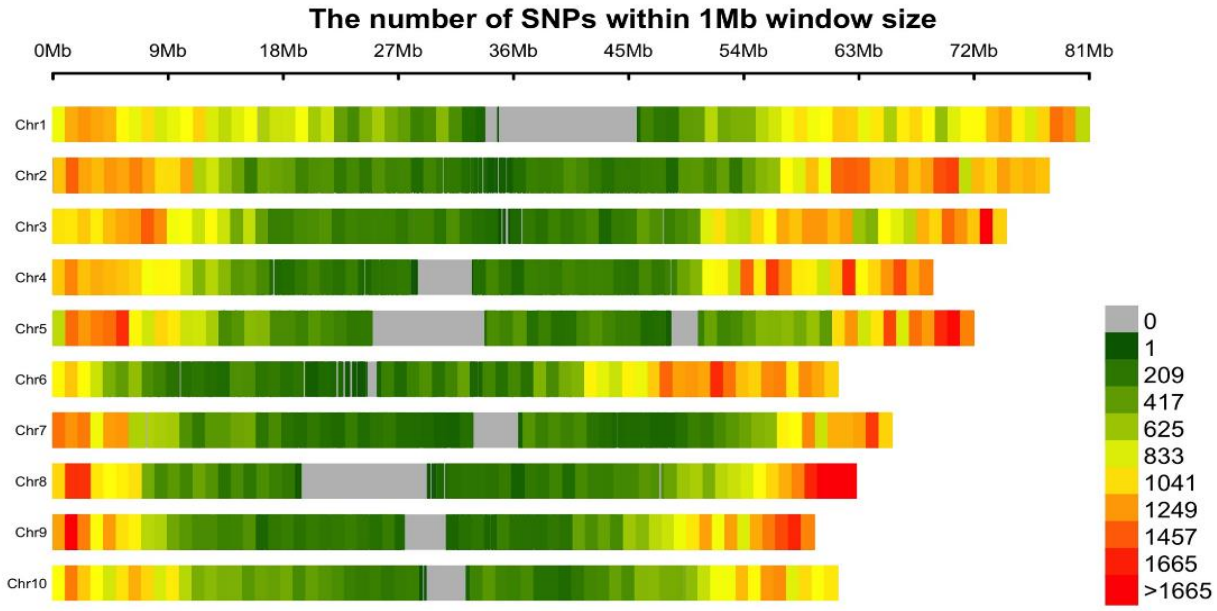


Figure 2.5. Genome-wide density of 410,717 single-nucleotide polymorphism (SNP) markers across the ten chromosomes of sorghum.

2.3.3.1 Effects of training population size on GEBV accuracy for anthracnose and grain mold disease resistance

The training population size was shown to have marked effect of prediction accuracy of foliar anthracnose reaction. For the 10% (95 accessions) training set, the mean prediction accuracy for the 2016 data was about 0.39 whereas it was 0.44 for the highest training set (90%, 851 accessions). For the 2015 data, prediction accuracy for 10% training population size was 0.33 and was increased to 0.39 when the training set was increased to 90%. The predication accuracy for different iterations for each training set was different with the range being highest for 90% training set (Figure 2.6).

Increasing the training population size from 10% to 60% raised the prediction accuracy by 9% in 2016 and 13% in 2015 data. For the 2016 data, prediction accuracy was increased by 18% when training population size was increased from 10% to 90% and only by 7% when it was increased from 30% to 90%.

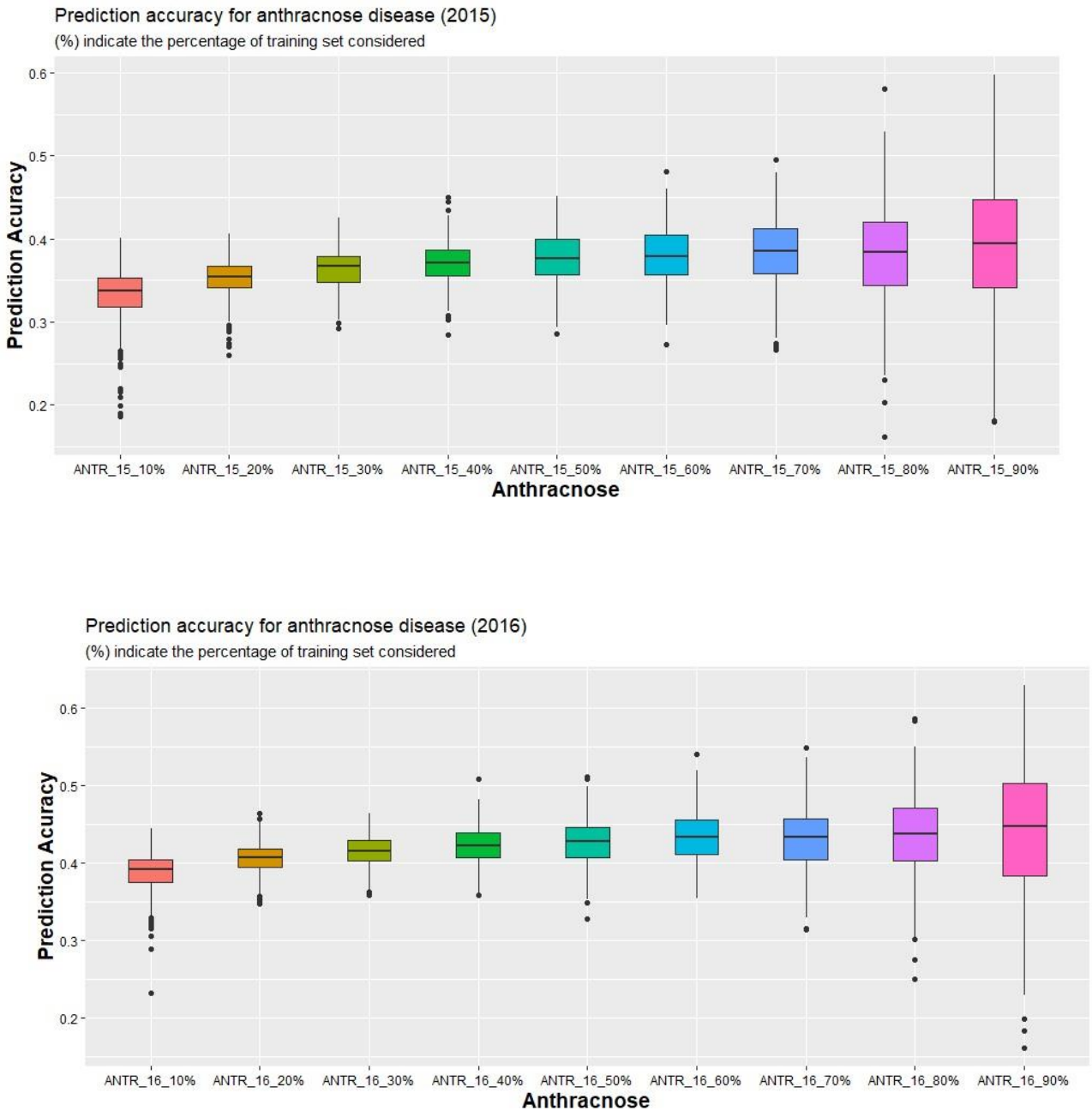


Figure 2.6. Prediction accuracy for anthracnose resistance for the year 2015 and 2016.

Similarly, marker density had moderate effect on prediction accuracy of foliar anthracnose disease reaction. The highest and lowest prediction accuracies of 0.44 and 0.33 were obtained for

the highest marker density (90,074 SNPS) and the lowest marker density (5,731 SNPs), respectively. Prediction accuracy slightly increased for increase in marker densities (Figure 2.7), and this was true for both 2015 and 2016 data at all training population sizes. Marker densities 5731 and 10458 had similar prediction accuracy at all training population sizes for both 2015 and 2016 data. The highest marker density, however, gave better accuracy across different training population sizes.

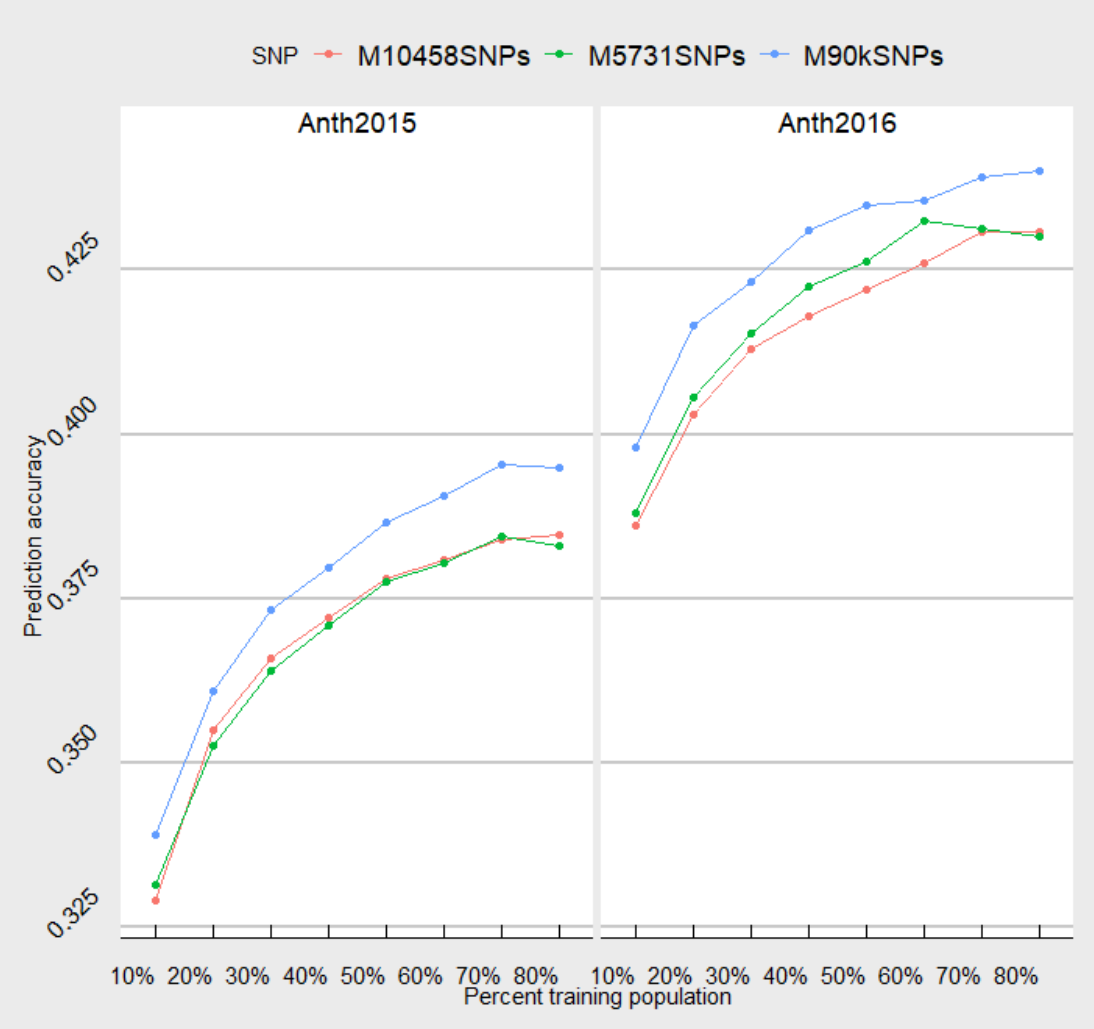


Figure 2.7. Prediction accuracy for anthracnose resistance as affected by number of SNPs for the years 2015 and 2016.

M10458SNPs = 10458 SNPs, M5734SNPs = 5731 SNPs, and M90kSNPs = 90074 SNPs

2.3.3.2 The effect of training population size and marker density on GEBV accuracy for grain mold resistance.

Genomic estimated breeding values for grain mold resistance were conducted using different training population size to observe its effect on prediction accuracy. The study also considered different marker densities to see its effect on perdition accuracy.

These populations were considered to determine the effect of the training population sizes on prediction accuracy. Like for the foliar anthracnose, the training population size was shown to have impact on prediction accuracy of grain mold reaction. Accordingly, the mean prediction accuracy of 0.38, 0.39, 0.40, 0.40, 0.40, 0.40, 0.40, 0.41, and 0.41 were obtained for training population sizes of 95 (10%), 189 (20%), 284 (30%), 378 (40%), 473 (50%), 567 (60%), 662 (70%), 757 (80%), and 851 (90%), respectively, at the lowest marker density (5731). The prediction accuracy increased as training population size increased for all marker densities. The prediction accuracies for the highest marker density (90,074 markers) ranged from 0.16 to 0.64 for the 200 iterations. With 10,458 markers and the same number of iterations, a prediction accuracy ranged from 0.16 to 0.59 the lowest being when 95 (10%) accessions were included in training set and the highest was for 757 (90%) accession in the training set (Figure 2.8 and Figure 2.9).

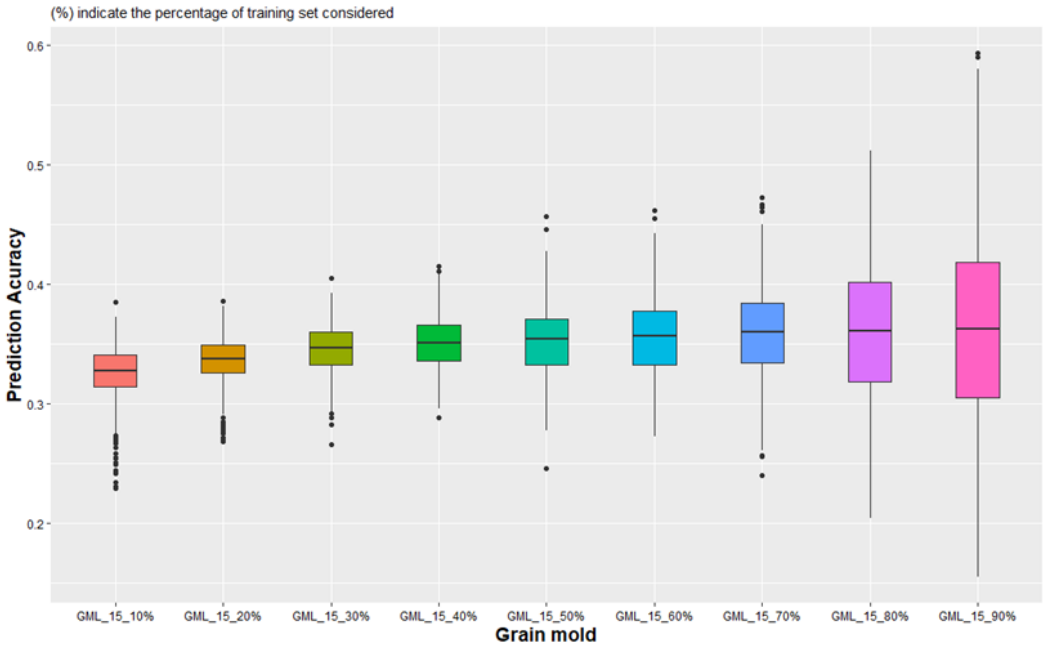


Figure 2.8. Prediction accuracy for grain mold resistance for the year 2015

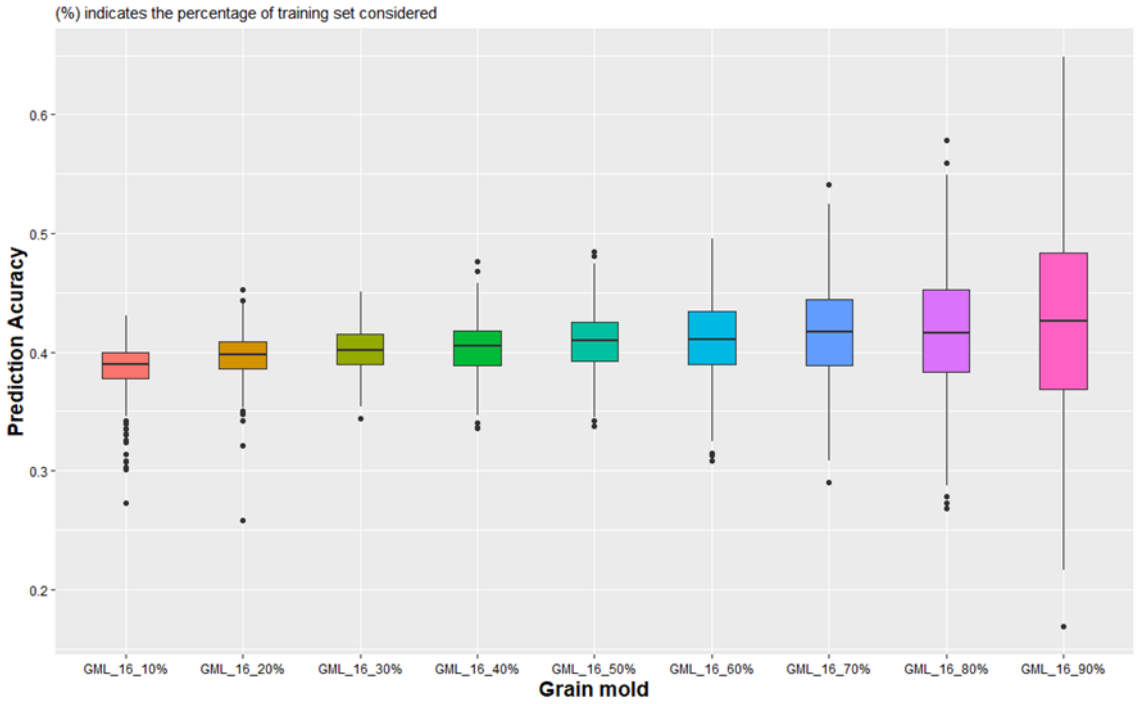


Figure 2.9. Prediction accuracy for grain mold for the year 2016.

The effects of marker densities on prediction accuracy of grain mold reaction were not as evident as the anthracnose for both 2015 and 2016 data. This was different for training population sizes with the effect of marker density being more evident when training population of about 30% was considered (Figure 2.10).

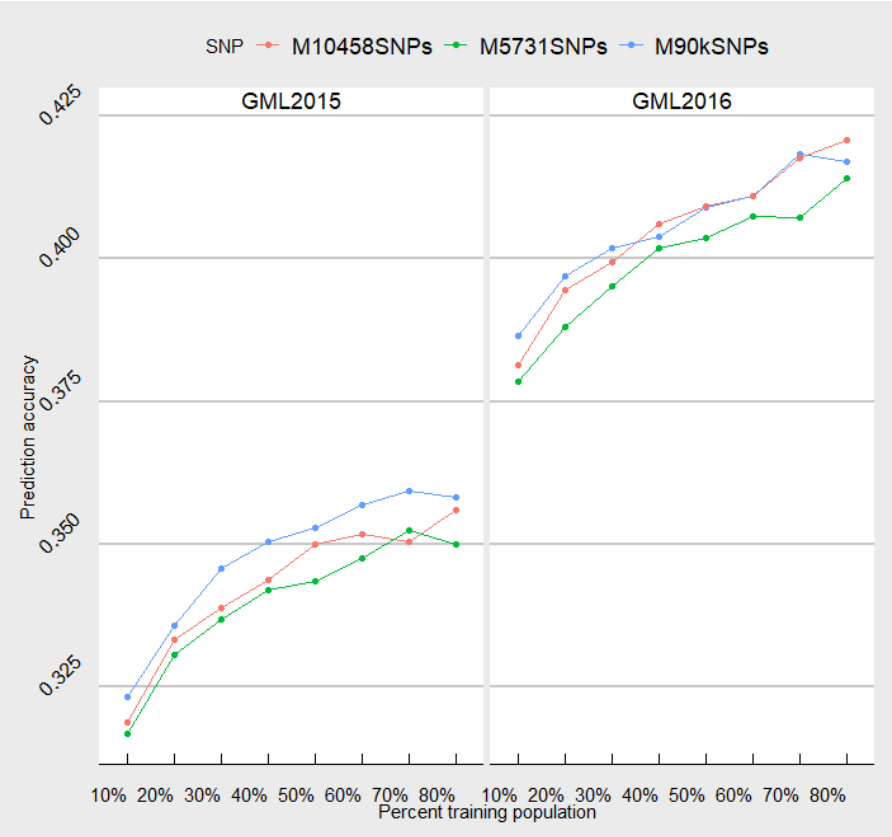


Figure 2.10. Prediction accuracy with different marker densities for grain mold resistance.

2.4 Discussion

Sorghum is endowed with diverse genetic resources that are of enormous value for improvement of the crop. However, they are not adequately exploited in breeding programs either due to limited access to the resources or due to lack of funding to dissect the traits harbored in them. The advent of marker technology has added new breeding tools to support the conventional breeding effort. Marker assisted breeding has enabled breeders to rapidly incorporate traits of economic value to breeding lines. But the application of MAS was limited to quantitative trait locus (QTLs) with large effect and hence has limited use for genes with smaller effects. Thus, genomic selection, based on next generation sequencing, has allowed exploitation of markers which were otherwise difficult to capture with MAS techniques. Breakthroughs in genomic research, advances in bioinformatic tools and a rapid decline in genotyping cost in recent years has created an enabling environment for the use of genomic approaches to exploit minor effect alleles to meet the ever-increasing human needs.

Genomic selection covers the whole genome and accounts for all loci with major and minor effects. GS exploits the population by dividing them into two sets, the training and testing set. The training set consists of individuals that have both phenotypic and genotypic data while the testing set only has genotypic data. Genomic selection thus utilizes phenotypic and genotypic data from the training set to predict the performance of the entire population based on their genotype only and thus eliminates the need for phenotyping the whole population.

In this study, the potential of genomic selection to enhance the utilization of sorghum germplasm in crop improvement was assessed. The effect of different training population sizes and different marker densities on prediction accuracy in genomic selection were evaluated. A total

of 946 sorghum accessions sampled from the Ethiopian Biodiversity Institute were evaluated for resistance to leaf anthracnose and grain mold diseases.

The results showed that the training population sizes, and marker densities have marked effect on prediction accuracy substantiating previous reports (Howard et al., 2014; Wang et al., 2018). The overall prediction accuracies for both leaf anthracnose and grain mold were fairly low, only about 0.4 at higher training population sizes. This may be attributable to various factors including the nature of the population. The accessions used in this study come from diverse geographic regions where their structure is shaped by combination of geophysical conditions of the regions and the unique social and cultural backgrounds of the native population in these regions that may have resulted in the emergence of unique set of germplasm, especially in isolated regions and communities. Both grain mold and anthracnose are complex diseases caused by multitudes of pathogens that thrive under different environments, hence there is high chance that certain genes and pathogen strains may be unique to certain environments and thus resistance alleles to those pathogens may be found only in that or similar environments.

We expect the prediction accuracies to be higher if the population were gathered from geographies where traditional seed exchange was common or if they were derived from same founding parents and then subjected to random mating. In such situations training populations set of any size might better represent the testing population and provide higher prediction accuracy compared to population of accessions collected from distant environments that had little or no history of genetic recombination. Nevertheless, though the prediction accuracy of this level may be considered low and do not allow to effectively hand pick the best accessions, it may allow to sieve through and remove a chunk of accessions of low prospect from future testing activities thus significantly reducing cost and time.

Considering 90,074 markers, the average prediction accuracies of leaf anthracnose incidence for the training population sizes of 95, 189, 284, 378, 473, 567, 662, and 757 were 0.40, 0.42, 0.42, 0.43, 0.43, 044, 0.44 and 0.44, respectively. With 10,458 markers, the prediction accuracies for the same population sizes were 0.39, 040, 0.41, 0.42, 0.42, 0.43, 0.43, and 0.43, respectively, and was 0.39, 0.41, 0.42, 0.42, 0.43, 0.43 and 0.43 when 5,731 markers were used for the same training population sizes. The study shows that marker density for the specified range did not significantly alter prediction accuracy. It is also important to note that increasing the training population size beyond 60% (567 accessions) did not increase the prediction accuracy indicating that the prediction accuracy reached a plateau at 60%. In a similar study conducted on canola, the prediction accuracy reached plateau when 70% training population was considered (Jan et al., 2016).

The accuracy of genomic estimated breeding value for grain mold resistance had the same trend as that of the anthracnose. For training populations sizes 95, 189, 284, 378, 473, 567, 662, and 757, and 90,074 markers, the prediction accuracies were 0.39, 0.40, 0.40, 0.40, 0.41 0.41, 0.42, and 0.42. Whereas for the 5,731 markers, prediction accuracies were 0.38, 0.39, 0.40, 0.40, 0.40, 0.41, 0.41, and 0.41 for the respective training population sizes. This revealed that training population sizes 189, 284, and 378 had no significate difference for prediction accuracy. Thus, the result demonstrates that, for populations of this nature, increasing marker numbers and training population sizes beyond 30% doesn't necessarily increase prediction accuracy and thus lower number of markers and modest size training set can be used for genomic prediction which saves cost and time for testing and data analysis.

2.5 Conclusion

The result indicated that genomic prediction has potential to estimate the performance of sorghum germplasm resources based on their genetic information per se without going through the rigors of field evaluation for all the available germplasm. This is true even for distant accessions with little or no opportunity for genetic recombination. This makes genomic prediction a powerful tool not only in breeding programs but also for germplasm characterization by the gene banks for cataloguing them for specific uses. This new tool will be of significant practical application for cryptic traits such as nutritional quality that are difficult to track in the field as well as for polygenic traits such as grain yield that are getting expensive to generate enough data from conventional field testing.

References

Badu-Apraku, B., Talabi, A. O., Fakorede, M. A. B., Fasanmade, Y., Gedil, M., Magorokosho, C., & Asiedu, R. (2019). Yield gains and associated changes in an early yellow bi-parental maize population following genomic selection for Striga resistance and drought tolerance. *BMC Plant Biology*, *19*(1). https://doi.org/10.1186/s12870-019-1740-z

Daetwyler, H. D., Bansal, U. K., Bariana, H. S., Hayden, M. J., & Hayes, B. J. (2014). Genomic prediction for rust resistance in diverse wheat landraces. *Theoretical and Applied Genetics*, *127*(8), 1795–1803. https://doi.org/10.1007/s00122-014-2341-8

De Roos, A. P. W., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, *183*(4), 1545–1553. https://doi.org/10.1534/genetics.109.104935

dos Santos, J. P. R., Pires, L. P. M., de Castro Vasconcellos, R. C., Pereira, G. S., Von Pinho, R. G., & Balestre, M. (2016). Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics*, *17*(1). https://doi.org/10.1186/s12863-016-0392-3

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019379

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Fristche-Neto, R., Akdemir, D., & Jannink, J. L. (2018). Accuracy of genomic selection to predict maize single crosses obtained through different mating designs. *Theoretical and Applied Genetics*, *131*(5), 1153–1162. https://doi.org/10.1007/s00122-018-3068-8

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high-capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090346

Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomes, Genetics*, *4*(6), 1027–1046. https://doi.org/10.1534/g3.114.010298

Huang, M., Balimponya, E. G., Mgonja, E. M., McHale, L. K., Luzi-Kihupi, A., Wang, G. L., & Sneller, C. H. (2019). Use of genomic selection in breeding rice (Oryza sativa L.) for resistance to rice

blast (Magnaporthe oryzae). *Molecular Breeding*, *39*(8). https://doi.org/10.1007/s11032-019-1023-2

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., & Snowdon, R. J. (2016). Genomic prediction of testcross performance in canola (Brassica napus). *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0147769

Jannink, J. L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics and Proteomics*, *9*(2), 166–177. https://doi.org/10.1093/bfgp/elq001

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Rutkoski, J. E., Poland, J. A., Bergstrom, G. C., & Sorrells, M. E. (2017). Comparison of Models and Whole-Genome Profiling Approaches for Genomic-Enabled Prediction of Septoria Tritici Blotch, Stagonospora Nodorum Blotch, and Tan Spot Resistance in Wheat. *The Plant Genome*, *10*(2). https://doi.org/10.3835/plantgenome2016.08.0082

Lorenz, A. J., Smith, K. P., & Jannink, J. L. (2012). Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science*, *52*(4), 1609–1621. https://doi.org/10.2135/cropsci2011.09.0503

Lozada, D. N., Mason, R. E., Sarinelli, J. M., & Brown-Guedira, G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genetics*, *20*(1). https://doi.org/10.1186/s12863-019-0785-1

Lozada, D. N., Ward, B. P., & Carter, A. H. (2020). Gains through selection for grain yield in a winter wheat breeding program. *PLoS ONE*, *15*(4). https://doi.org/10.1371/journal.pone.0221603

Mace, E. S., Buhariwalla, K. K., Buhariwalla, H. K., & Crouch, J. H. (2003). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant molecular biology reporter*, *21*, 459-460.

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499-511. https://doi.org/10.1038/nrg2796

Mirdita, V., He, S., Zhao, Y., Korzun, V., Bothe, R., Ebmeyer, E., Reif, J. C., & Jiang, Y. (2015). Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population. *Theoretical and Applied Genetics*, *128*(12), 2471–2481. https://doi.org/10.1007/s00122-015-2602-1

Newell, M. A., & Jannink, J. L. (2014). Genomic selection in plant breeding. *Crop breeding: Methods and protocols*, 117-130.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., … Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, *457*(7229), 551–556. https://doi.org/10.1038/nature07723

Prom, L. K., Perumal, R., Erpelding, J., Isakeit, T., Montes-Garcia, N., & Magill, C. W. (2009). A Pictorial Technique for Mass Screening of Sorghum Germplasm for Anthracnose (Colletotrichum sublineolum) Resistance. *The Open Agriculture Journal*, *3*(1), 20–25. https://doi.org/10.2174/1874331500903010020

Rio, S., Mary-Huard, T., Moreau, L., & Charcosset, A. (2019). Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. *Theoretical and Applied Genetics*, *132*(1), 81–96. https://doi.org/10.1007/s00122-018-3196-1

Sallam, A. H., & Smith, K. P. (2016). Genomic selection performs similarly to phenotypic selection in barley. *Crop Science*, *56*(6), 2871–2881. https://doi.org/10.2135/cropsci2015.09.0557

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., Mason, R. E., Babar, A., Harrison, S., Sutton, R., Griffey, C. A., & Brown-Guedira, G. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theoretical and Applied Genetics*, *132*(4), 1247–1261. https://doi.org/10.1007/s00122-019-03276-6

Thakur, R. P., Reddy, B. V. S., & Mathur, K. (2007). Screening techniques for sorghum diseases: Info Bull N. 76 Pp 1-92. *Patancheru*, *502*, 324.

Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, *6*(4), 330-340. https://doi.org/10.1016/j.cj.2018.03.001

Yabe, S., Yoshida, H., Kajiya-Kanegae, H., Yamasaki, M., Iwata, H., Ebana, K., Hayashi, T., & Nakagawa, H. (2018). Description of grain weight distribution leading to genomic selection for grain-filling characteristics in rice. *PLoS ONE*, *13*(11). https://doi.org/10.1371/journal.pone.0207627

Chapter 3 - Genome-wide association study of sorghum anthracnose disease (*Colletotrichum sublineolum*): Unraveling genetic variations and disease associations

Abstract

Like all other crops, sorghum also suffers from various biotic agents to a variable degree. Among the major biotic constraints in humid and sub-humid agroecologies of Ethiopia are foliar anthracnose and grain mold. In this study, we investigated the degree of SNP marker trait association for reaction to sorghum anthracnose. The study revealed significant associations on chromosomes 1, 2, 3, 4, 6, 8, and 10. Notably, chromosome 8 presented the most substantial genetic associations, with nine loci (S8_16230498, S8_16389715, S8_37466666, S8_39664335, S8_42468301, S8_42983043, S8_43776246, S8_7908407, and S8_570182) showing significant association with the anthracnose disease. Likewise, on chromosome 1, several loci (S1_20098885, S1_49271328, S1_49271328, S1_49454848, S1_50111275, S1_50294000, S1_50533418, S1_30455882, and S1_50721682) were significantly associated with the disease. Chromosome 2 also revealed two significant loci (S2_55407842 and S2_69275221) while only one locus (S3_53970498) on chromosome 3 displayed a significant association. Chromosomes 4 and 6 had two loci each (S4_62789639, S4_7538383, S6_25755784, and S6_55783758) that were significantly associated with the disease. The GWAS results shed light on the genetics of sorghum anthracnose disease reaction and highlight potential genomic regions influencing disease outcomes. Further investigation may be needed to validate the results and understand the underlying molecular mechanisms to facilitate the development of genetic tools for use in resistance breeding.

3.1 Introduction

Sorghum (*Sorghum bicolor*) is an important cereal crop grown worldwide for food, fodder, and as feedstock for industrial products. Millions of people in the semi-arid tropics of Africa and Asia rely on sorghum as a staple food. Ethiopia is among the largest producers and consumers of sorghum with a large proportion of the produce used for local consumption by primary producers. The crop is grown in almost all agroecological zones of the country with the semi-arid mid and low-altitude zones being the major sorghum belt primarily because other crops do not do so well under such conditions. Sorghum is also grown in several other milder agroecological zones along with other crops such as in the west and southwestern parts of the country. Despite the conducive weather and rich soils, sorghum production in these areas face peculiar challenges and one such challenge is a massive infection by the fungal pathogen *Colletotrichum sublineolum*. The disease often called leaf anthracnose is responsible for substantial loss of grain yield, and forage value leading to significant economic losses and food security concerns.

Because meager research funds are allocated for the most pressing problems, not much attention has been paid to solving the disease problem in sorghum. Breeding efforts by the Ethiopian national program have focused mainly on enhancing yield, drought tolerance, and resistance to the parasitic weed Striga and have registered modest success. Globally, conventional breeding has resulted in consistent genetic improvement contributing to an average yield increase of 27kg ha^{-1} yr^{-1} between 1963 and 2017 (Demarco et al., 2023). Recent discoveries in DNA technology have enhanced the understanding of plant genome architecture and provided additional tools for exploring genetic factors associated with traits of economic importance including host plant resistance to a variety of diseases such as anthracnose.

The advancements in genomics, matched with the rapid decline in genotyping cost, have opened new avenues to explore the genetic construct of complex traits, including disease resistance in plants. One important area where the application of genomic technology contributed to crop improvement is genome-wide association studies (GWAS). This approach has emerged as a valuable tool to uncover associations between genetic markers and specific phenotypic traits in diverse germplasm populations. While GWAS was developed and widely used in human genetics to uncover genetic factors associated with various human diseases, it has also gained popularity in plant genetics (Kumari et al., 2022).

Over the years, there has been a substantial body of research, delving into the broader application of GWAS. These studies have been conducted to uncover links between allelic or genotypic frequencies and a diverse range of plant traits, including biotic and abiotic stress resistance, yield-associated traits, and metabolic composition in annual and perennial crop species (Kumari et al., 2022).

Most GWAS studies conducted on crop biotic and abiotic stress resistance focused on wheat and rice ((Zhu et al., 2008; Muleta et al., 2017; Xiao et al., 2018; Galagedara et al., 2020; Satturu et al., 2020; Liu et al., 2021). More than 200 GWAS works have been published for different agronomic traits in wheat alone (Sehgal and Dreisigacker, 2022). Fungal diseases including stripe rust and fusarium head blight in wheat and rice blast were among the well-studied (Muleta et al., 2017; Li et al., 2019; Zeng et al., 2020; Ruan et al., 2020). In a study on 190 Ethiopian bread wheat lines, Muleta et al. (2017) identified 15 and 9 loci associated with stripe rust and stem rust resistance, respectively, that were mapped close to the already identified rust resistance genes. A total of thirty associated loci were identified for a rice blast disease in a study that involved 366 indica rice (Wang et al., 2014).

Some of the GWAS works conducted on sorghum include grain yield components, traits related to forage, fodder quality and biofuel, agroclimatic traits, anthracnose response in sorghum association panel, stalk rot, grain mold, striga resistance, and starch content. GWAS study conducted on sorghum diversity panel (390 entries) identified significant marker trait associations for yield per panicle, grain number per panicle and thousand grain weight (Boyles et al., 2016). Significant associations were also detected for panicle architecture and biomass traits (Spindel et al., 2018; Zhang et al., 2019). The study on 163 Senegalese sorghum accessions for sorghum anthracnose response has not identified any significant association for this trait (Ahn et al., 2021) while Cruet-Burgos et al. (2020) were able to detect two loci associated with anthracnose response in 114 recombinant inbred lines (RILs) derived from sorghum line SC 112-14.

In this study, we attempted to apply GWAS to unravel the genetic variations in Ethiopian sorghum landraces and discover marker-trait association for anthracnose resistance to identify candidate sources of resistance for use in breeding programs. Because the disease is caused by a complex of pathogens and our source germplasm consisted of diverse materials collected from different agroecologies, the GWAS approach was thought a good fit to identify multiple resistance sources at the same time. Hence, the objective of the study was to characterize the genetic architecture of sorghum landraces of Ethiopian origin, identify sources of resistance and map genomic regions underlying resistance to sorghum anthracnose.

3.2 Materials and methods

3.2.1 Genetic materials

The study was conducted using a large set of diverse local sorghum germplasm (946 accessions) collected from different parts of Ethiopia. The accessions were selected from the sorghum collections maintained by the Ethiopian Biodiversity Institute, the Ethiopian Institute of Agricultural Research, and other research institutions in the country. The materials also included local cultivars widely grown by farmers in different regions, released varieties as well as cultivars known for their unique attributes such as nutritional value, Striga resistance, and utilization.

3.2.2 The field experimental site

The accessions were evaluated for two consecutive seasons in 2015 and 2016 at Bako (9°08'N, 37°03'E), Oromia, Ethiopia. The study site was located at an elevation of 1642 m a.s.l. with a mean annual rainfall of 1281 mm and annual temperature of 19.7°C. Most of the precipitation comes during the growing period (June to September) when the temperature is milder than the annual average. The precipitation subsides after September, the temperature rises, and relative humidity remains high through the grain-filling period creating a uniquely favorable environment for leaf disease and grain mold development.

3.2.3 Experimental design and field management

The accessions were planted in a single-row plot of 3 m length. The distance between rows was 80cm while the between plant distance was 20 cm (Girma et al., 2019; Nida et al., 2019). The plots were fertilized according to the recommendation for the area. Phosphorus fertilizer was applied at planting at the rate of 100 kg/ha (DAP) (Diammonium phosphate) and nitrogen was applied at 50 kg/ha (urea) when the plants were at a knee height. Weeds were controlled manually.

3.2.4 Field data collection

In every row, five plants representing the accession were selected and tagged at flowering and all data were collected on those plants. Disease ratings were conducted as described by (Prom et al., 2009). In this method, a rating scale of 1 to 5 was used where 1 represents no signs of infection. A rating of 2 was assigned to accessions where hypersensitive reactions were visible with localized lesions and no acervuli were noticeable. A score of 3 was assigned to accessions with infected lower leaves and acervuli were present. Those accessions with infections progressed to middle leaves and visible acervuli were assigned a score of 4. A score of 5 was given to those with infections reaching up to the flag leaf with visible acervuli.

3.2.5 DNA extraction, genotyping, SNP calling and filtering, and data imputation.

DNA was extracted from a week-old seedling using cetyltrimethylammonium bromide (CTAB) DNA extraction protocol modified for 96 well plates (Mace et al., 2003). The DNA samples were digested with restriction enzyme *Ape*KI (recognition cite: 5'...G|CWGC...3') and 10 libraries were constructed for the 946 accessions (Girma et al., 2019; Nida et al., 2019). Genotyping of the libraries was conducted at the DNA Sequencing Facility of the University of Wisconsin with the Illumina HiSeq 2500 lane sequencing machine (Illumina Inc., San Diego, CA, USA). Sequence reads were aligned to the BTx623 sorghum reference genome version 3. Poorly performing SNPs and genotypes were removed by implementing quality control procedures. TASSEL-GBS 5.0 (Glaubitz et al., 2014) was used to call and filter SNP markers. The data was filtered to remove markers missing in more than 80% of the accessions and those with minor allele frequencies of < 1%. After filtering, a total of 410,727 SNPs were obtained. The remaining SNPs were imputed for missing data with BEAGLE V 4.1 (Browning and Browning, 2016). Finally, 33,981 SNPs were retained and used for GWAS analysis.

3.2.6 Genome-wide association studies (GWAS)

GWAS was conducted on 33,981 SNPs from 946 sorghum accessions evaluated for reaction to anthracnose disease. The mixed linear model (MLM) in the genome association and prediction integrated tool (GAPIT) with kinship matrix K + Q was used for the analysis (Lipka et al., 2012). The rrBLUP package was used in R statistical software (Endelman, 2011), by applying the following model (Yu et al., 2006).

$$y = x\beta + Zu + S\tau + e$$

where y is the vector of observed values of phenotypes (BLUEs), $x, Z, and\ S$ are design matrices (observations of the variables in the linear model), β is a vector of levels of fixed effects, u is a vector of levels of random effects (genetic marker effects), τ models the additive SNP effect as a fixed effect, and e is the residual effect. For association analysis, the disease score was used as the quantitative trait. The genome-wide significance threshold for 2015 data was set at a p-value of $< 6 \times 10^{-5}$ and $< 2 \times 10^{-5}$ for Benjamini-Hochberg values of 0.1 and 0.05 respectively. For the 2016 data, P-values of 3×10^{-5} and 1.3×10^{-5} respectively were used.

3.2.7 Candidate gene identification

SNPs significantly associated with anthracnose resistance were annotated by aligning to the sorghum bicolor reference genome version 3.11 in *Phytozome 13* (Goodstein et al., 2012). Genes within 100kb of the significant SNP were considered candidates. A literature search was conducted to validate the identified candidate genes for known functions or putative functions in disease resistance/susceptibility in sorghum or other crops.

3.3 Results

3.3.1 Anthracnose resistance in the Ethiopian sorghum germplasm collection

A substantial number of the accessions were shown to have resistance to foliar anthracnose disease. Of the 946 accessions included in the study, 187 had a score of 1 or 2 on a scale of 1-5 with 1 representing resistance. Whereas the bulk of the accessions (551) had a mean disease score rating of 2 to 3 and only 14 were rated 4 and above (Figure 3.1). So, nearly 20% of the accessions expressed resistance to the disease. No attempts were made to identify the strains or other foliar disease-causing pathogens other than *Colletotrichum sublineolum.* Broad-sense heritability was estimated to be 0.63.

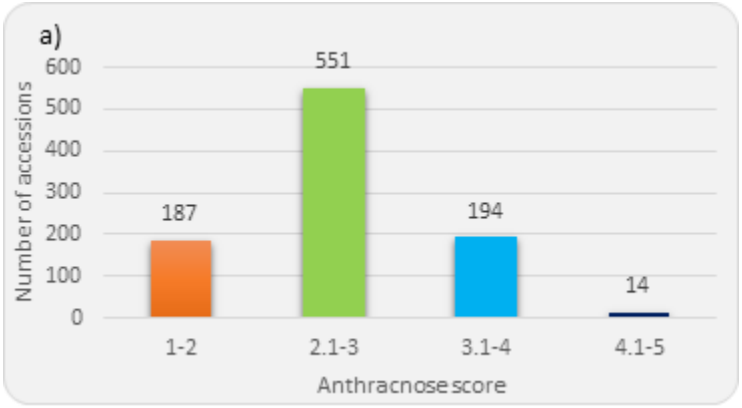


Figure 3.1. Frequency distribution of anthracnose score.
(1 = represents no signs of infection, 2 = the plants show hypersensitive reactions with local lesions and no acervuli, 3 = the bottom leaves were infected and acervuli were present, 4 = infections that progressed to middle leaves and with visible acervuli, 5 = infections reaching up to the flag leaf with visible acervuli).

3.3.2 Population structure

We noted that different sorghum accessions show different levels of resistance to fungal pathogens such as *Colletotrichum sublineolum*. This variation may mirror the genetic diversity present in the population. The co-evolutionary dynamics between sorghum and the pathogens may

be revealed by the genetic structure of accessions. We conducted ADMIXTURE analysis in R software using a hypothetical clusters/subpopulation size of K=20. The results indicated a continuous sharp decrease in cross-validation error as K was dropped and reached its lowest point at K=13 and then a slight increase at K=14 which was then followed by a continuous decrease (Figure 3.2). Thus, K=13 was considered the optimum number of clusters (Figure 3.3).

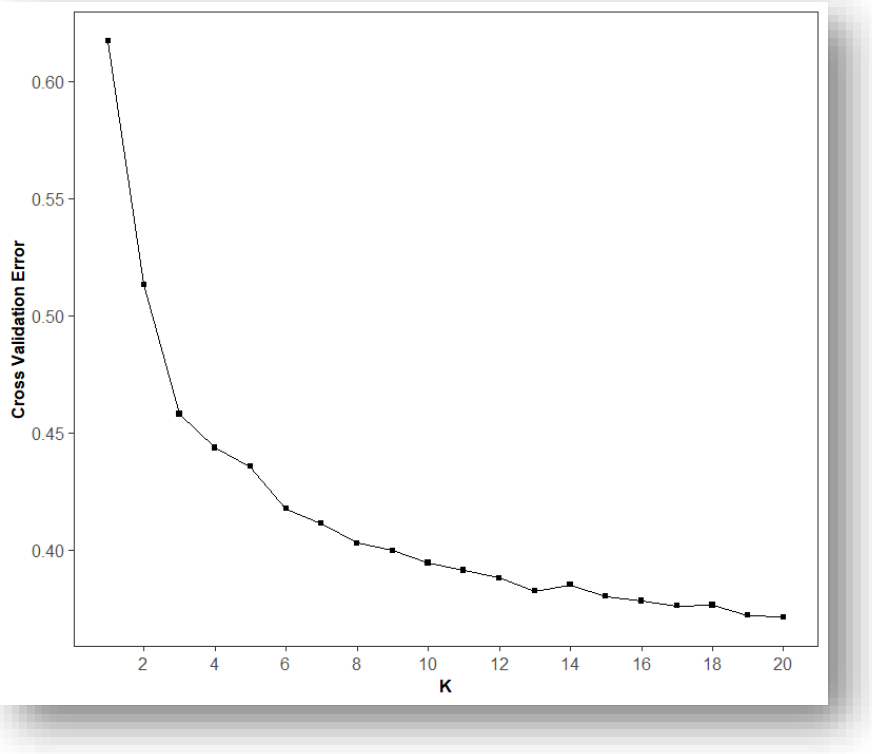


Figure 3.2. Cross-validation error rates and K values

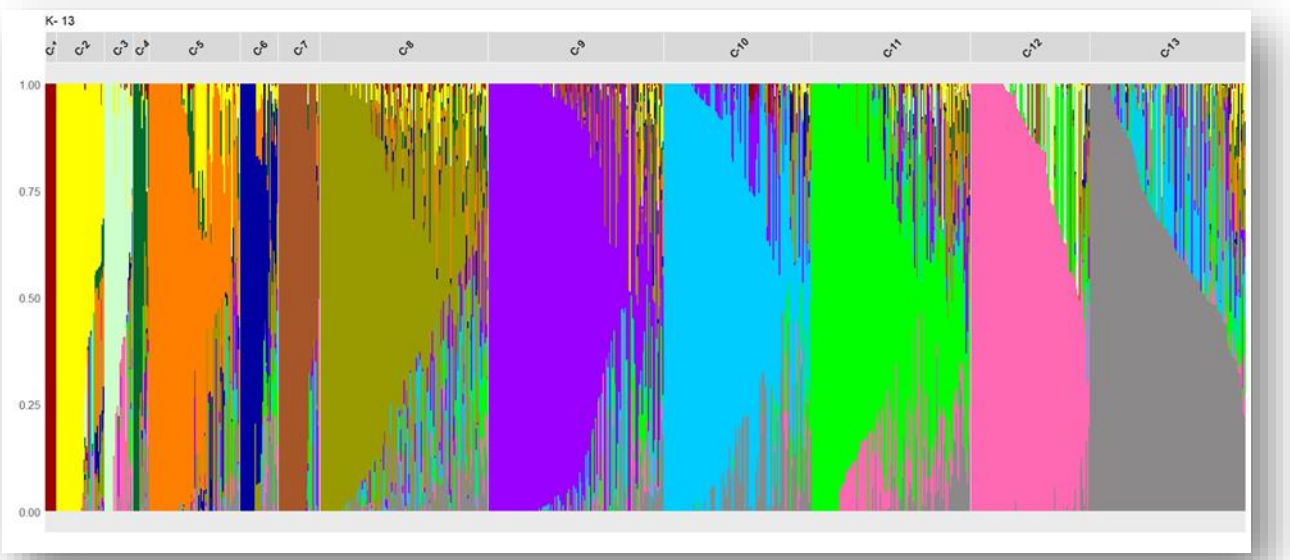


Figure 3.3. Population genetic structure of 946 sorghum accessions (C1 to C13 represent cluster 1 to cluster 13)

3.3.3 Identification of significant SNPs associated with sorghum anthracnose resistance.

A genome-wide association study (GWAS) was conducted to identify significant marker trait associations for resistance to anthracnose disease. In total, 23 SNPs strongly associated with anthracnose resistance were discovered (Bonferroni p-value < 0.05). However, results between the two test environments were not consistent. In 2015, most of the significant SNPs occurred on chromosome 1 with chromosomes 2, 4, 6, 8, and 10 also having 1-2 significant SNPs. Whereas in 2016 season, most of the significant SNPs were on chromosome 8 with chromosome 3 also harboring two significant SNPs. Overall, chromosomes 1, 2, 3, 4, 6, 8, and 10 carried SNPs significantly associated with leaf anthracnose in 2015 and 2016 seasons (Figure 3.4).

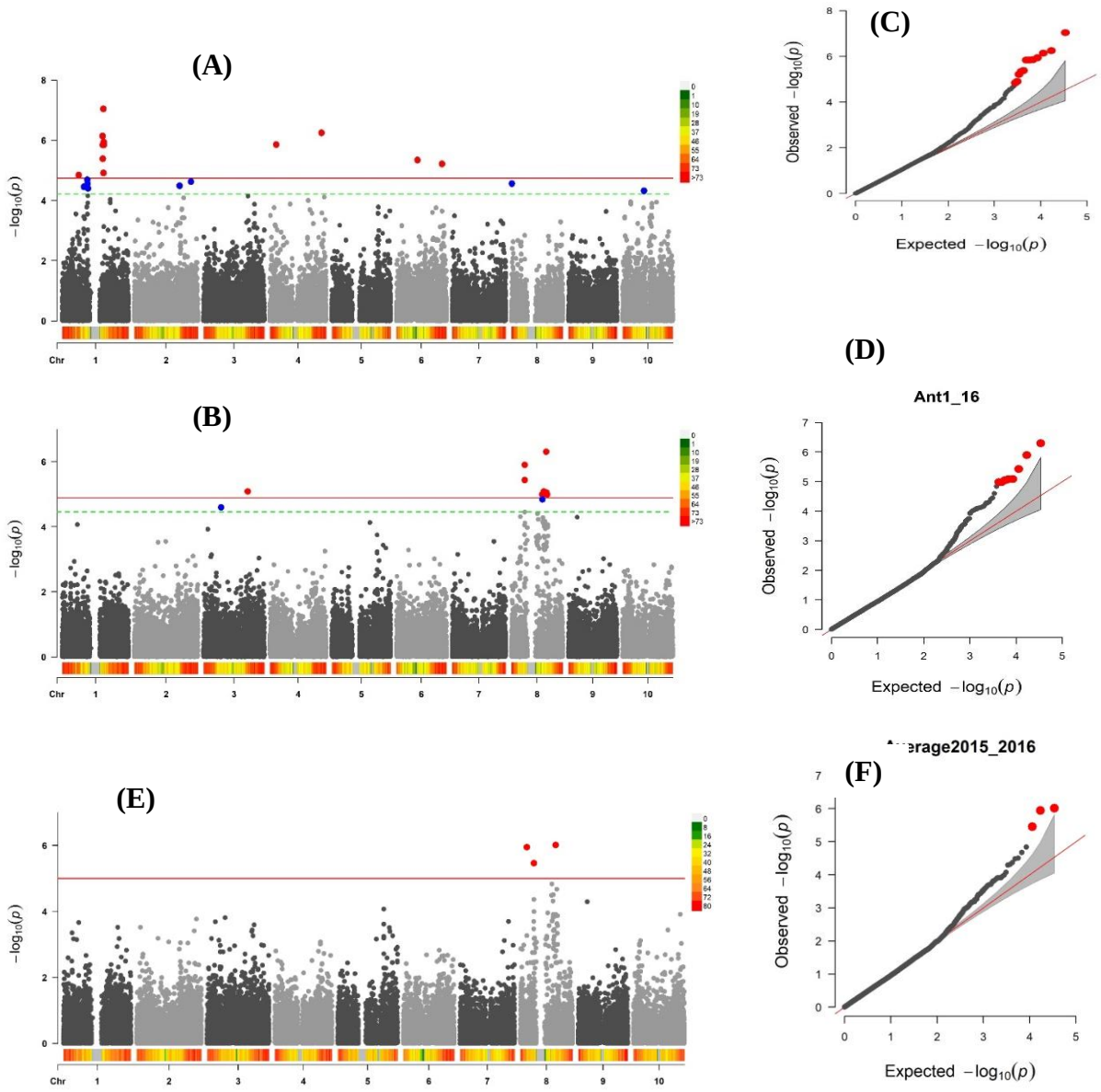


Figure 3.4. Manhattan and QQ plots for 2015, 2016, and the mean of the two years.

(A) and (B), Manhattan plots showing the significant SNPs of the genome-wide association of sorghum anthracnose resistance for 2015 and 2016, respectively. (C) and (D) QQ-plot (2015 and 2016 respectively). (E) and (F) represent the Manhattan plot and the QQ plot for the mean of the two years respectively. The horizontal red and green lines represent the FDR-adjusted p-values of <0.01 and <0.05 respectively.

The significant SNP loci on chromosome 1 are S1_20098885, S1_49271328, S1_49271328, S1_49454848, S1_50111275, S1_50294000, S1_50533418, S1_30455882, and S1_50721682. Two SNPs on chromosome 2, S2_55407842 and S2_69275221 were also significant at $P < 0.05$ but only one significant locus (S3_53970498) was identified on chromosome 3. Two SNPs on chromosome 4, S4_62789639 and S4_7538383, and another two on chromosome 6, S6_25755784 and S6_55783758, were identified. We also found significant associations at 9 loci on chromosome 8 (S8_16230498, S8_16389715, S8_37466666, S8_39664335, S8_42468301, S8_42983043, S8_43776246, S8_7908407, and S8_570182). Sixteen of these loci were known to have a significant role in plant response to pathogens. These are six loci on chromosome 1 (S1_30455882, S1_30455882, S1_30787810, S1_31259197, S1_50294000, S1_50721682), two on chromosome 2 (S2_55407842, S2_69275221) one each on chromosomes 3 and 4 (S3_53970498, S4_7538383), two on chromosome 6 (S6_55783758, S6_25755784), and 4 on chromosome 8 (S8_16230498, S8_16389715, S8_39664335, S8_570182) (Table 3.1).

Table 3.1. Genome-wide significant markers for anthracnose resistance in sorghum ($p < 0.05$).

Chr	Position	Genes	p-value	Annotated function	Distance (bp)
1	S1_30455882	Sobic.001G257300	2.04E-05	F-Box proteins in plants	72485
1	S1_30455882	Sobic.001G257800	2.04E-05	Myosin	18243
1	S1_30787810	Sobic.001G258500	2.75E-05	Zinc Finger FYVE Domain Containing Protein	36217
1	S1_31259197	Sobic.001G259700	3.99E-05	Jasmonate ZIM domain-containing protein (JAZ),	85867
1	S1_50294000	Sobic.001G265900	1.23E-05	Protein kinase domain (Pkinase) // Leucine Rich Repeat (LRR_1) // Leucine-rich repeat (LRR_8), like Proline-rich protein precursor	2000
1	S1_50721682	Sobic.001G266500	1.44E-06	Probable lipid transfer (LTP_2)	64535
2	S2_55407842	Sobic.002G175700	3.22E-05	Mitogen-Activated Kinase Kinase, Cascades in Plant Disease Resistance Signaling	69101
2	S2_69275221	Sobic.002G321100	2.38E-05	Leucine-rich repeat-containing Protein, Ras GTPases	0
2	S2_69275221	Sobic.002G320700	2.38E-05	Ankyrin Repeat Domain-Containing Protein 50	57499
3	S3_53970498	Sobic.003G207600	8.15E-06	F-box domain	92873
4	S4_7538383	Sobic.004G087750	1.4E-06	Zinc-binding in reverse transcriptase (zf-RVT),	105184
6	S6_55783758	Sobic.006G208400	6.1E-06	Nuclear pore complex protein Nup88	33204
6	S6_25755784	Sobic.006G040550	4.55E-06	Zinc Finger CCHC Domain Containing Protein,	99928
8	S8_16230498	Sobic.008G084600	3.7E-06	Nucleotide Kinase/like Adenylate kinase B, F-box protein	11983
8	S8_16389715	Sobic.008G084900	1.26E-06	Like D-mannose binding lectin family protein	46115
8	S8_39664335	Sobic.008G092300	8.22E-06	Zinc Finger CCCH Domain-Containing Protein 43,	4166
8	S8_570182	Sobic.008G006800	1.58E-06	Like Leucine Rich Repeat family protein, expressed	163
8	S8_570182	Sobic.008G007400	1.59E-06	Protein tyrosine kinase (Pkinase_Tyr) // Leucine-rich repeat N-terminal domain (LRRNT_2) // Leucine-rich repeat (LRR_8): like bacterial blight resistance protein XA26	23037
8	S8_570182	Sobic.008G007500	1.59E-06	Zinc Finger HIT Domain Containing Protein 1	26702
8	S8_570182	Sobic.008G008100	1.59E-06	Kelch motif (Kelch_1) // Development and cell death domain (Dev_Cell_Death)	50485
8	S8_570182	Sobic.008G008500	1.59E-06	Like Protein phosphatase 2c, putative, expressed	86663

Further analysis of the candidate genes within these regions indicated several putative genes that might have a role in anthracnose resistance. These genes include the zinc finger protein domains found near S1_20098885, S1_30455882, S4_62789639, S8_43776246, and S8_39664335, F-box domains located close to S1_20098885 and S1_30455882. Jasmonate ZIM domain-containing protein (JAZ), myosin, Ras GTPases, and Serine Protease Family S10 Serine Carboxypeptidase near S1_30455882. Leucine-rich repeat (LRR) was identified near S1_50294000 and S8_43776246 and Probable lipid transfer (LTP_2) near S1_50294000. B3 DNA binding domain (B3) // Auxin response factor (Auxin_resp), eukaryote-specific dsRNA binding protein, non-specific serine/threonine protein kinase/ threonine-specific protein kinase, nucleotide kinase, and salt stress response/antifungal (Stress-antifung) // protein tyrosine kinase (Pkinase_Tyr) were identified near S1_20098885. ADP-ribosylation factor 3, Kelch motif (Kelch_1) Development and cell death domain (Dev_Cell_Death), and protein phosphatase 2C 65-related were detected near S8_43776246.

3.4 Discussion

Under natural environments, plants are always exposed to a multitude of biotic and abiotic stresses which through time resulted in the evolution of complex defense systems against these stresses (Dixon and Lamb, 1990; Dangl and Jones, 2001). An important component of crop improvement is the manipulation of this natural defense system to understand these mechanisms and identify the underlying genes for use to develop disease and pest-resistant crop cultivars.

Results from our current study revealed multiple putative genes associated with anthracnose resistance in sorghum. The loci for the significant associations on chromosome 1 are S1_20098885, S1_49271328, S1_49271328, S1_49454848, S1_50111275, S1_50294000, S1_50533418, S1_30455882, and S1_50721682. Also, S2_55407842 and S2_69275221 on chromosome 2, S3_53970498 on chromosome 3, S4_62789639 and S4_7538383, on chromosome 4, and S6_25755784 and S6_55783758 on chromosome 6 were identified. We also found significant associations at 9 loci on chromosome 8 (S8_16230498, S8_16389715, S8_37466666, S8_39664335, S8_42468301, S8_42983043, S8_43776246, S8_7908407, and S8_570182). Previous studies on sorghum identified sites associated with anthracnose resistance on chromosomes 1, 4, 5, 6, 8, and 9 (Cuevas et al., 2018; Prom et al., 2019; Mengistu et al., 2021). This finding indicates that anthracnose is a truly complex disease with resistance genes found on almost all chromosomes suggesting significant variability among anthracnose-causing pathogens. A controlled study based on artificial inoculations might narrow down the complexity and allow pinpointing of specific genes/genomic regions associated with resistance to a specific pathogen species/strain.

The identification of candidate genes for each SNP was conducted in the Phytozome database (Goodstein et al., 2012). The putative genes identified in this study include leucine-rich

repeat (LRR), leucine-rich repeat receptor-like protein kinase, zinc finger protein domains, F-box domains, Jasmonate ZIM domain-containing protein (JAZ), myosin, Ras GTPases, and Serine Protease Family S10 Serine Carboxypeptidase, Probable lipid transfer (LTP_2), B3 DNA binding domain (B3) // Auxin response factor (Auxin_resp), eukaryote specific dsRNA binding protein, non-specific_serine/threonine protein kinase/threonine-specific protein kinase, nucleotide kinase, salt stress response/antifungal (Stress-antifung) // protein tyrosine kinase (Pkinase_Tyr), ADP-ribosylation factor 3, Kelch motif (Kelch_1) // Development and cell death domain (Dev_Cell_Death), protein phosphatase 2C 65-related, and protein kinase domains (Pkinase). These genes may be too many to work with for breeding programs. The structure of the population and the fact that the study was carried out under natural infection conditions where numerous pathogens and pathogen strains may be involved in causing the disease might have contributed to the large number of significant SNPs. The working population are accessions of diverse genetic background with limited chance of recombination as revealed by structure analysis (Figure 3.3).

Besides the likelihood of multiple accessions within the working population having unique resistance alleles and the possibility of several pathogen involved, the multiple layers of interaction between the host and pathogens before and after infection may have contributed to the detection of relatively large number significant SNPs. Plants have evolved mechanisms for recognizing potentially dangerous pathogens and responding to them before they cause damage. These mechanisms are of multiple layers and are deployed in phases. The first layer is innate immunity, also called basal resistance, and is responsible for protecting plants against most pathogens (Freeman and Beattie, 2008). Plants induce innate immunity when they detect microbe-associated molecular patterns (MAMPs). Chitin in fungi, and flagellin in bacteria, represent these highly conserved molecular patterns (Freeman and Beattie, 2008; Boller and Felix, 2009). Microbes

develop effector proteins that act to suppress these defenses in plants (Bittel and Robatzek, 2007; Newman et al., 2013; Ye and Murata, 2016). To tackle this, plants have evolved *R* proteins that help them detect the pathogen effectors and hence activate stronger defenses (Bent and Mackey, 2007). Common features of most of the disease resistance genes (R-genes) cloned so far are that they contain an amino-terminal domain, a nucleotide-binding site (NBS), and a leucine-rich-repeat (LRR) domain (Bent and Mackey, 2007). Other domains such as F-box domains (as in *Arabidopsis suppressor of nim1-1* (SON_1), and the ankyrin (ANK) domain-containing proteins also play a vital role in plant disease resistance (Kim and Delaney, 2002; Wu et al., 2009; Mou et al., 2013).

Leucine-rich repeat (LRR) proteins are involved in specific protein–protein interactions and are known to have a significant role in plant defenses. The SNPs (genes) that were associated with LRR were S1_50294000 (Sobic.001G265900), S2_69275221 (Sobic.002G321100), S8_570182 (Sobic.008G006800), S8_570182 (Sobic.008G007400). Nucleotide-binding leucine-rich repeat immune receptors (NLR) play an important role in protecting plants against pathogens. Here plants trigger programmed cell death (PCD) also called hypersensitive response (HR) to cease the spread of the infection (Heath, 2000; Caplan et al., 2008; Deng et al., 2018). Three types of NLRs were identified in plants based on their N-terminal domains. There are the TIR (Toll/interleukin-1 receptor) _NLRs (TNLs) and two non-TNLs: (coiled-coil (CC) motif, and a resistance to powdery mildew 8 (RPW8) domain (Shao et al., 2016; Zhong and Cheng, 2016; Maruta et al., 2022).

First described in animals as a protein-protein interaction model, the Toll/interleukin-1 receptor (TIR) is known to function in the immunity response of all forms of life (Essuman et al., 2022; Li et al., 2023).

F-box domains have been reported to play a significant role in plants' response to biotic and abiotic stresses. The *Arabidopsis* F-box protein MAX2 is known to contribute to resistance against bacterial infection (Piisilä et al., 2015). In our study, S3_53970498 located on chromosome 3 emerged as a significant gene contributing to resistance against one or more fungal pathogens causing anthracnose.

The ANK protein domain is reported to play important roles in biotic and abiotic responses in pepper (*Capsicum*), *Arabidopsis*, rice (*Oryza sativa* L.*)*, potato (*Solanum tuberosum* L.), and maize (*Zea mays* L.) (Yan et al., 2002; Seong et al., 2007; Sakamoto et al., 2008; Wu et al., 2009; Mou et al., 2013; Jiang et al., 2013).

SNP S2_69275221is 57499 bp away from Sobic.002G320700, which is Ankyrin Repeat Domain-Containing Protein 50. ANK protein groups may be involved in antioxidant metabolism regulation when pathogens attack and in the presence of other stresses as in the case of *Arabidopsis AKR2* protein ((Yan et al., 2002; Sakamoto et al., 2008). CaKR1 (*Capsicum annuum* ankyrin-repeat domain C3H1 zinc finger protein (CaKR1)), ANK protein in pepper, is reported to play an important role in biotic and abiotic stress responses (Seong et al., 2007). Accelerated cell death6-1 (acd6-1), a novel ANK protein, leads to spontaneous cell death and increased disease resistance. Responsiveness to salicylic acid, a major defense signal, is known to increase in the presence of acd6-1 (Lu et al., 2003). Ankyrin-repeat proteins have also been reported in maize. These proteins were also reported to regulate the jasmonic and salicylic acid pathways which play an important role in basal defense against *Magnaporthe oryzae* in rice and late blight in potatoes (Wu et al., 2009; Mou et al., 2013; Jiang et al., 2013).

The genes Sobic.001G257300, Sobic.003G207600 and Sobic.004G088300 associated with GBS SNPs S1_30455882, S3_53970498 and S4_7538383, respectively, were known to be F-

box/LRR-Repeat Protein 15-related. These proteins are one of the largest protein superfamilies in plants (Xu et al., 2009). A study on sorghum anthracnose in the US sorghum association panel also reported F-box domains as putative genes (Cuevas et al., 2018). F-box proteins target specific pathogen effectors or host proteins involved in pathogen virulence for degradation and hence act as key regulators of immune responses (Göhre et al., 2008). The pathogen is prevented from manipulating the host immunity by this degradation. On the other hand, these proteins can also enhance the ability of plants to recognize and respond to pathogens. In a study on *Arabidopsis thaliana*, F-box protein 22 (FBP22) was reported to be involved in plant defense by targeting the bacterial effector protein AVrPtoB, known to suppress plant immune responses (Göhre et al., 2008).

The other domains identified in this study as putative genes were the zinc finger protein domains. This is a superfamily involved in plant growth and development. OH et al. (2005) were able to identify genes encoding putative EPF-type zinc-finger proteins that were expressed in hot pepper in the presence of *Xanthomonas axonopodis* pv. *glycines*. EPF-type zinc-finger proteins are reported to be important in plant defense and the expression of pathogenesis-related (PR) and defense-related genes (OH et al., 2005). WRKY zinc-finger families are among families that constitute transcription factors associated with plant defense response (Chen and Chen, 2002; Robatzek and Somssich, 2002).

3.5 Conclusion

The GWAS study highlights several specific loci on different chromosomes associated with sorghum anthracnose resistance with chromosomes 1 and 8 containing the largest number of significant loci. These findings provide valuable insights into the genetic basis of sorghum anthracnose resistance and could potentially contribute to sorghum breeding endeavors. Understanding these molecular mechanisms can pave the way for further research to improve the resistance of sorghum and other crops to anthracnose and other diseases, thus contributing to agricultural sustainability and food security.

References

Ahn, E., Prom, L. K., Hu, Z., Odvody, G., & Magill, C. (2021). Genome-wide association analysis for response of Senegalese sorghum accessions to Texas isolates of anthracnose. *Plant Genome*, *14*(2). https://doi.org/10.1002/tpg2.20097

Bent, A. F., & Mackey, D. (2007). Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology*, *45*, 399–436. https://doi.org/10.1146/annurev.phyto.45.062806.094427

Bittel, P., & Robatzek, S. (2007). Microbe-associated molecular patterns (MAMPs) probe plant immunity. In *Current Opinion in Plant Biology* (Vol. 10, Issue 4, pp. 335–341). https://doi.org/10.1016/j.pbi.2007.04.021

Boller, T., & Felix, G. (2009). A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology*, *60*, 379–407. https://doi.org/10.1146/annurev.arplant.57.032905.105346

Boyles, R. E., Cooper, E. A., Myers, M. T., Brenton, Z., Rauh, B. L., Morris, G. P., & Kresovich, S. (2016). Genome-wide association studies of grain yield components in diverse sorghum germplasm. *The Plant Genome*, *9*(2). https://doi.org/10.3835/plantgenome2015.09.0091

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *American Journal of Human Genetics*, *98*(1), 116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Caplan, J., Padmanabhan, M., & Dinesh-Kumar, S. P. (2008). Plant NB-LRR immune receptors: From recognition to transcriptional reprogramming. In *Cell Host and Microbe* (Vol. 3, Issue 3, pp. 126–135). Cell Press. https://doi.org/10.1016/j.chom.2008.02.010

Chen, C., & Chen, Z. (2002). Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced Arabidopsis transcription factor. *Plant Physiology*, *129*(2), 706–716. https://doi.org/10.1104/pp.001057

Cruet-Burgos, C. M., Cuevas, H. E., Prom, L. K., Knoll, J. E., Stutts, L. R., & Vermerris, W. (2020). Genomic dissection of anthracnose (*Colletotrichum sublineolum*) resistance response in sorghum differential line SC112-14. *G3: Genes, Genomes, Genetics*, *10*(4), 1403–1412. https://doi.org/10.1534/g3.120.401121

Cuevas, H. E., Prom, L. K., Cooper, E. A., Knoll, J. E., & Ni, X. (2018). Genome-wide association mapping of anthracnose (*Colletotrichum sublineolum*) resistance in the U.S. sorghum association panel. *The Plant Genome*, *11*(2), 0. https://doi.org/10.3835/plantgenome2017.11.0099

Dangl, J., & Jones, J. (2001). Plant pathogens and integrated defense responses to infection. *Nature*, *411*, 826–833. www.nature.com

Demarco, P. A., Mayor, L., Rotundo, J. L., Prasad, P. V. V., Morris, G. P., Fernandez, J. A., Tamagno, S., Hammer, G., Messina, C. D., & Ciampitti, I. A. (2023). Retrospective study in U.S. commercial sorghum breeding: II. Physiological changes associated to yield gain. *Crop Science*, *63*(2), 867–878. https://doi.org/10.1002/csc2.20845

Deng, Y., Wang, J., Tung, J., Liu, D., Zhou, Y., He, S., Du, Y., Baker, B., & Li, F. (2018). A role for small RNA in regulating innate immunity during plant growth. *PLoS Pathogens*, *14*(1). https://doi.org/10.1371/journal.ppat.1006756

Dixon, R. A., & Lamb, C. J. (1990). Molecular communication in interactions between plants and microbial pathogens. In *Annu. Rev. Plant Physiol. Plant Mol. Bioi* (Vol. 41). www.annualreviews.org

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Essuman, K., Milbrandt, J., Dangl, J. L., & Nishimura, M. T. (2022). Shared TIR enzymatic functions regulate cell death and immunity across the tree of life. In *Science* (Vol. 377, Issue 6605). American Association for the Advancement of Science. https://doi.org/10.1126/science.abo0001

Freeman, B., & Beattie, G. (2008). An overview of plant defenses against pathogens and herbivores. *The Plant Health Instructor*. https://doi.org/10.1094/phi-i-2008-0226-01

Galagedara, N., Liu, Y., Fiedler, J., Shi, G., Chiao, S., Xu, S. S., Faris, J. D., Li, X., & Liu, Z. (2020). Genome-wide association mapping of tan spot resistance in a worldwide collection of durum wheat. *Theoretical and Applied Genetics*, *133*(7), 2227–2237. https://doi.org/10.1007/s00122-020-03593-1

Girma, G., Nida, H., Seyoum, A., Mekonen, M., Nega, A., Lule, D., Dessalegn, K., Bekele, A., Gebreyohannes, A., Adeyanju, A., Tirfessa, A., Ayana, G., Taddese, T., Mekbib, F., Belete, K., Tesso, T., Ejeta, G., & Mengiste, T. (2019). A large-scale genome-wide association analysis of Ethiopian sorghum landrace collection reveals loci associated with important traits. *Frontiers in Plant Science*, *10*, 449171. https://doi.org/10.3389/fpls.2019.00691

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high-capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090346

Göhre, V., Spallek, T., Häweker, H., Mersmann, S., Mentzel, T., Boller, T., de Torres, M., Mansfield, J. W., & Robatzek, S. (2008). Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. *Current Biology*, *18*(23), 1824–1832. https://doi.org/10.1016/j.cub.2008.10.063

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1). https://doi.org/10.1093/nar/gkr944

Heath M. C. (2000). Hypersensitive response-related death. *Plant molecular biology*, *44*(3), 321–334. https://doi.org/10.1023/a:1026592509060

Jiang, H., Wu, Q., Jin, J., Sheng, L., Yan, H., Cheng, B., & Zhu, S. (2013). Genome-wide identification and expression profiling of ankyrin-repeat gene family in maize. *Development Genes and Evolution*, *223*(5), 303–318. https://doi.org/10.1007/s00427-013-0447-7

Kim, H. S., & Delaney, T. P. (2002). Arabidopsis SON1 is an F-box protein that regulates a novel induced defense response independent of both salicylic acid and systemic acquired resistance. *Plant Cell*, *14*(7), 1469–1482. https://doi.org/10.1105/tpc.001867

Kumari, M., Muduli, L., Meher, P. K., & Pradhan, S. K. (2022). Genome-Wide Association Study (GWAS) for Trait Analysis in Crops. In *Genomics of Cereal Crops* (pp. 295-307). New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-2533-0_15

Li, H., Thrash, A., Tang, J. D., He, L., Yan, J., & Warburton, M. L. (2019). Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant Journal*, *98*(5), 853–863. https://doi.org/10.1111/tpj.14282

Li, S., Manik, M. K., Shi, Y., Kobe, B., & Ve, T. (2023). Toll/interleukin-1 receptor domains in bacterial and plant immunity. In *Current Opinion in Microbiology* (Vol. 74). Elsevier Ltd. https://doi.org/10.1016/j.mib.2023.102316

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, Q., Lan, G., Zhu, Y., Chen, K., Shen, C., Zhao, X., Zhang, F., Xu, J., & Li, Z. (2021). Genome-wide association study on resistance to rice black-streaked dwarf disease caused by rice black-streaked dwarf virus. *Plant Disease*, *105*(3), 607–615. https://doi.org/10.1094/PDIS-10-19-2263-RE

Lu, H., Rate, D. N., Song, J. T., & Greenberg, J. T. (2003). ACD6, a novel Ankyrin protein, is a regulator and an effector of salicylic acid signaling in the Arabidopsis defense response. *Plant Cell*, *15*(10), 2408–2420. https://doi.org/10.1105/tpc.015412

Mace, E. S., Buhariwalla, K. K., Buhariwalla, H. K., & Crouch, J. H. (2003). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant molecular biology reporter*, *21*, 459-460.

Maruta, N., Burdett, H., Lim, B. Y. J., Hu, X., Desa, S., Manik, M. K., & Kobe, B. (2022). Structural basis of NLR activation and innate immune signaling in plants. In *Immunogenetics 74*(1), 5-26. *Springer Science and Business Media Deutschland* GmbH. https://doi.org/10.1007/s00251-021-01242-5

Mengistu, G., Shimelis, H., Assefa, E., & Lule, D. (2021). Genome-wide association analysis of anthracnose resistance in sorghum [Sorghum bicolor (L.) Moench]. *PLoS ONE*, *16*(12). https://doi.org/10.1371/journal.pone.0261461

Mou, S., Liu, Z., Guan, D., Qiu, A., Lai, Y., & He, S. (2013). Functional analysis and expressional characterization of Rice Ankyrin repeat-containing protein, OsPIANK1, in basal defense against Magnaporthe oryzae attack. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059699

Muleta, K. T., Rouse, M. N., Rynearson, S., Chen, X., Buta, B. G., & Pumphrey, M. O. (2017). Characterization of molecular diversity and genome-wide mapping of loci associated with resistance to stripe rust and stem rust in Ethiopian bread wheat accessions. *BMC Plant Biology*, *17*(1). https://doi.org/10.1186/s12870-017-1082-7

Newman, M. A., Sundelin, T., Nielsen, J. T., & Erbs, G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. In *Frontiers in Plant Science*, 4, 139. Frontiers Research Foundation. https://doi.org/10.3389/fpls.2013.00139

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

Oh, S.-K., Park, J. M., Joung, Y. H., Lee, S., Chung, E., Kim, S.-Y., Yu, S. H., & Choi, D. (2005). A plant EPF-type zinc-finger protein, CaPIF1, involved in defense against pathogens. *Molecular Plant Pathology*, *6*(3), 269–285. https://doi.org/10.1111/J.1364-3703.2004.00284.X

Piisilä, M., Keceli, M. A., Brader, G., Jakobson, L., Jöesaar, I., Sipari, N., Kollist, H., Palva, E. T., & Kariola, T. (2015). The F-box protein MAX2 contributes to resistance to bacterial phytopathogens in Arabidopsis thaliana. *BMC Plant Biology*, *15*(1). https://doi.org/10.1186/s12870-015-0434-4

Prom, L. K., Ahn, E., Isakeit, T., & Magill, C. (2019). GWAS analysis of sorghum association panel lines identifies SNPs associated with disease response to Texas isolates of *Colletotrichum sublineola*. *Theoretical and Applied Genetics*, *132*(5), 1389–1396. https://doi.org/10.1007/s00122-019-03285-5

Robatzek, S., & Somssich, I. E. (2002). Targets of AtWRKY6 regulation during plant senescence and pathogen defense. *Genes and Development*, *16*(9), 1139–1149. https://doi.org/10.1101/gad.222702

Ruan, Y., Zhang, W., Knox, R. E., Berraies, S., Campbell, H. L., Ragupathy, R., Boyle, K., Polley, B., Henriquez, M. A., Burt, A., Kumar, S., Cuthbert, R. D., Fobert, P. R., Buerstmayr, H., & DePauw, R. M. (2020). Characterization of the genetic architecture for fusarium head blight resistance in durum wheat: The complex association of resistance, flowering time, and height genes. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.592064

Sakamoto, H., Matsuda, O., & Iba, K. (2008). ITN1, a novel gene encoding an ankyrin-repeat protein that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance in Arabidopsis thaliana. *Plant Journal*, *56*(3), 411–422. https://doi.org/10.1111/j.1365-313X.2008.03614.x

Satturu, V., Vattikuti, J. L., Durga Sai, J., Kumar, A., Singh, R. K., Srinivas Prasad, M., Zaw, H., Jubay, M. L., Satish, L., Rathore, A., Mulinti, S., Ishwarya Lakshmi, V. G., Chakraborty, A., & Thirunavukkarasu, N. (2020). Multiple genome wide association mapping models identify quantitative trait nucleotides for brown planthopper (*Nilaparvata lugens*) resistance in MAGIC Indica population of rice. *Vaccines*, *8*(4), 1–16. https://doi.org/10.3390/vaccines8040608

Sehgal, D., & Dreisigacker, S. (2022). GWAS case studies in wheat. In *Methods in Molecular Biology* 2481, 341–351. Humana Press Inc. https://doi.org/10.1007/978-1-0716-2237-7_19

Seong, E. S., Cho, H. S., Choi, D., Joung, Y. H., Lim, C. K., Hur, J. H., & Wang, M. H. (2007). Tomato plants overexpressing CaKR1 enhanced tolerance to salt and oxidative stress. *Biochemical and Biophysical Research Communications*, *363*(4), 983–988. https://doi.org/10.1016/j.bbrc.2007.09.104

Shao, Z. Q., Xue, J. Y., Wu, P., Zhang, Y. M., Wu, Y., Hang, Y. Y., Wang, B., & Chen, J. Q. (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiology*, *170*(4), 2095–2109. https://doi.org/10.1104/pp.15.01487

Spindel, J. E., Dahlberg, J., Colgan, M., Hollingsworth, J., Sievert, J., Staggenborg, S. H., Hutmacher, R., Jansson, C., & Vogel, J. P. (2018). Association mapping by aerial drone reveals 213 genetic associations for Sorghum bicolor biomass traits under drought. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-5055-5

Wang, C., Yang, Y., Yuan, X., Xu, Q., Feng, Y., Yu, H., Wang, Y., & Wei, X. (2014). Genome-wide association study of blast resistance in indica rice. *BMC Plant Biology*, *14*(1). https://doi.org/10.1186/s12870-014-0311-6

Wu, T., Tian, Z., Liu, J., Yao, C., & Xie, C. (2009). A novel ankyrin repeat-rich gene in potato, star, involved in response to late blight. *Biochemical Genetics*, *47*(5–6), 439–450. https://doi.org/10.1007/s10528-009-9238-2

Xiao, N., Gao, Y., Qian, H., Gao, Q., Wu, Y., Zhang, D., Zhang, X., Yu, L., Li, Y., Pan, C., Liu, G., Zhou, C., Jiang, M., Huang, N., Dai, Z., Liang, C., Chen, Z., Chen, J., & Li, A. (2018). Identification of genes related to cold tolerance and a functional allele that confers cold tolerance. *Plant Physiology*, *177*(3), 1108–1123. https://doi.org/10.1104/pp.18.00209

Xu, G., Ma, H., Nei, M., & Kong, H. (2009). Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. *Proceedings of the National Academy of Sciences*, *106.3*, 835–840.

Yan, J., Wang, J., & Zhang, H. (2002). An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant Journal*, *29*(2), 193–202. https://doi.org/10.1046/j.0960-7412.2001.01205.x

Ye, W., & Murata, Y. (2016). Microbe associated molecular pattern signaling in guard cells. In *Frontiers in Plant Science*, 7, 179654. Frontiers Media S.A. https://doi.org/10.3389/fpls.2016.00583

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., Tang, T., Wang, Y., Zha, S., Gao, M., Li, C., Shu, C., Wei, Z., Qimei, W., Basang, Y., Dunzhu, J., Li, Z., Bai, L., Shi, J., … Nyima, T. (2020). Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. *Molecular Plant*, *13*(1), 112–127. https://doi.org/10.1016/j.molp.2019.10.009

Zhang, R., Zhou, Y., Yue, Z., Chen, X., Cao, X., Ai, X., Jiang, B., & Xing, Y. (2019). The leaf-air temperature difference reflects the variation in water status and photosynthesis of sorghum under waterlogged conditions. *PLoS ONE*, *14*(7). https://doi.org/10.1371/journal.pone.0219209

Zhong, Y., & Cheng, Z. M. (2016). A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. *Scientific Reports*, *6*. https://doi.org/10.1038/srep32923

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, *1*(1). https://doi.org/10.3835/plantgenome2008.02.0089

Chapter 4 - Unlocking genetic synergy for sorghum anthracnose resistance: A combining ability perspective

Abstract

Sorghum anthracnose, caused by the fungal pathogen *Colletotrichum sublineolum*, poses a significant threat to sorghum production worldwide. Anthracnose can lead to a devastating yield loss, rendering it imperative to enhance anthracnose resistance in sorghum varieties. The objective of the current study was to identify stable sources of resistance to the disease and determine their potential as parental sources to breed resistant varieties/hybrids. Four seed parents and ten pollinators of variable disease reaction were intercrossed using a Design II mating scheme. The resulting hybrids and the inbred parents were evaluated under five environments. Data were collected on various agronomic traits, including flowering time (DTF), plant height (PTH), panicle weight (PKW), grain yield per panicle (GYP), hundred-grain weight (HGW), and resistance to anthracnose (ANT). The disease was initiated using both artificial inoculation at one of the locations and natural infection in others. Disease rating was performed using a scale of 1 to 5 (1 representing resistant and 5 susceptible). The results showed significant effect of genotypes and its components for all traits except inbred effect for HGW and M vs F inbred effect for DTF and HGW. The environment also had a significant effect on all traits except HGW. The interaction of environment with genotypes and its components were also significant for all traits except male vs. F × E, F × E and I vs. H × E for GYP, PTH and PKW, respectively. Combining ability analysis identified pollinator parents, R245, R604, and R633 and seed parents B539, and B553 as having significant GCA for anthracnose resistance. Likewise, hybrids A539 × R669, A527 × R672, A531 × R672 and A539 × R588 had highly significant specific combining ability (SCA) for the trait. This study provides useful insight on resistance and combining ability of elite sorghum parental

lines. These lines are currently being used by seed industries for commercial hybrid development and the study provides additional resources for seed producers and growers on potential adaptation of hybrids derived from these lines to high foliar disease environments.

4.1 Introduction

Sorghum (*Sorghum bicolo*r L. Moench) is the main food for more than half a billion people in tropical and subtropical regions of Africaand Asia. As the center of origin and diversity for the crop, numerous forms of cultivated sorghum and their wild relatives occur in sympatry across the sub-Saharan Africa (De Wet, 1978; Ananda et al., 2020). Sorghum is also a major source of feed and biofuel feedstock. In the United States, animal feed and biofuel production are the two major sectors where sorghum is mostly utilized. Moreover, its gluten-free attribute has been gaining momentum and sorghum is poised to become an important food grain especially with the growing concern of sustainability due to climate change (Taylor et al., 2006; Stefoska-Needham et al., 2015).

Globally, about sixty-one million tons of sorghum was produced on 40 million hectares in 2021 with an average productivity of 1.5 tons per hectare (FAO, 2021). It is cultivated in more than one hundred countries, with the top 10 producers being the USA, Nigeria, India, Ethiopia, Mexico, Sudan, Argentina, China, Brazil, and Australia (FAO, 2021). These countries together contribute about 77% of world sorghum production (Aruna and Cheruku, 2018; Upadhyaya et al., 2019; Hao et al., 2021). The United States is the largest producer of grain sorghum with 11.4 million tons produced on 2.6 million hectares in 2021 (FAO, 2021). Sorghum ranks as the fourth most important grain crop in the United States, trailing behind the primary agricultural staples corn *(Zea mays* L.), soybean [*Glycine max* (L.) Merr.], and wheat (*Triticum aestivum* L.). Kansas is the top sorghum-producing state accounting for 64% of the total grain volume produced in the nation.

Other major states producing sorghum in the order of total production are Texas, Colorado, South Dakota, Nebraska, Oklahoma, Missouri, New Mexico, Illinois, and Georgia.

Although sorghum is known for its relative resistance/tolerance to pests and diseases and wide ecological adaptability, various biotic and abiotic stresses constantly challenge its cultivation (Tesso et al., 2012; Huang, 2018; Zhang et al., 2019). Anthracnose caused by the fungal pathogen *Colletotrichum sublineolum* is a significant threat for sorghum production in environments where warm weather and high rainfall converge (Chala et al., 2011). Yield losses due to anthracnose is significant and losses of up to 70% or more have been reported in Malawi and Brazil (Thomas et al., 1996; Cota et al., 2017). This yield loss can be mitigated by implementing proper management practices including the use of resistant varieties, cultural practices, and fungicides (Hess, 2002; Mofokeng et al., 2017). The use of fungicide as a control option has significant economic implications and poses risk to environment and human health. Host-plant resistance is deemed most effective, sustainable, and environmentally friendly management option (Abreha et al., 2021). Thus, development of resistant cultivars has always been an important breeding target for this and other economic diseases affecting crop plants.

Resistance breeding involves identification of stable sources of resistance, preferably in agronomically superior background, and evaluation of the sources for their suitability as breeding parent. Resistance sources that combine well with series of other lines are often preferred as breeding parents. The Kansas State sorghum breeding program has developed a series of parental lines that produce outstanding hybrids. Many of these lines have superior grain quality and moderate resistance to the emerging insect pest, sugar cane aphid (sorghum aphid). Several years of observations under high disease environment in Puerto Rico revealed that some of these lines also possess strong resistance to foliar diseases. Thus, the objective of the current study was to

identify stable sources of resistance among the existing elite parental lines and determine their potential as parental sources for breeding resistant varieties/hybrids. In other words, it was to confirm the disease reaction of selected parental lines and evaluate their combining ability for the trait. Sources with significant general combining ability (GCA) are considered stable as the resistance is likely under additive genetic control and the trait is heritable, hence, are preferred as breeding parents. Provided they have the desired agronomic merit, such lines can be directly used as a parent for production of disease resistant hybrids for use by farmers.

Past studies on combining ability for anthracnose resistance identified lines and hybrids with good GCA and SCA for the trait (Kenga et al., 2004b; Mengistu et al., 2010, 2020; Jane Da Rocha et al., 2018). We anticipate the current study to generate useful information on anthracnose disease reaction of the current releases from Kansas State University and determine the mode of inheritance of the resistance trait.

4.2 Material and methods

4.2.1 Plant materials

The experiment comprised four female parental lines (B527, B531, B539, and B553) and ten pollinator lines (R-lines) which included R245, R588, R604, R611, R618, R633, R651, R653, R669, and R672. The materials were selected depending on their varied responses to anthracnose when grown under disease pressure in tropical winter nursery in Puerto Rico. The materials were developed by Kansas State University from diverse pedigrees at different times, tested in hybrid combination across seasons and locations, and selected and released based on the combined data on agronomic performance. The parental lines were intercrossed in Design II mating scheme to generate a total of 40 hybrids. The hybrids along with the parents, altogether 54 entries, were evaluated across locations under natural and artificial disease pressure.

4.2.2 Experimental design and field management

The experiment was laid out in a randomized complete block design (RCBD). The evaluation was carried out under five disease environments: the 2018/2019 winter nursery and 2019 spring nursery at Yela Farms, Guayanilla, and USDA-ARS, Tropical Agriculture Research Station in Isabela, Puerto Rico, and summer 2020 at Tifton, Georgia. The tests were conducted in three replications using a 5m long single row plot. The plots were spaced 0.75m apart and the average distance between plants was 20 cm. Standard sorghum agronomic procedures were followed for seed preparation, planting, and other managements throughout the cropping seasons.

Disease initiation and development at Guayanilla and Isabella locations in Puerto Rico were left for nature to take course. Whereas at Tifton, Georgia, the plots were artificially inoculated using *C. sublineolum*-infected grains following the procedures described by Prom et al. (2009). The inoculum was prepared from a mixture of local isolates collected by the USDA-ARS-Tifton.

4.2.3 Data collection and statistical analysis

Data were collected on common agronomic parameters including days to flowering (DTF), plant height (PTH), and yield components such as panicle weight (PKW), grain yield per panicle (GYP) and hundred-grain weight (HGW) and anthracnose disease severity (ANT). Plant height and days to flowering were recorded on plot basis whereas yield components and ANT were recorded on five randomly tagged main tillers in a plot and taking their average. Individual parameters were measured as follow:

DTF was recorded as the number of days from planting to when 50% of the plants in a plot reached half bloom.

PTH the length (cm) of the plant measured from the base (soil surface level) to the top of the panicle.

PKW was estimated as the mean weight (g) of five panicles harvested from the tagged plants.

GYP is the average weight (g) of grains harvested from the five random panicles in a plot.

HGW is the weight (g) of 100 grain samples taken from the bulk seeds of five random panicles harvested in a plot.

ANT is the anthracnose disease severity scored using a standardized rating system on a scale of 1 to 5 indicating the level of anthracnose infection and disease severity. Accessions with leaves showing no sign of infection were assigned a score of 1. Score 2 was assigned to accessions showing hypersensitive reactions with local lesions but no visible acervuli. A score of 3 were assigned to accessions with infected bottom leaves with acervuli and a score of 4 was assigned to accessions where middle to bottom leaves were infected and acervuli were visible, while score 5 was given to accessions where infections reached up to the flag leaf and acervuli were observed Prom et al. (2009). All tagged plants were bagged after flowering to prevent bird attack in Puerto Rico field experiments (Figure 4.1).

Figure 4.1. Tags (orange) and sorghum heads covered with mesh bags at the Yela Farm**,** Puerto Rico field experiment.

The data were then subjected to a comprehensive statistical analysis using SAS software version 9.4 for Windows (SAS Institute Inc.) and R statistical software. Environments and replications were treated as random effects whereas the entry and its components (inbreds, hybrids, male, female, and male × female) were considered fixed. The interactions of entry and its components with the environment were treated as random. Estimates for GCA and SCA were calculated following the methods described in (Griffing, 1956; Hallauer and Miranda, 1988; Bernardo, 2020). The following model was used to calculate combining ability.

$Y_{ijk} = \mu + G_i + G_j + S_{ij} + eijk$, where:

Y_{ijk} is the observed trait value for the k^{th} hybrid resulting from the cross between the i^{th} and j^{th} parental lines, μ is the population mean, G_i is the general combining ability effect of the i^{th} parental line, G_j is the general combining ability effect of the j^{th} parental line, S_{ij} is the specific

combining ability effect between the cross of the i^{th} and the j^{th} parental lines, and *eijk is* the random error term associated with the observation.

Statistical analysis was performed on individual environment and combined dataset to assess the effects of the entry. Entry effects were further partitioned into components related to inbred and hybrid characteristics, as well as inbred versus hybrid distinctions. In the same way, the hybrid effect was disaggregated into male, female, and male × female interaction effects corresponding to the general combining ability (GCA) for males and females as well as the specific combining ability (SCA) effects. The difference between the grand mean and the marginal means of the lines was used to compute the GCA estimates of each parental line. A two-tailed t-test was used to determine the significance of the effects. Pearson correlation coefficients and graphs were generated using R statistical software using combined data from all environments.

4.3 Results

4.3.1 Analysis of variance

The combined analysis of variance for agronomic traits and yield components is presented in Table 4.1. The ANOVA shows that the environment effect was significant for all traits measured except the HGW. Similarly, the entry effect was significant for all traits studied while its inbred component was significant only for PHT, PKW, GYP and ANT. The hybrid component, on the other hand, was significant for all traits except DTF (Table 4.1). The inbred vs. hybrid effect was also significant for all traits except the ANT.

The male, female, and male vs. female components of the inbred were significant for all traits except DTF and HGW. The male, female and M x F components of the hybrid were also significant for all traits except the female effect was not significant for HGW.

The interaction between entry and its inbred and hybrid components was significant for all traits. The interaction of the inbred male with environment was also significant for all traits. However, the inbred female by environment interaction was significant only for PHT and ANT and the inbred male vs. female by environment interaction was significant only for DTF and PKW. On the other hand, the hybrid by environment interaction effect and the interaction between environment and all components of the hybrid were significant for all traits. The inbred vs. hybrid by environment interaction effect was also significant for all traits except PHT and PKW (Table 4.1).

4.3.2 Performance of genotypes

As presented in the ANOVA, the entry, inbred parent, and hybrid effects were significant for majority of the traits studied (Table 4.1). Days to flowering among inbred parents ranged from about 63 to 69 d with the mean of 66.5 d (Table 4.2). The range was from 63 to 67 d among the female inbreds and 65 to 69 d among the male inbreds. Similarly, there was significant variation for plant height with the shortest line R669 being 82 cm tall and the tallest line R604 measuring 106 cm. The female parents were in the range of 83 to 91 cm.

Yield components, panicle weight and panicle yield, were also significantly different among the inbreds with lines R245 being the top followed by R618 and R633 as having the second and third largest scores for both traits, in that order. Whereas lines B531 followed by B539 and B553 have the least scores (Table 4.2). For anthracnose disease rating, those lines show the reverse trend.

Table 4.1. Combined analysis of variance for agronomic parameters.

Source of variation	DF	Mean Squares					
		DTF	PTH	PKW	GYP	HGW	ANT
Environment (E)	4	3523.6**	25931.3**	6925.4**	1836.0**	0.9^{ns}	41.0**
Replication	2	33.5**	350.5**	86.5^{ns}	2520.7**	0.3^{ns}	1.3**
Entry	53	62.3**	1430.7**	2607.6**	1382.6**	1.1**	6.1**
Inbred (I)	13	75.4^{ns}	871.3**	4452.7**	2306.9**	0.6^{ns}	9.5**
Male (M)	9	62.6^{ns}	852.9**	2219.6**	1189.3**	0.5^{ns}	8.9**
Female (F)	3	36.7^{ns}	285.9^{ns}	199.7*	140.2*	0.1^{ns}	4.6**
M vs. F	1	317.4^{ns}	2789.2*	37291**	18735**	3.2^{ns}	29.6**
Hybrid	39	21.0^{ns}	900.7**	1818.0**	984.2**	1.2**	4.9**
Male	9	30.1**	2281.6**	4051.8**	2213.3**	2.3**	4.7*
Female	3	52.4**	1182.7**	4722.2**	1774.3**	3.5^{ns}	40.0**
M × F	27	14.0**	412.3**	699.5**	476.2**	0.5**	1.1**
I vs. H	1	1530.2*	29418**	9347.7*	4761.3*	2.3*	5.1^{ns}
Entry × E	212	22.4**	159.1**	226.2**	138.7**	0.4**	0.7**
Inbred × E	52	40.2**	93.5**	231.9**	145.8**	0.5**	0.7**
M × E	36	40.5**	79.5**	261**	169.0**	0.6**	0.7**
F × E	12	14.7^{ns}	120.6**	41.8^{ns}	34.9^{ns}	0.3^{ns}	0.6**
M vs. F × E	4	115.0**	138.5^{ns}	548.6*	260.6^{ns}	0.4^{ns}	0.2^{ns}
H × E	156	14.3**	176.2**	211.0**	126.5**	0.4**	0.7**
M × E	36	19.0**	430.0**	439.0**	231.9**	0.4**	0.8**
F × E	12	14.0**	187.0^{ns}	327.9**	199*	1.6**	3.4**
M × F × E	108	13**	90.4**	119.1**	84.2**	0.2**	0.3**
I vs. H × E	4	106.7**	347.9^{ns}	683.8^{ns}	490.6*	2.7**	3.7**
Error	530	4.82	22.1	57.7	44.4	0.2	0.1

DTF = Days to flowering, PTH = Plant height, PKW = Panicle weight, GYP = Grain yield per panicle, HGW = Hundred grain weight, ANT = Anthracnose score. *** significant at p<0.001, ** significant at p<0.01 and * significant at p<0.05.

Those R-lines that had the highest panicle weight and panicle yield had the lowest disease score and the B-lines with the least yield components were more susceptible to the disease and hence had among the highest anthracnose disease score. Those other lines with intermediate panicle yield and panicle weight also had intermediate disease scores. Lines R245 clearly stood

out as the most resistant line with disease rating of 1.73 followed by R604 and R633 with disease ratings of 2.11 and 2.14, respectively. This was consistent with visual field observation in Figure 2 obtained from artificially inoculated field in Tifton Georgia where R245 stood green in the middle of highly devastated field. Although R604 was resistant to the disease, its hybrid A531 x R604 appeared highly susceptible to the disease (Figure 2). Hundred grain weight, on the other hand, does not seem to have been affected by disease reaction of the lines. All the lines, regardless of their disease reaction, had above average grain weight with the overall mean of 3.4 g.

Table 4.2. Mean performance of the parents evaluated across environments for each trait.

Genotypes	DTF	PTH (cm)	PKW (g)	GYP (g)	HGW (g)	ANT (1-5)
B527	66.80	91.60	43.85	31.85	3.64	3.91
B531	64.47	91.03	33.97	23.83	3.57	4.37
B539	63.64	91.75	38.29	27.57	3.57	3.65
B553	63.21	82.76	37.20	29.01	3.70	3.04
R245	69.47	95.22	92.11	66.76	3.39	1.73
R588	67.07	91.70	51.06	35.40	3.63	3.63
R604	69.36	106.14	71.24	51.95	3.37	2.11
R611	64.93	105.47	56.21	39.53	3.01	3.65
R618	64.53	104.30	80.29	56.64	3.43	2.71
R633	65.07	97.71	74.27	53.61	3.19	2.14
R651	65.80	102.06	61.03	45.87	3.18	3.53
R653	69.27	93.85	63.75	44.75	3.15	3.47
R669	69.13	81.68	69.87	48.57	3.35	2.41
R672	68.67	95.38	60.65	46.33	3.54	3.75
Mean	**66.53**	**95.05**	**59.56**	**42.98**	**3.41**	**3.15**
CV	**9.3**	**14.2**	**19.4**	**23.3**	**16.5**	**18**
LSD	**4.5**	**9.7**	**14.8**	**7.2**	**0.4**	**0.4**

Figure 4.2. Field experiment at Tifton, Georgia. Resistant male parental line, R245 (middle) and susceptible hybrid, A531 × R604 (left).

Data shown in Tables 4.2 and 4.3 seem to be strongly supported by a photoshoot of the field in Tifton Georgia (Figure 4.2) demonstrating the variation in disease reaction of the entries. This could have been even more discernable had there been shots from aerial view. The pollinator line R45 that had a mean disease score of 1.73, stands almost clean amid other entries that appear to have suffered various degrees of foliar damage by the infection. There is evidence of premature death of plants for susceptible entries such as A531 x R604 hybrid (Figure 4.2). R604 has been among the least disease score as inbred per se but was not able to hold up the resistance when crossed to some of the susceptible females.

Table 4.3. Mean performance of the hybrids across environments for each trait

Genotypes	DTF	PTH (cm)	PKW (g)	GYP (g)	HGW (g)	ANT (1-5)
A527xR245	63.53	103.64	83.01	60.41	3.03	3.25
A527xR588	62.71	107.09	52.38	36.29	3.33	4.01
A527xR604	63.07	98.73	68.45	50.58	3.16	3.62
A527xR611	63.73	118.86	68.27	50.59	3.05	3.87
A527xR618	63.53	117.32	43.87	33.19	3.54	3.26
A527xR633	63.80	111.78	48.67	32.97	3.49	3.67
A527xR651	63.87	102.41	46.14	33.87	2.95	4.11
A527xR653	63.53	102.64	56.79	39.74	3.15	3.93
A527xR669	64.00	113.18	68.40	50.91	3.07	3.97
A527xR672	65.60	91.24	66.96	53.01	3.27	3.57
A531xR245	62.93	123.17	89.45	62.10	3.09	3.39
A531xR588	61.93	109.48	57.71	40.23	3.34	4.11
A531xR604	63.64	117.87	77.50	58.08	3.01	3.63
A531xR611	61.13	108.35	74.96	54.21	3.07	3.98
A531xR618	61.87	122.62	66.03	46.67	3.34	3.72
A531xR633	63.29	109.76	72.84	49.59	3.24	3.40
A531xR651	62.21	105.78	63.81	45.27	2.92	4.15
A531xR653	61.00	103.87	69.15	51.65	3.46	3.97
A531xR669	63.21	120.18	63.80	44.59	2.95	4.14
A531xR672	62.71	106.79	65.89	45.49	3.44	3.68
A539xR245	64.60	110.84	76.83	56.26	3.27	2.48
A539xR588	63.87	102.30	59.59	44.09	3.40	2.89
A539xR604	64.87	112.51	73.74	52.71	2.94	2.57
A539xR611	63.20	114.49	67.87	50.83	3.06	3.17
A539xR618	62.27	121.34	61.37	45.19	3.68	3.43
A539xR633	61.64	105.34	75.82	51.73	3.61	2.69
A539xR651	64.87	115.64	64.81	48.07	3.05	3.25
A539xR653	63.73	96.30	66.85	46.25	2.91	3.17
A539xR669	62.93	111.41	59.52	43.38	3.11	2.60
A539xR672	63.07	101.72	54.05	39.39	3.17	3.76
A553xR245	62.07	108.93	90.65	63.78	3.12	2.37
A553xR588	63.93	107.31	64.14	44.87	3.52	2.80
A553xR604	65.80	100.55	87.56	65.79	3.61	2.31
A553xR611	63.27	114.71	72.15	52.25	3.05	3.31
A553xR618	61.47	118.64	78.96	55.35	3.74	2.69
A553xR633	62.67	108.19	78.21	53.63	3.58	2.33
A553xR651	64.07	108.75	69.49	50.75	3.27	3.61
A553xR653	63.33	102.10	75.41	55.48	4.00	2.72
A553xR669	63.73	103.88	60.15	42.87	3.24	3.03
A553xR672	66.67	94.01	57.77			
Mean	**63.33**	**108.84**	**67.48**	**48.77**	**3.26**	**3.35**
CV	**8.0**	**13.3**	**17.9**	**19.7**	**13.8**	**23.2**
LSD	**3.7**	**10.4**	**8.7**	**6.9**	**0.3**	**0.6**

4.3.3 Correlation among agronomic traits and anthracnose

As one can see from the correlation coefficients, there was a high degree of association between two or more of the traits, some of which were highly significant. Plant height was significantly correlated with all yield components, panicle weight (r=0.33), panicle yield (r=0.31) and HGW

(r=-0.31). Anthracnose disease severity did not have any correlation with plant height. As expected, there was strong and positive correlation between panicle weight and panicle yield (r=0.99) but neither trait was correlated with HGW.

The correlation between anthracnose disease severity and other traits was significant only for panicle weight (r=-0.54) and panicle yield (r=-0.53) (Figure 4.3). This observation was already noted from Tables 4.2 and 4.3 where inbreds and hybrids with higher panicle weight and panicle yield clearly stood out as having the least disease score and vice versa.

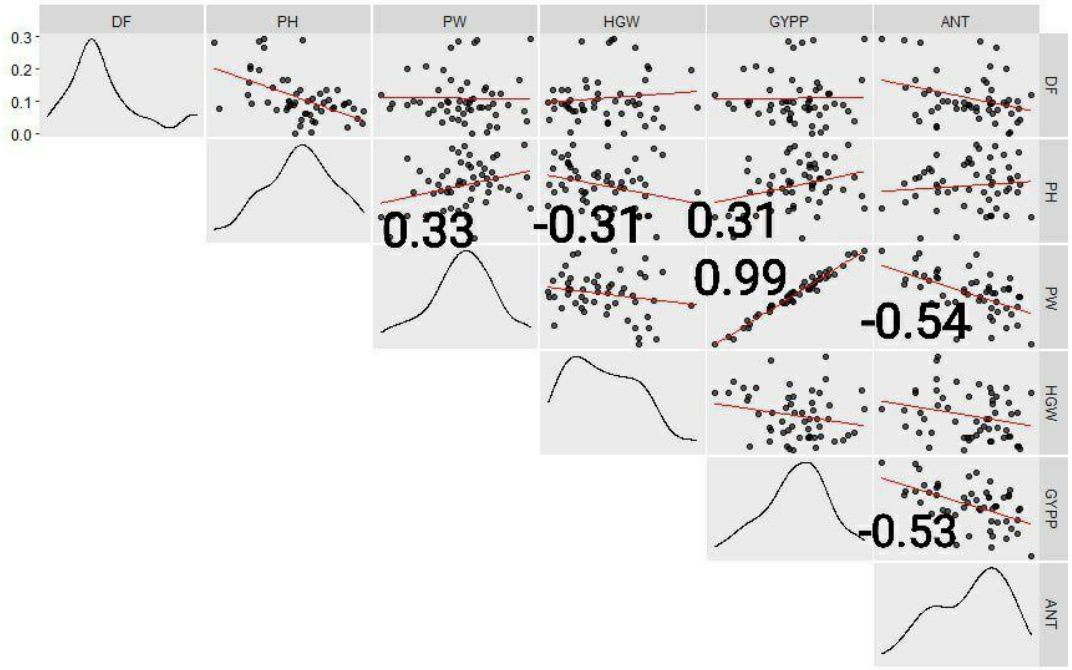


Figure 4.3. Correlation between traits.
The numbers shown are for significant correlations between the traits. Days to flowering (DTF), plant height (PTH), panicle weight (PKW), hundred-grain weight (HGW), grain yield per panicle (GYP), and anthracnose rating (ANT).

4.3.4 Combining ability analysis

Parental lines B531, B539, R588, R611, R618, R633, R653, R245 and R669 exhibited significant negative GCA effects for DTF. The highest negative GCA was for R618 (-1.88)

followed by A531 (-1.68). No significant positive GCAs were recorded for any of the genotypes for DTF. All the parents showed either a positive or negative significant GCA effect for PTH where significant negative GCA were observed for R653 and R672 while B527, B531, B539, B553, R245, R604, R611, R618, R633, R651, and R669 exhibited significant positive GCA effects.

The GCA effect of male and female parents were shown in. For PKW, parental lines B531, B553, R245, R604, R611 and R633, showed significant positive GCA effects with the highest being R245 (19.5) followed by R604 (11.33) and B553 (7.97). Parental lines B527, R588, R651, and R672 showed significant negative GCA. B527 and R588 had the highest negative GCA for PKW. For GYP, R245 recorded significant GCA with the highest effect of 13.51. Parental lines B531, B553, R604, and R611 had significant GCA followed by R245. On the contrary, B527, R588, R651, and R672 had significant negative GCA for GYP.

For HGW, R618 (0.26), B553 (0.19), R633 (0.17) and R672 (0.15) showed significant GCA while R651 (-0.26), R611 (-0.25), R245 (-0.18), R604 (-0.12), B531 (-0.12), and B527 (-0.1) recorded significant negative GCA. For anthracnose rating, parents A553 (-0.5), R245 (-0.41), A539 (-0.29), R633 (-0.26), and R604 (-0.25) exhibited significant negative GCA, indicative of the resistance of these lines to anthracnose. Whereas A531 (0.52), R651 (0.48), A527 (0.43), R611 (0.29), and R588 (0.16) exhibited significant positive GCA for anthracnose rating.

Table 4.4. Mean for male and female parents and general combining ability (GCA) estimates.

	DTF		PTH		PKW		GYP		HGW		ANT	
Genotype	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA	Mean	GCA
B527	63.76	-0.41*	106.69	1.43***	60.3	-5.18***	44.26	-2.85***	3.21	-0.1***	3.73	0.43***
B531	62.48	-1.68***	112.65	7.39***	70.18	4.7***	49.81	2.68***	3.19	-0.12***	3.82	0.52***
B539	63.55	-0.62***	109.19	3.94***	66.04	0.56	47.79	0.66	3.22	-0.09	3	-0.29***
B553	63.74	-0.43*	106.71	1.45***	73.45	7.97***	52.45	5.32***	3.51	0.19***	2.78	-0.5***
R245	63.44	-0.72**	111.65	6.39***	84.98	19.5***	60.64	13.51***	3.13	-0.18***	2.88	-0.41***
R588	63.15	-1.01***	106.54	1.29*	58.46	-7.02***	41.37	-5.75***	3.4	0.08	3.45	0.16***
R604	64.39	0.21	107.42	2.16***	76.81	11.33***	56.79	9.67***	3.18	-0.12**	3.03	-0.25***
R611	62.83	-1.33***	114.1	8.85***	70.81	5.33***	51.97	4.84***	3.06	-0.25***	3.58	0.29***
R618	62.28	-1.88***	119.98	14.72***	62.56	-2.92**	45.1	-2.01*	3.58	0.26***	3.27	-0.01
R633	63.02	-1.15***	108.77	3.51***	68.89	3.4***	46.98	-0.14	3.48	0.17***	3.02	-0.26***
R651	63.8	-0.37	108.15	2.89***	61.06	-4.41***	44.49	-2.63**	3.05	-0.26***	3.78	0.48***
R653	62.9	-1.27***	101.23	-4.02***	67.05	1.57	48.28	1.15	3.38	0.06	3.45	0.15**
R669	63.4	-0.77**	111.82	6.57***	63.14	-2.33*	45.48	-1.64	3.1	-0.21***	3.43	0.14**
R672	64.6	0.42	98.44	-6.81***	61.18	-4.29***	44.67	-2.44**	3.47	0.15**	3.43	0.13**

***significant at p<0.001, ** significant at p<0.01, * significant at p<0.05, DTF = Days to flowering, PTH = plant height, PKW = panicle weight, HGW = hundred-grain weight, GYP = grain yield per panicle, and ANT = Anthracnose rating.

Table 4.5. Specific Combining ability (SCA) estimates for hybrid combinations for the traits studied.

Female	Male	DTF	PTH	PKW	HGW	GYP	ANT
A527	R245	0.49	-9.45	3.20***	0.01**	2.63***	-0.06
A527	R588	0.13**	-0.9	-0.89***	0.03	-2.23***	0.12***
A527	R604	-0.90**	-10.12***	-3.18	0.08	-3.35*	0.15**
A527	R611	1.3	3.32***	2.64	0.09**	1.48*	-0.14***
A527	R618	1.65	-4.10***	-13.51***	0.07*	-9.05***	-0.45
A527	R633	1.19	1.57***	-15.03***	0.11	-11.15***	0.21***
A527	R651	0.48	-7.17*	-9.74***	0.00**	-7.77***	-0.11***
A527	R653	1.04	-0.03*	-5.08***	-0.13	-5.68***	0.04***
A527	R669	1.01	-0.08***	10.44	0.08*	8.28*	0.10***
A527	R672	1.42***	-8.64***	11.02	-0.05	12.29***	-0.29**
A531	R245	1.41	4.12***	-0.24***	0.09*	-1.22***	-0.01
A531	R588	0.45***	-4.47***	-5.45***	0.07	-3.82***	0.13***
A531	R604	1.12	3.06***	-4.02***	-0.05**	-1.40***	0.07***
A531	R611	-0.03***	-13.15*	-0.56***	0.13**	-0.45***	-0.13***
A531	R618	1.261***	-4.76***	-1.23	-0.11	-1.12	-0.08***
A531	R633	2.21	-6.41***	-0.75***	-0.12	-0.08	-0.15
A531	R651	0.26***	-9.76	-1.96	-0.01***	-1.9	-0.15***
A531	R653	-0.22***	-4.76	-2.6	0.2	0.68**	-0.01***
A531	R669	1.19**	-0.41***	-3.34	-0.01***	-3.4	0.18***
A531	R672	0.14**	0.95	0	0.09	-1.87	-0.27***
A539	R245	1.77	-4.75***	-8.72***	0.23	-5.05***	-0.11***
A539	R588	1.33	-8.18*	0.57**	0.09	2.05	-0.27***
A539	R604	1.11	1.15***	-3.64***	-0.15***	-4.75**	-0.18***
A539	R611	0.98*	-3.55***	-3.5	0.09**	-1.81*	-0.13
A539	R618	0.60***	-2.58***	-1.75**	0.20***	-0.58	0.44
A539	R633	-0.35***	-7.37	6.37***	0.22**	4.08**	-0.04***
A539	R651	1.69	3.56***	3.18	0.10**	2.91	-0.24
A539	R653	1.45	-8.86***	-0.76	-0.38***	-2.7	0.02
A539	R669	0.15**	-4.35***	-4.19**	0.10*	-2.77*	-0.54***
A539	R672	-0.90**	-0.66**	-7.70***	-0.21	-5.95***	0.63***
A553	R245	-0.57***	-4.17**	-2.30***	-0.21*	-2.18***	0.00***
A553	R588	1.2	-0.69	-2.29	-0.08*	-1.83	-0.14***
A553	R604	1.85***	-8.32***	2.78***	0.23**	3.68***	-0.21***
A553	R611	0.86*	-0.85***	-6.64***	-0.20**	-5.05**	0.23
A553	R618	-0.39***	-2.80***	8.43***	-0.03***	4.92***	-0.08***
A553	R633	0.08***	-2.03*	1.36***	-0.10**	1.32***	-0.19***
A553	R651	0.7	-0.85**	0.46*	0.03	0.93*	0.33**
A553	R653	0.86	-0.58*	0.39***	0.42***	1.88***	-0.22***
A553	R669	0.76	-9.4	-10.96**	-0.06	-7.94*	0.10**
A553	R672	2.51***	-5.89***	-11.38***	0.28***	-10.30***	-0.23***

***significant at p<0.001, ** significant at p<0.01, * significant at p<0.05, DTF = Days to flowering, PTH = plant height, PKW = panicle weight, HGW = hundred-grain weight, GYP = grain yield per panicle, and ANT = Anthracnose rating.

The SCA effect for the different crosses is presented in (Table 4.5). Significant positive SCA were observed for HGW in crosses A539 × R618, A553 × R653, A553 × R672, A527 × R245, A527 × R611, A527 × R651, A531 × R611, A539 × R611, A539 × R633, A539 × R651, and A553 × R604. On the contrary, crosses A531 × R651, A531 × R669, A539 × R604, A539 × R653, and A553 × R618, A531 × R604, A553 × R604, A553 × R611, and A553 × R633 had significant negative SCA for the trait.

Crosses A527 × R245, A527 × R672, A553 × R604, A553 × R618, A553 × R633, A553 × R653 A531 × R653 and A539 × R633 had a positive significant SCA for GYP. Negative significant SCA were observed for A527 × R588, A527 × R618, A527 × R633, A527 × R651, A527 × R653, A531 × R245, A531 × R588, A531 ×R604, A531 × R611, A539 × R245, A539 × R672, A553 × R245, A553 × R672, A539 × R604 and A553 × R611.

Eighteen crosses, A527 × R611, A527 × R651, A531 × R611, A5531 × R618, A531 × R651, A531 × R653, A531 × R672, A539 × R245, A539 × R588, A539 × R604, A539 × R633, A539 × R669, A553 × R588, A553 × R604, A553 × R618, A553 × R633, A553 × R653, A553 × R672, and A527 × R672 exhibited negative significant SCA ($p<0.001$) for anthracnose disease rating showing that the hybrid combinations had resistance reaction. Positive significant SCA ($p<0.001$) were observed for crosses A527 × R588, A527 × R633, A527 × R653, A527 × R669, A531 × R588, A531 × R604, A531 × R669, A539 × R672, A553 × R245, A527 × R640, A553 × R651, and A553 × R669.

4.4 Discussion

Sorghum (*Sorghum bicolor*) faces a persistent challenge from biotic and abiotic stressors. Sorghum anthracnose, a devastating fungal disease caused by *Colletotrichum sublineolum,* is a major pathogen causing a significant yield loss of up to 50% (Harris et al., 1964; Thakur and Mathur, 2000; Cota et al., 2017). The current study comprised a diverse set of sorghum genotypes involving four female parents, ten male parents and the hybrids resulting from intercrossing of these parents. The parents were selected based on visual observations of foliar disease reactions in Puerto Rico in the preceding seasons. The hybrids and their parents were evaluated across five distinct environments, including three locations known to be hotspot areas for anthracnose and two others.

The ANOVA run on data collected from these environments revealed significant effects of genotype, environment, and their interaction for all traits studied. The results agree with similar findings reported earlier (Kenga et al., 2004; Mateo, 2007; Mengistu et al., 2010; and Kanbar et al., 2011). Given the differences in pedigree, yield potential and adaptation among the parents and hybrids, the results thus observed are expected. The significant entry effect and its components for yield related traits and the significant interaction of these components with environment highlights the impact of environment on overall crop performance and specific agronomic components. Moreover, analysis of the disease data corroborates the visual observation noted in the parental lines and the differences in disease environment.

The performance of the parental lines and their hybrids for yield components and reaction to anthracnose were variable. Lines R245, R604, R633, R618 and R669 had resistance reaction to foliar disease with disease ratings ranging from 1.73 in R245 to 2.71 in R618. Among the females, B527 and B531 had susceptible reaction to the disease while the other two females, B539 and

B553, were moderately resistant. This agrees with multiple years of observation we made in tropical winter nursery. The lines had similar grain weight with other lines but had significantly higher panicle weight and panicle yield. It is not clear if the higher yield components were functionally associated with improved disease resistance in these lines. But given that the potential number of grains per panicle were set long before the anthracnose infection occurred, (Roozeboom and Prasad, 2019), and that grain weight was not affected casts doubt on whether the higher yield components in the resistant lines has anything to do with the resistance reaction. These lines had larger stature (taller height or higher biomass) and/or longer maturity that are perhaps the main drivers for higher yield components.

The resistant R-lines when crossed to both susceptible and resistant females produced hybrids that are again higher in yield components than the rest of the lines. The overall mean panicle weight and yield per panicle of the entire hybrids was 67.48 and 48.77 g, respectively. Hybrids involving the resistant line R245 had panicle weight and yield per panicle ranging from 76.8 to 90.65 g and 56.26 to 63.78 g, respectively. However, the mean grain weight for these hybrids was not any different from the overall mean of the hybrids. The disease reaction of the hybrids, on the other hand, was significantly different. The reaction of hybrids involving resistant lines was not consistent across all females. The hybrids of the resistant lines with susceptible females (A527 and A531) were as susceptible as any other hybrids but these lines produced resistant hybrids when crossed with the moderately resistant females (A539 and A553).

For instance, R604 had resistant reaction similar to R245 when evaluated as inbred per se but it produced susceptible hybrid with A531. The field performance of the hybrid A531 x R604 was captured in Figure 4.2 where it was growing next to R245 showing a great contrast in performance with respect to disease reaction. All other resistant lines produced the same

susceptible hybrids when crossed to susceptible females, but they all produced fairly resistant hybrids when crossed with the moderately resistant females. All the susceptible R lines, however, produced susceptible hybrids with all females. Female A553 produced resistant hybrids (more resistant than the B553 per se) when crossed with all R-lines (Table 4.3) whereas female A539 produced resistant hybrids only with the resistant males. This shows that the resistant reaction is occurring at multiple levels with different genes acting together or independently. Such complexity may have to do with random nature of the pathogens and differential reaction of genotypes under different environments. Both additive and dominant genes at multiple loci may be responsible for the observed reactions while the likelihood of epistatic interaction contributing to resistance is also high. The result generally shows that there is a significant genetic component to reaction of sorghum genotypes to the disease; and a more controlled study that allows monitoring of the pathogen strain could provide better understanding of the genetic effect.

Nevertheless, the variable disease reaction of the resistant R-lines when evaluated in hybrid combination with resistant and susceptible females but a higher yield component regardless of disease reaction confirms that the higher yield components in the resistant hybrids and lines had less to do with disease reaction. This, however, doesn't mean that the foliar disease does not have impact on grain yield, but in this particular experiment, disease severity did not seem to have affected yield components except perhaps in artificially inoculated plots where the extent of foliar damage was remarkable (Figure 4.2). The significant association between disease reaction and yield components does not reflect the cause-and-effect relationship.

Individual lines have different GCA for the different traits considered. Almost all lines except R618 have significant GCA for resistance to anthracnose. Lines A527 and A531 had significant positive GCA indicating susceptibility, and this was true from the performance of

individual hybrids involving these females. Whereas both A539 and A553 had significant negative GCA indicating resistance to the disease. Among the R-lines, R245 had the highest negative GCA followed by R633 and R604. Line R651 had the highest GCA for susceptibility. The highest significant GCA for yield per panicle was observed for R245 followed by R604 and the least in R588. Among the females the highest positive GCA was noted in A533 followed by A531 with A527 being the least. The yield per panicle followed similar trend. Although their reaction in different crosses differ, lines that combine high GCA for the desired traits (yield components and disease reaction) have potential for use as breeding materials or directly as parents for producing resistant hybrids. Accordingly, lines R245, R604 and A553 stood out as the most desirable genetic materials for sorghum improvement.

4.5 Conclusion

This study evaluated the agronomic performance and disease reaction of a set of hybrid combinations with the aim of determining the potential of the lines as breeding materials. The result showed that a number of lines included in the cross have superior agronomic characteristics that can be utilized in breeding programs to enhance agronomic traits. Many of the lines have been previously evaluated for hybrid performance, and the current study confirmed this potential, and those lines can be directly used for commercial hybrid development. Those same elite lines also had remarkable disease resistance. In cross combination with parents with moderate resistance, the lines can produce hybrids with acceptable yield potential that also has the desired disease reaction. Because the disease rating was performed under natural infection condition, further study may be needed to understand which pathogen species and strains these lines react to and in what way. But the current study makes practical sense in that it tested the performance of hybrids under natural infection where multitudes of pathogens act at the same time. In conclusion, sources such R245, R604 and A/B553 hold great promise for developing hybrids with improved resistance to foliar diseases.

References

Abreha, K. B., Ortiz, R., Carlsson, A. S., & Geleta, M. (2021). Understanding the sorghum–*Colletotrichum sublineola* interactions for enhanced host resistance. *Frontiers in Plant Science*, 12, 641969. Frontiers Media S.A. https://doi.org/10.3389/fpls.2021.641969

Ananda, G. K. S., Myrans, H., Norton, S. L., Gleadow, R., Furtado, A., & Henry, R. J. (2020). Wild Sorghum as a Promising Resource for Crop Improvement. *Frontiers in Plant Science*, 11, Frontiers Media S.A. https://doi.org/10.3389/fpls.2020.01108

Aruna, C., & Cheruku, D. (2018). Genetic improvement of grain sorghum. In *breeding sorghum for diverse end uses* (157–173). Elsevier. https://doi.org/10.1016/B978-0-08-101879-8.00010-3

Bernardo, R. (2020). *Breeding for quantitative traits in plants* (3rd ed.). Stemma Press.

Chala, A., Tronsmo, A. M., & Brurberg, M. B. (2011). Genetic differentiation and gene flow in *Colletotrichum sublineolum* in Ethiopia, the center of origin and diversity of sorghum, as revealed by AFLP analysis. *Plant Pathology*, *60*(3), 474–482. https://doi.org/10.1111/j.1365-3059.2010.02389.x

Cota, L. V., Souza, A. G. C., Costa, R. V., Silva, D. D., Lanza, F. E., Aguiar, F. M., & Figueiredo, J. E. F. (2017a). Quantification of yield losses caused by leaf anthracnose on sorghum in Brazil. *Journal of Phytopathology*, *165*(7–8). https://doi.org/10.1111/jph.12582

De Wet, J. M. J. (1978). Systematics and evolution of sorghum Sect. Sorghum (Gramineae). *American Journal of Botany*, 65(4), 477-484. https://www.jstor.org/stable/2442706

FAO. (2021). *Available at: https://www.fao.org/faostat/en/#data/QCL (Accessed August 26, 2023).*

Griffing, B. R. U. C. E. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Australian journal of biological sciences*, *9*(4), 463-493.

Hallauer, A. R., & Miranda, J. D. (1988). *Quantitative Genetics in Maize Breeding 2nd edn. By A. R. Hallauer & J. B. Miranda Filho. xii468 pages. Ames Iowa State University Press. 1988.* Iowa State University Press.

Hao, H., Li, Z., Leng, C., Lu, C., Luo, H., Liu, Y., Wu, X., Liu, Z., Shang, L., & Jing, H. C. (2021). Sorghum breeding in the genomic era: opportunities and challenges. *Theoretical and Applied Genetics*, *134*(7), 1899–1924. https://doi.org/10.1007/s00122-021-03789-z

Hess, D. E., Bandyopadhyay, R., & Sissoko, I. (2002). Pattern analysis of sorghum genotype× environment interaction for leaf, panicle, and grain anthracnose in Mali. *Plant Disease*, *86*(12), 1374-1382.

Huang, R. D. (2018). Research progress on plant tolerance to soil salinity and alkalinity in sorghum. *Journal of Integrative Agriculture*, 17(4), 739–746. Chinese Academy of Agricultural Sciences. https://doi.org/10.1016/S2095-3119(17)61728-3

Jane Da Rocha, M., Airton, J., Nunes, R., Augusto Da Costa Parrella, R., Sherma Da, P., Leite, S., Maria, G., Lombardi, R., Luiza, M., Moura, C., Schaffert, R. E., & Bruzi, A. T. (2018). General and specific combining ability in sweet sorghum. *Crop Breeding and Applied Biotechnology*, *18*, 365–372. http://dx.doi.org/10.1590/1984-70332018v18n4a55

Kanbar, O., Kanbar, A., & Shehab, S. (2011). Combining ability and heterosis for some yield traits in sorghum (*Sorghum bicolor* [L.] Moench) using (line × tester) design. *Journal of Plant Production*, *2*(8), 1009–1016. https://doi.org/10.21608/jpp.2011.85633

Kenga, R., Alabi, S. O., & Gupta, S. C. (2004). Combining ability studies in tropical sorghum (Sorghum bicolor (L.) Moench). *Field Crops Research*, *88*(2–3), 251–260. https://doi.org/10.1016/j.fcr.2004.01.002

Mateo Moncada, R. A. (2007). Genetic diversity and combining ability among sorghum conversion lines. (Doctoral dissertation, Texas A&M University).

Mengistu, G., Ayana, A., & Belete, K. (2010). Combining Ability for Yield and its Components in Ethiopian Sorghum (Sorghum bicolor [L.] Moench) Landraces. *East African Journal of Sciences*, *4*(1), 34–40.

Mengistu, G., Shimelis, H., Laing, M., Lule, D., & Mashilo, J. (2020). Combining ability and heterosis among sorghum (*Sorghum bicolor* [L.] Moench) lines for yield, yield-related traits, and anthracnose resistance in western Ethiopia. *Euphytica*, *216*(2). https://doi.org/10.1007/s10681-020-2563-6

Mofokeng, M. A., Shimelis, H., Laing, M., & Shargie, N. (2017). 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*(9), 1078–1085. https://doi.org/10.21475/ajcs.17.11.09.pne347

Prom, L. K., Perumal, R., Erpelding, J., Isakeit, T., Montes-Garcia, N., & Magill, C. W. (2009). A pictorial technique for mass screening of sorghum germplasm for anthracnose (*Colletotrichum sublineolum*) resistance. *The Open Agriculture Journal*, *3*(1), 20–25. https://doi.org/10.2174/1874331500903010020

Roozeboom, K. L., & Prasad, P. V. V. (2019). Sorghum growth and development. In *Sorghum: State of the Art and Future Perspectives* (pp. 155–172). Wiley. https://doi.org/10.2134/agronmonogr58.2014.0062

Stefoska-Needham, A., Beck, E. J., Johnson, S. K., & Tapsell, L. C. (2015). Sorghum: an underutilized cereal whole grain with the potential to assist in the prevention of chronic disease. *Food Reviews International*, *31*(4), 401-437. http://ro.uow.edu.au/smhpapers/3019

Taylor, J. R. N., Schober, T. J., & Bean, S. R. (2006). Novel food and non-food uses for sorghum and millets. *Journal of Cereal Science*, 44(3), 252–271. https://doi.org/10.1016/j.jcs.2006.06.009

Tesso, T., Perumal, R., Little, C. R., Adeyanju, A., Radwan, G. L., Prom, L. K., & Magill, C. W. (2012). Sorghum pathology and biotechnology-A fungal disease perspective: Part II. Anthracnose, stalk rot, and downy mildew. *The European Journal of Plant Science and Biotechnology*, *6*(Special Issue 1), 31–44.

Thomas, M. D., Sissoko, I., & Sacko, M. (1996). Development of leaf anthracnose and its effect on yield and grain weight of sorghum in West Africa. *Plant Disease*, *80*, 151–153.

Upadhyaya, H. D., Vetriventhan, M., Asiri, A. M., Azevedo, V. C. R., Sharma, H. C., Sharma, R., Sharma, S. P., & Wang, Y. H. (2019). Multi-trait diverse germplasm sources from mini core collection for sorghum improvement. *Agriculture (Switzerland)*, *9*(6). https://doi.org/10.3390/agriculture9060121

Zhang, R., Zhou, Y., Yue, Z., Chen, X., Cao, X., Ai, X., Jiang, B., & Xing, Y. (2019). The leaf-air temperature difference reflects the variation in water status and photosynthesis of sorghum under waterlogged conditions. *PLoS ONE*, *14*(7). https://doi.org/10.1371/journal.pone.0219209

Appendix A Reactions of some genotypes for anthracnose disease.