A GENOME-WIDE ASSOCIATION STUDY OF THE QUANTITATIVE RESISTANCE TO *STRIGA HERMONTHICA* AND PLANT ARCHITECTURE OF *SORGHUM BICOLOR* IN NORTHWESTERN ETHIOPIA

by

Megan Khangura

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agronomy

West Lafayette, Indiana

December 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Gebisa Ejeta, Chair

Department of Agronomy, Purdue University

Dr. Clifford Weil

Department of Agronomy, Purdue University

Dr. Tesfaye Mengiste

Department of Botany and Plant Pathology, Purdue University

Dr. Rebecca W. Doerge

Department of Biological Sciences, Department of Statistics and Data Science, Carnegie Mellon University

Dr. Alex Lipka

Department of Crop Science, University of Illinois Urbana-Champaign

Approved by:

Dr. Ronald Turco

Head of the Graduate Program

With a heavy heart, this dissertation is dedicated to Dr. Sharon Gray for her devotion to bettering the world at home and around the world. Dr. Gray, please rest peacefully knowing you will stay forever young in our minds. Our memory of you serves as a steadfast reminder of the fragility of life as we follow in your footsteps, embarking on international research in hopes of leaving the world a little better than we have found it. The vigilance I acquired after you lost your life, saved mine, and for that, I am forever grateful.

Dr. Sharon Gray

November 25th 1985 - October 4th 2016

ACKNOWLEDGMENTS

When spiderwebs unite, they can tie up a lion.
-Ethiopian Proverb

After spending a year and a half in India studying agronomy with a focus on soil science, I decided that I needed to change my career path. For a number of reasons, I have concluded that if I wanted to be part of the solution in feeding the hungry, I would need to change my career focus to plant breeding and genetics. It is one thing to have a goal that you work at every day, but it is another entirely to have a person share the belief. My advisor Dr. Ejeta shares the belief that we could achieve global food security, and that plant breeding was a part of the solution

Dr. Ejeta has provided endless support in facilitating an environment where I could acquire the transdisciplinary skills that are needed to address global hunger. Dr. Ejeta supported me while I lived in Ethiopia for two full growing seasons where I not only did my fieldwork but also got an gained an understanding of a wonderful culture that intend on spending my life to serve. I would like to sincerely thank you Dr. Ejeta for having faith in me, facilitating my growth as a scientist and teaching me the value of diligence, kindness and respect.

The monumental international research effort that encompassed my Ph.D. research was not accomplished in isolation. Dr. Patrick Rich, thank you for always having an open door, pushing me to improve my writing and always being an advocate when I needed it most. Dr. Rebecca Doerge, you have been with me since week one at Purdue University. Thank you for always telling me what I needed to hear sprinkled with a compassion and kindness. I am so lucky to have a female mentor like you in my life. To the members of my committee, Dr. Tesfaye Mengiste, Dr. Clifford Weil and Dr. Alex Lipka, thank you all for your time, dedication and steadfast support. Thank you to Mr. Terry Lemming and all of the Ejeta laboratory members, past and present for their hard work in the field.

Dearest brother, Matthew Fenton, thank you for helping me pack seed and do any other crazy task I asked all weekend long to help me with my research in exchange for a burger. You make me very proud, and I am so excited to have my brother as a partner in the field of plant breeding and genetics. Of course, I must thank my family, supporting me all of these years, being stressed at home waiting for the phone to ring just to know I am fine. To my mother, father and extended family thank you for everything and I am sorry for the sleepless night. There is one man who deserves a true medal for the endless support he provided during the course of my graduate work. Dr. Rajdeep Singh Khangura, I could not ask for a better partner in life and science. I am blessed to have you as a husband. I am so excited to see what we will accomplish together.

Above all I must thank the group of more than thirty women and a few young men in Ethiopia that toiled with me in the hot sun for two years. Gebremedhn Gebregergs Abay you and Kiflom are the hardest working individuals I have ever met and this research would have never been completed without you both. Gebremedhn your contributions have not been forgotten and I am forever grateful. Selam, Meeri and Turye worked with me for two years and have decided to pursue a career in agricultural science, I wish you all the best, and give you my endless support. To all the researchers, staff and daily laborers of the Humera Agricultural Research Center, thank you so much for graciously welcoming me into your world and tirelessly supporting this work. You all were so influential in shaping my career goals and ethos moving forward in international development. Thank you to the Bill and Melinda Gates Foundation for funding this research.

TABLE OF CONTENTS

LIST OF TABLES 7
LIST OF FIGURES 8
ABSTRACT 9
CHAPTER 1. A REVIEW OF *STRIGA HERMONTHICA* PARASITISM IN SORGHUM 11
1.1 Introduction 11
1.2 The parasitic life cycle *S. hermonthica* 13
1.2.1 Seed dormancy and conditioning 14
1.2.2 Strigolactone induced germination and perception 14
1.2.3 Haustorial formation and attachment 16
1.2.4 Growth and maturation 17
1.3 Quantitative resistance *S. hermonthica* in sorghum 18
1.3.1 Indirect analysis of *S. hermonthica* resistance using laboratory assays 18
1.3.2 Direct analysis of *S. hermonthica* resistance in infested fields 19
1.4 Future research for understanding quantitative resistance to *S. hermonthica* 20
1.4.1 Linkage mapping 20
1.4.2 Association mapping 22
1.4.3 Association mapping in MAGIC populations 23
1.4.4 Mapping for *Striga hermonthica* Resistance in Sorghum 23
CHAPTER 2. INVESTIGATING A CEREAL KILLER: EXPLORING QUANTITATIVE RESISTANCE TO THE PARASITE *STRIGA HERMONTHICA* IN SORGHUM 28
2.1 Abstract 28
2.2 Introduction 28
2.3 Materials and Methods 32
2.3.1 Germplasm 32
2.3.2 Genotypic data 32
2.3.3 Genotyping for *LGS1* 33
2.3.4 *Striga hermonthica* field resistance phenotypes 34
2.3.5 *Striga hermonthica* germination stimulant activity 36
2.3.6 Association analysis 36
2.4 Results 37

2.4.1 Structure of the PP37 population 37
2.4.2 Genome-wide association of *S. hermonthica* germination stimulant activity 38
2.4.3 Genome-wide association of *S. hermonthica* field resistance 39
2.5 Discussion 40
2.5.1 An improved understanding of LGS1 and mapping structural variants 40
2.5.2 The challenge remains phenotyping *Striga hermonthica* resistance 41
CHAPTER 3. EXPLORING THE GENETIC ARCHITECTURE OF PLANT ARCHITECTURE IN SORGHUM 58
3.1 Abstract 58
3.2 Introduction 59
3.3 Materials and Methods 60
3.3.1 Germplasm 60
3.3.2 Genotypic data 61
3.3.3 Field experimental design 62
3.3.4 Phenotyping for plant architecture 62
3.3.5 Population structure and association analysis 63
3.4 Results and Discussion 64
3.4.1 Structure of the PP37 population 64
3.4.2 Exertion is not associated with *dw3* 64
3.4.3 Novel locus associated with exertion and panicle length 65
CHAPTER 4. ACHIEVING GLOBAL FOOD SECURITY: THE BROADER IMPACTS OF PLANT SCIENCE RESEARCH 75
4.1 Abstract 75
4.2 Introduction 75
4.3 What are the broader impacts? 76
4.4 Translational research 76
4.5 Germplasm collections 78
4.6 Building human capacity in plant breeding and genetics 79
4.7 Responsibly implementing international research for sustainable impact 81
REFRENCES 86

LIST OF TABLES

Table 1.1: The previously detected QTL conferring field resistance to *S. hermonthica.* 26
Table 2.1: The founding accessions of the PP37 MAGIC population..... 45
Table 2.2: The primers used in the LGS1 marker screening of the PP37 MAGIC population. ... 46
Table 2.3: The phenotypic traits collected in 2016 and 2017. 46
Table 2.4: Candidate gene list for ST-SH3.1 47
Table 3.1: Candidate genes for qHT7.1 associated with all plant architecture phenotypes. 67
Table 3.2: Candidate genes for association novel association on chromosome 1 68
Table 4:1 The Sustainable Development Goals, developed by the United Nations. 85

LIST OF FIGURES

Figure 1.1: The parasite *in situ*. 27
Figure 2.1: *Striga hermonthica* attachment to the host root 48
Figure 2.2: The effect of *S. hermonthica* on BTx623. 49
Figure 2.3: Population structure of the PP37 MAGIC founders. 50
Figure 2.4: Population structure of the PP37 MAGIC population. 51
Figure 2.5: Germination stimulant activity of the PP37 MAGIC population. 52
Figure 2.6: A Manhattan plot of the MGD score in a subset of the PP37 MAGIC 53
Figure 2.7: Allelic distribution of *LGS1* in the PP37 MAGIC population. 54
Figure 2.8: The effect of *LGS1* on the *S. hermonthica* count. 55
Figure 2.9: Detecting *LGS1* in the GWAS 56
Figure 2.10: A novel GWAS peak for *S. hermonthica* resistance 57
Figure 3.1: Phylogenetic tree of the PP37 MAGIC population founders. 69
Figure 3.2: Phylogenetic tree of the PP37 MAGIC population 70
Figure 3.3: Manhattan plots for plant architecture on chromosome 7. 71
Figure 3.4: A syntenic dotplot between maize and the sorghum. 72
Figure 3.5: A schematic of the syntenic region in sorghum and maize influencing plant height. 73
Figure 3.6: Manhattan plots for significant associations with plant architecture. 74
Figure 4.1: The number of malnourished people in the world is increasing. 84

ABSTRACT

Sorghum (*Sorghum bicolor)* is a well-known agronomic crop of global importance. The demand for sorghum as a food crop makes it the fifth most important cereal in the world. The grain of sorghum is utilized for food and feed, whereas the sorghum biomass may have many other uses such as for fodder, bioenergy or even for construction. Globally, sorghum is consumed as a food crop and used for home construction primarily in the developing world. The grain and biomass yield of sorghum is drastically reduced by the parasitic plant *Striga hermonthica* which is endemic to Sub-Saharan Africa. To date, only one sorghum gene, *LGS1*, has been characterized as a genetic mechanism that reduces *S. hermonthica* parasitism by altering the strigolactone composition of the host root exudates which results in a reduction of the parasites ability to germinate. To establish more durable resistance additional genetic variation needs to be identified that reduces the *S. hermonthica* parasitism in sorghum, but also reduces the parasitic weed seed bank by promoting suicidal germination. To that end, the PP37 multi-parent advanced generation inter-cross (MAGIC) population was developed, originally as a recurrent selection population that was developed to recombine sorghum accessions with different putative resistance mechanisms to *S. hermonthica.* Whole genome sequences were developed for approximately 1,006 individuals of the PP37 MAGIC population. The population was phenotyped for *S. hermonthica* resistance during the 2016 and 2017 growing season in Northwestern Ethiopia. There was significant spatial variation in the *S. hermonthica* natural infestations that were partially attenuated for with artificial inoculation. The data was used to conduct a genome-wide association study that detected several subthreshold peaks, including the previously mapped *LGS1*. The highly quantitative nature of *S. hermonthica* resistance confounded with the complex spatial variation in the parasite infestations across a given location make it difficult to detect highly heritable variation across years and environments.

In addition to *S. hermonthica* resistance, the plant architecture of the PP37 MAGIC was also assessed at a location in Northwestern Ethiopia that is free of the parasite, as it significantly reduces plant height. To asses plant architecture the total plant height, the height of the panicle base, flag leaf height, and pre-flag leaf height were collected using a relatively high-throughput barcoded measurement system. Sorghum head exertion and panicle length were derived from this data. The actual measures of plant architecture and the derived traits were used to conduct a genome-wide association study. The high heritability of this trait demonstrated the statistical

power of the PP37 mapping population. Highly significant peaks were detected that resolved the *dwarf3* locus and an uncharacterized qHT7.1 that had only been previously resolved using a recombinant inbred line population. Furthermore, a novel significant locus was associated with exertion on chromosome 1. The random mating that was utilized to develop the PP37 MAGIC has broken the population structure that when present can hinder our ability associate regions of the genome to a given phenotype. As a result, novel candidate gene lists have been developed as an outcome of this research that refined the potential genes that need to be explored to validate qHT7.1 and the novel association on chromosome 1.

This research demonstrated the power of MAGIC populations in determining the genomic regions that influence complex phenotypes, that facilitates future work in sorghum genetic improvement through plant breeding. This research however also demonstrates a large international research effort. The nuisances and lessons learned while conducting this international research project are also discussed to help facilitate and guide similar research projects in the future. The broader impacts of this research on the society at large are also discussed, to highlight the unique potential broader impacts of international research in the plant sciences. The broader impacts of this research include germplasm development and extensive human capacity building in plant breeding genetics for developing country students and aspiring scientists. Overall this research attempts to serve as a model for highlighting the interdisciplinary nature and complexity of conducting international plant science research, while also making significant strides in improving our understanding the genetic architecture of quantitative traits of agronomic importance in sorghum.

CHAPTER 1. A REVIEW OF *STRIGA HERMONTHICA* PARASITISM IN SORGHUM

1.1 Introduction

Parasitic plants rely on the host plant to complete their lifecycle as they can feed directly from the host plant (1). Approximately one percent of all angiosperms utilize parasitism in their lifecycle (Westwood, Yoder, Timko, & dePamphilis, 2010). Holoparasitic plants have no photosynthetic activity and obtain all of their reduced carbon from the host through the haustorial connection, where hemiparasitic plants undergo varying degrees of photosynthetic activity (2). Perhaps the most notable members of the family parasitic Orobanchaceae are the *Striga* species as they parasitize essential crop species resulting in a significant negative impact on crop productivity worldwide (3). The value of the estimated annual yield loss from *Striga* worldwide is approximately $7 billion USD (3,4). *Striga* species have specific host preferences, and *Striga hermonthica* preferentially parasitizes host plants in the Poaceae family, notably maize (*Zea mays*), rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*) (5).

Striga hermonthica is a hemiparasitic plant that attacks sorghum and can result in a 40 to 100% yield reduction (Gebisa Ejeta, 2007). All parasitic plants connect to the host using an organ known as the haustorium, which in the case of *S. hermonthica* allows for a xylem-to-xylem connection between the two organisms (1). *S. hermonthica* parasitism results in significant yield reduction because it extracts water and nutrients from the host and thus results in chlorosis, wilting and even necrosis of the host (Figure 1.1). In the developing countries where *S. hermonthica* is prevalent, there are not established systems to produce herbicides to control the pest. Furthermore, *S. hermonthica* inflicts the majority of the damage on the plant before emerging from the ground, rendering post-emergent herbicides as ineffective. Therefore, the ideal method of control is to develop sorghum that has genetic resistance to *S. hermonthica* parasitism.

Sorghum displays natural genetic variation for quantitative resistance and tolerance to *S. hermonthica*. Quantitative resistance refers to a plant having a reduction in an infection that is conferred by multiple genes across the genome (6). Tolerance is a degree of susceptibility in which tolerant plants exhibit less reduction in yield than other susceptible plants with similar levels of infection (7). As early as 1936 researchers reported on the sorghum cultivars Dobbs and Kano, noting that even under *S. hermonthica* parasitism these cultivars were tolerant enough to produce

grain unlike the more susceptible landraces (8). Further analysis on the sorghum cultivar Dobbs corroborated this observation by demonstrating that this genotype produced a higher yield under *S. hermonthica* parasitism, but also that it supported a fewer number of attached parasitic plants than the susceptible check indicating resistance (9). Due to the severity of the *S. hermonthica* problem international efforts were made to screen sorghum germplasm collections for additional resistant cultivars (10). This research documents a significant hurdle in screening for *S. hermonthica* resistance, namely the lack of uniformity of natural parasite infestations across testing locations. Like many weed species the spatial distribution of *S. hermonthica* across testing locations, including inoculated sick-plots is not uniform (10–13). Screening the natural variation of sorghum germplasm for *S. hermonthica* field resistance can be cumbersome and requires extensive replication. To overcome the constraints presented by field variation on *S. hermonthica*, experimental designs are implemented that utilize dense placement of susceptible checks to quantify the variation of the parasite infestations in the field (Gilliver et al. 1986). These limitations have substantially hindered the ability to screen and discern levels of parasitism and thus have limited the understanding of quantitative resistance and tolerance to *S. hermonthica* in sorghum.

However, other unique strategies have been employed to overcome the limitations of phenotyping sorghum for *S. hermonthica* resistance and tolerance. The majority of our current understanding of quantitative resistance to *S. hermonthica* in sorghum was greatly facilitated by a research program at Purdue University which sought to deconstruct the life-cycle and interaction between the parasite and host into key events that could be analyzed in detail with individual laboratory assays (4,14,15). This strategy allowed for the decomposition of the multiple putative quantitative resistance mechanisms into qualitative assays that permitted efficient phenotyping for *Striga* parasitism. The ability to map these resistance mechanisms was improved further by advances in molecular markers and genome sequencing technology (16). The genomics era has brought a new level of hope to unraveling the genetics of *S. hermonthica* resistance in sorghum.

The intricate relationship between sorghum and *S. hermonthica*m will be reviewed here with significant detail given to the novel advancements made in the last decade that have permitted a greater understanding of the relationship. Even though, this improved understanding of the parasite-host interaction and sequencing technology facilitated the detection of the one and only known genetic resistance mechanism in sorghum, *LOW GERMINATION STIMULANT 1* (*LGS1*).

However, our understanding of the genetic architecture of quantitative resistance to *S. hermonthica* in sorghum remains limited. To date, the full potential of genome-wide sequencing tools and analysis have yet to be realized. With high-density genome coverage explaining the genetic variation, the main limitation remains the ability to phenotype accurately. In what follows is a demonstration of how the ability to phenotype can be improved by having a clear understanding of the host-parasite interaction, careful experimental design and utilizing appropriate populations when implementing genome-wide assessments for *S. hermonthica* resistance in sorghum.

1.2 The parasitic life cycle *S. hermonthica*

In Latin the term *Striga* translates to mean a witch or evil spirit. In a seemingly cursed manner, the host plants become wilted, chlorotic and stunted when parasitized by *Striga*, which is commonly known as witchweed. The tropical savannah spanning from the northwestern corner of Ethiopia to southern Sudan is home to the greatest biodiversity of *Striga.* This area is endemic with *S. hermonthica*, the *Striga* species with the largest impact across sub-Saharan Africa. *S. hermonthica* parasitizes plant species from the Poaceae family, predominantly staple food crops such as sorghum, maize, and rice. Infection of these crops by *S. hermonthica* can result in up to a total yield loss where infestations are high (17). Specifically, *S. hermonthica* is an obligate root hemi-parasite. The parasitic eudicot *S. hermonthica* is a member of the Orobanchaceae of the order Lamiales and only infects members of the Poaceae family (18). The whole genome sequence of *S. hermonthica* remains publicly unavailable to date. However, limited information is available on the *S. hermonthica* genome. Flow cytometry revealed that the *S. hermonthica* is a diploid (n=19) that has an estimated genome size of 1,801 Mbp (19). An expressed sequence tag (EST) analysis identified 17,137 putative genes of which 79% are homologs to genes of known function in model plant species (19). *S. hermonthica* is an obligate outcrossing plant with significant genetic variation (20) which was confirmed when looking at EST from *S. hermonthica* types collected from diverse locations and hosts (19). Further transcriptomic work on *S. hermonthica* has eluded to the core set of genes required for parasitism, specifically regarding the formation of parasitic plant organs such as the haustorium (21). These genomic and transcriptomic studies have facilitated a greater genetic understanding of the *S. hermonthica*.

The successful completion of the parasitic life-cycle necessitates a successful connection to the host but can be decomposed into several key stages. The *S. hermonthica life-cycle* involves

the following key stages; below ground (1) dormancy (2) conditioning, (3) germination, (4) haustorium formation, (5) attachment to the host and the above ground (6) growth and maturation. Here the current state of knowledge on each of the stages will be discussed to provide a thorough understanding of the *S. hermonthica* parasitic cycle, with detail being focused on advances made in the last decade as prior reviews have explained the lifecycle concisely (4). As will be described further, an understanding of the life cycle is vital as each stage represents a unique opportunity for phenotyping quantitative resistance mechanisms to *S. hermonthica* in sorghum and other host crops of global importance.

1.2.1 Seed dormancy and conditioning

Seed dormancy is an innate suspension in germination in an otherwise viable seed after maturity. *S. hermonthica* displays two types of dormancy, a primary dormancy and a secondary dormancy that is also known as a wet dormancy. The primary dormancy of *S. hermonthica* typically lasts for six months after seed maturity (22). Under conducive conditions, *S. hermonthica* will transition from primary dormancy to stage known as conditioning. During the conditioning phase, the *S. hermonthica* seeds must be exposed to warm and moist environmental conditions for 2 to 14 days, initiating metabolic processes that facilitates detection of an exogenous germination stimulant produced by the host (23). If the exogenous germination stimulant is not perceived *S. hermonthica* can desiccate and enter into a secondary dormancy (22,24). This secondary dormancy allows the parasite to persist in the soil until a host plant is available to support its lifecycle.

1.2.2 Strigolactone induced germination and perception

As early as 1933 it was reported that *Striga* seeds would only germinate when they were near a host root or when irrigated with water that had percolated across maize roots (25). Thus, it was concluded that direct contact between *Striga* and the host is not required and that an exogenous compound exuded from the host root stimulates germination. The exudate of the sorghum host also germinated *S. hermonthica* which indicated a host specificity of the parasite to the Poaceae family and established that the putative compound responsible for germination contained a pentose sugar (26). The chemical structure of *Striga* germination stimulant, strigol, was determined

utilizing exudates from the roots of cotton (*Gossypium hirisutum*) which is a non-host plant (27) This research demonstrated that strigol solutions of 10^{-11} M concentrations could provide 50% germination rates in *Striga*, eluding to the remarkable sensitivity of the strigol receptors in the parasite. By 1986, sorgoleone was reported as the first *Striga* germination stimulant isolated from root exudates of sorghum, a *Striga* host species (28). However, further studies indicated that sorgoleone was on par in sorghum cultivars that were known as susceptible and resistant in laboratory and field reaction to *Striga* (29). This discrepancy necessitated further study regarding the *Striga* germination stimulants produced by host plants. Sorgolactone was confirmed to be the primary *Striga* germination stimulant from sorghum root exudates (30). Likewise, strigol was confirmed as the primary germination stimulant from the maize and porso millet root exudates (31). Strigol and sorgolocatone are both sesquiterpene lactones that are active as *Striga* germination stimulants and thus began to be referred to collectively as strigolactones, as proposed by Dr. Larry Butler at Purdue University (32). Analysis of sorghum under laboratory conditions revealed that *S. hermonthica* susceptible cultivars tended to exude the specific strigolactone 5-deoxystrigol, while resistant cultivars exuded orobanchol (33). A similar association of *Striga* susceptibility with 5-deoxystrigol exudation occurs in maize (34).

With the chemical nature of the *S. hermonthica* determined it remained unclear on how the parasite perceived these host-derived signaling molecules. Recently the perception of strigolactones by *S. hermonthica* received needed attention as the mechanism of how the parasite perceived strigolactone germination stimulants were not previously well understood. As early as 1972 it was demonstrated that even picomolar concentrations of strigol could stimulate *Striga* germination (27). Research on the endogenous perception of strigolactones in *Arabidopsis thaliana* identified a group of genes encoding α/β hydrolases as the putative strigolactone receptors, specifically AtDWARF14 (*AtD14*) and HYPOSENSISTIVE TO LIGHT (*AtHTL)*/KARRIKIN INSENSITIVE2 (35–39). Further examination revealed that *S. hermonthica* contains 12 genes that are homologous to *AtD14 and AtHTL,* further designated as *ShD14* and *ShHTL1* to *ShHTL1* (40). A greater understanding of the affinity of given *S. hermonthica* receptors for synthetic and natural strigolactones was facilitated by the development of Yoshimulactone Green (YLG), a fluorescent small molecule reporter that indicates strigolactone perception by *S. hermonthica.* This research demonstrated that GR24 and 5-deoxystrigol produced the greatest inhibitory concentration with YLG, meaning that 5-deoxystrigol displays the highest binding affinity to the

majority of the putative *S. hermonthica* strigolactone receptors (40). The demonstrated affinity of the strigolactone receptors to 5-deoxystrigol provides evidence for the mechanism behind the laboratory and field observations that indicate host plants producing 5-deoxystrigol have higher germination stimulant activity and thus are more susceptible to *S. hermonthica* in the field.

Interestingly, *ShHTL6* and *ShHTL7* appear to be promiscuous receptors in that they have an affinity for a range of strigolactoness including 5-deoxystigol and orobanchol, unlike *ShHTL2-5* and *ShHTL8* which have preferential affinity for only 5-deoxystrigol (40). The putative strigolactones receptors *ShHTL1-9* were inserted as transgenes in the *Arabidopsis htl-3* genetic background and the resulting seeds germinated to varying degrees with exogenous application of GR24 (41). This research found that only *ShHTL7* perceives GR24 at the picomolar concentrations observed in *S. hermonthica*, and thus the researchers concluded *ShHTL7* to be the most functionally sensitive receptor for germination of the parasite in response to exogenous strigolactones signals from the host (41).

Strigolactones are not the only plant hormone required for successful germination. It is now known that *S. hermonthica* will initiate but not complete the phases of germination in response to exogenous GR24 and 5-deoxystrigol when ethylene biosynthesis is inhibited (40,42). By observing the fluorescence of YLG in *S. hermonthica* embryos, we now know that strigolactone perception during germination occurs in three distinct wavelike phases; (1) wake-up phase, (2) pre-germination pause, and (3) the elongation tide (40). Strigolactone perception occurs approximately twenty minutes after exogenous application of YLG in a period known as the wake-up phase followed by a pre-germination pause. Following this pause, there is a second wave of fluorescence concentrated at the *S. hermonthica* root tip and subsequent elongation of the root (40). This concentration of strigolactone perception activity indicates that *S. hermonthica* utilizes chemotaxis towards a strigolactone concentration gradient to develop the radical towards the host root while in its free-living state.

1.2.3 Haustorial formation and attachment

S. hermonthica follows the characteristic signature of all parasitic plants in that it can directly feed on the host plant, here specifically the roots, via a specialized organ of parasitism called the haustorium (1). It is vital that the germinated *S. hermonthica* find and penetrate the host as soon as possible because it has minimal reserves in the small seed. The haustorium allows *S.*

hermonthica to attach to the roots of the host plant by invading the root system and making a direct connection to the xylem. For the haustorium to form an additional signal must be perceived from the host root. Studies on *S. asiatica* demonstrated that 2,6-dimethoxy-p-benzoquinone (DMBQ) is the principal haustorium inducing factor exuded from sorghum roots (43). DMBQ was demonstrated to be the product of cell well oxidation suggesting that the parasite produces phenol oxidases to detect the host (44). This has not been demonstrated in *S. hermonthica* but is considered to be the putative mode utilized by the parasite to initiate the formation of the haustorium. Upon perception of the haustorium inducing factor, the root cap of the *S. hermonthica* radical will swell, and the epidermal cells and haustorial hairs will form. It is assumed that cell wall degrading enzymes exuded by *S. hermonthica* facilitate haustorial penetration into the host. Transcriptomic studies have revealed that genes encoding cell wall degrading enzymes, specifically glycosyl hydrolases and pectin lyases are upregulated in haustorial tissues from *S. hermonthica* (Yang et al. 2015). Subsequently, the epidermal cells will differentiate into intrusive cells that invade the host plant and become xylem cells forming a xylem-to-xylem connection between the host and parasite (45,46).

1.2.4 Growth and maturation

With a successful connection to the xylem, the establishment of parasitism is complete, and the parasite can successfully extract water and dissolved nutrients from the host. Once a successful attachment is established, *S. hermonthica* will grow underground for four to seven weeks before emerging as a photosynthetically active plant (47). It is approximated that at least 35% of the parasites carbon is derived from the host (48). The pre-emergent stage of *S. hermonthica* causes significant damage to the host plant, and the symptoms include wilting, chlorosis and severe stunting which in many cases prevent the host plant from reaching maturity. Postemergence, the *S. hermonthica* plant will photosynthesize, but it is not capable of surviving without continued attachment to the host (25). *S. hermonthica* will flower within three to four weeks after emergence, and each plant is capable of producing up to a half a million seeds upon maturity that have the potential to remain viable in the soil for over a decade (49).

1.3 Quantitative resistance *S. hermonthica* in sorghum

The improved understanding of the *S. hermonthica* lifecycle has facilitated additional discoveries regarding quantitative resistance to the parasite in the sorghum host. The interaction between the host and the parasite illustrates the putative quantitative resistance mechanisms. It was hypothesized that each critical stage of the *S. hermonthica* life cycle could be dissected into a qualitative laboratory assays that could be used to study quantitative resistance in sorghum without utilizing cumbersome field observations (15). By analyzing the interaction of *S. hermonthica* and the host, laboratory assays have been developed to study the following putative resistance mechanisms; low germination stimulant, low haustorial factor, and the hypersensitive response. This approach has identified sorghum genotypes with these mechanisms of *S. hermonthica* resistance and facilitated the detection of genomic regions that confer resistance to the parasite.

1.3.1 Indirect analysis of *S. hermonthica* resistance using laboratory assays

An agar gel assay was developed to examine the variation in *Striga* germination stimulant activity in host plants (29). The premise being that indirect phenotyping of germination stimulant activity would translate to increased resistance of indirectly selected sorghum accessions that have low germination stimulant activity. By utilizing the agar gel assay, sorghum accessions having high and low *Striga* germination stimulant activity are distinguishable qualitatively. The agar gel assay was used to phenotype sorghum a bi-parental mapping population and associate the loci resulting in low *Striga* germination stimulant activity to a 30-gene region on chromosome five, known as *LOW GERMINATION STIMULANT 1* (*LGS1*) (50). The agar gel assay together with chemical characterization of the strigolactones present in root exudates from a variety of natural variants at *LGS1* were used to identify the causative gene as *Sobic.005G213600,* which encodes an uncharacterized protein with a sulfotransferase domain (51). As noted, *S. hermonthica* germinates when it perceives 5-deoxystrigol, a strigolactone that is exuded by the roots of the sorghum host plant (33). In *lgs* mutants where *LGS1* has been deleted the sorghum plants exude an alternate form of strigolactone known as orobanchol that does not facilitate the germination of *S. hermonthica* to wild-type levels. To date, the deletion of *LGS1* is the only known mechanism of *S. hermonthica* resistance that has been characterized.

A modified agar gel assay known as the extended agar gel assay was used to characterize haustorial initiation, haustorial attachment, and signatures of a hypersensitive response reaction.

As discussed, after germination *S. hermonthica* requires an additional host-derived signal to form a haustorium. Some sorghum genotypes to facilitate germination of the parasite but the haustorium of *S. hermonthica* may not form, and this is known as a low haustorial factor resistance mechanism. The use of the extended agar gel assay identified the wild sorghum accession P78 as having the low haustorial factor resistance mechanism, and segregation analysis indicated that this resistance mechanism might be inherited through a single dominant gene (4). The genomic region or associated genes conferring low haustorial factor resistance was mapped to 19.3 centimorgans (cM) from the SSR marker txp358 (52), which resides on chromosome nine of the sorghum genome (53). The extended agar gel assay facilitated further understanding of the genetic mechanisms that influence the hypersensitive response reaction by sorghum to *S. hermonthica* parasitism. In the *S. hermonthica* and sorghum context, the hypersensitive response is characterized by the formation of necrotic lesions at the haustorial attachment sites (54). The locus that influences this hypersensitive response was reported to be 7.5 cM from the SSR marker txp96 (52), which is on sorghum chromosome two (53).

1.3.2 Direct analysis of *S. hermonthica* resistance in infested fields

Studies that analyzed the field reaction of sorghum to *S. hermonthica* parasitism *in situ* have also provided limited insight into the genomic regions that influence resistance. The field reaction of sorghum to *S. hermonthica* is characterized primarily using the terms resistance and tolerance. In this context, a sorghum accession with resistance would host significantly fewer *S. hermonthica* plants than a susceptible accession, whereas tolerance indicates an accession that can support a susceptible level number of *S. hermonthica* plants but can still produce an acceptable yield (55). Field screening for *S. hermonthica* quantitative resistance measures resistance and tolerance, which may consist of the assayable resistance mechanisms such as low germination stimulant, low haustorial factor and the hypersensitive response that can't be resolved under field conditions. Effective field screening for *S. hermonthica* quantitative resistance is hindered because natural field infestations of the parasite are highly heterogeneous and are strongly influenced by environmental variation. This hindrance can be overcome to some extent by inoculating the testing location with *S. hermonthica* seeds and utilizing appropriate experimental designs that are replicated (11,12,56). The predominant phenotypes used to measure quantitative resistance are counts of the number of *S. hermonthica* per experimental unit over a period of time

which can be used to construct derived traits such as the area under the *Striga* number process curve (ASNPC) (57). Some studies phenotyped sorghum germplasm and identified accessions with field resistance to *S. hermonthica* (9,10,58). The sorghum that was identified as resistant to *S. hermonthica* we later utilized to construct bi-parental mapping populations that facilitated quantitative trait loci (QTL) analysis (59). The *S. hermonthica* resistant accessions IS9830 and N13 were crossed to a common susceptible parent, E36-1. The ASNPC was utilized to map a total of 29 QTL across the two populations that are associated with *S. hermonthica* resistance (Table 1.1) (59). This research indicates that every chromosome in the sorghum genome is potentially contributing to *S. hermonthica* quantitative resistance. Given that this QTL analysis used a low-marker density, these intervals span large genomic regions which require further study for the actual *S. hermonthica* resistance genes to be identified.

1.4 Future research for understanding quantitative resistance to *S. hermonthica*

In 2009, the first version of the sorghum genome was published (60). An additional 49 diverse sorghum genomes were sequenced to provide a genome-wide understanding of the genetic variation present in sorghum (61). The sorghum reference genome was recently improved as version 3 was recently released, providing further genomic insight(62). Studies have also analyzed the transcriptome of sorghum to assess the differential gene expression, some under certain abiotic and biotic stresses (62–64) These genomic and transcriptomic resources have facilitated a greater genetic understanding of several quantitative traits such as plant architecture, abiotic stress tolerance in sorghum, flowering time, kernel composition, yield and stalk rot resistance (65–71). With the exception of mapping *LGS1*, there have been no other studies to date that utilize these genomic and transcriptomic resources to improve our understanding of quantitative resistance to *S. hermonthica.* Improved sequencing technology can facilitate the exploration of the sorghum genome and transcriptome to resolve our knowledge of quantitative resistance to *S. hermonthica.*

1.4.1 Linkage mapping

Quantitative traits are controlled by the collective action of many genes and their interaction with the environment. The majority of agronomically important traits, including yield, abiotic stress tolerance and even *S. hermonthica* resistance in sorghum are considered to be quantitative. A QTL is a region of the genome that explains a statistically significant portion of

the phenotypic variation of a quantitative trait of interest, likely containing allelic variation responsible for the segregation of the phenotype. Linkage or QTL mapping involves locating genomic regions that influence a specific trait. The general common procedures of linkage mapping in plants will be mentioned here, however, the concise details of linkage mapping methodology have been concisely reviewed (72). In plants, the general procedure for linkage mapping involves producing a bi-parental mapping population by crossing two parent lines that show segregation for the trait of interest. This initial cross would be further advanced via self-pollination to produce hundreds of recombinant inbred lines (RIL) that display segregation for the trait of interest. Each RIL is phenotyped for the trait of interest and is genotyped using genetic markers. The genotypic data is then used to create a genetic map, which is the linear ordering of markers within a linkage group. Generally, linkage mapping techniques involve statistically associating a region between two markers with phenotypic variation of trait at a given level of significance. Overall these studies have indicated that in plant populations, quantitative traits are controlled by a limited set of small-effect QTL, in addition to very few QTL that produces large effects in the phenotypic variation of a trait (73). Family-based linkage mapping studies have been hindered because they have limited genomic resolution and are restricted to the allelic diversity found in a bi-parental population (74). As sequencing technology becomes increasingly affordable, it is now routine to genotype a bi-parental mapping population using SNP markers. This has facilitated high-density linkage mapping which improves the genomic resolution of a QTL due to an increased genetic marker density across the genome (75). However, this resolution is still limited by a lack of recombination events. To incorporate greater allelic diversity in the analysis, nested association mapping (NAM) populations have been designed in which several bi-parental populations are developed that all have a common female parent, that are crossed to a diverse set of male parents (76,77). Linkage mapping in NAM populations can be executed through a Joint QTL linkage mapping (78). The allelic diversity in these populations is relatively limited, for example there are only 26 founders in the maize NAM and 11 founders in the sorghum NAM which limits the usefulness of these populations if the founders do not have phenotypic variation for the trait of interest (76,77).

1.4.2 Association mapping

Association mapping typically makes phenotype to genotype associations in natural populations, commonly referred to as association panels, in contract to linkage mapping which using structured populations. Association panels consist of individuals that are selected to represent the diversity of the organism, such as the sorghum association panel (SAP) (79). Association mapping in association panels rely on ancestral recombination to resolve phenotype to genotype associations in a process called association mapping. Recombination limits resolution of a QTL on average to 10-20 cM because of the limited number of recombination events that occur in constructing the bi-parental or backcross populations (72,80). Association mapping is an improvement over linkage mapping in that it improves the resolution of the QTL, reduces the research time needed in population development and allows a greater allelic diversity to be analyzed (81). The increased resolution achieved in association mapping is a function of the structure of the linkage disequilibrium or the non-random association of loci across the genome. In a genome-wide association study (GWAS), association mapping is performed between phenotypes and genotypes across the entire genome (82). In sorghum GWAS in natural populations such as association panels have improved genetic understanding of several quantitative traits such as plant architecture, flowering time, kernel composition, yield and stalk rot resistance (Adeyanju, Little, Yu, & Tesso, 2015; Boyles et al., 2016; Chopra, Burow, Burke, Gladman, & Xin, 2017; Higgins, Thurber, Assaranurak, & Brown, 2014; Morris et al., 2013; Rhodes et al., 2017). However, all of these studies as with others using natural populations had to statistically control for population structure which limits the statistical power of these analyses.

Population structure is present in natural populations, for example association panels and even structured populations such as the NAM. Population structure is the presence of subpopulations with distinct allele frequencies due to non-random mating. Population structure can cause spurious phenotype to genotype associations in GWAS because it acts as a confounding variable. In plants non-random mating due to geographical origin, local adaptation and breeding history can result in population structure (81,83). Statistical corrections can be made to control for population structure but this limits the power to detect the causative genomic regions that influence the phenotype of interest (84,85). Another option to overcome the hindrance of population structure is to utilize a population design that incorporates enough random mating that

the population becomes unstructured, such as multi-parent advanced generation intercross (MAGIC) populations(86).

1.4.3 Association mapping in MAGIC populations

Generally, the development of a MAGIC population includes (1) selecting founder lines that segregate for the trait of interest, (2) random mate the founders for several generations, (3) randomly select a population of random mated individuals and advance them via self-fertilization. This procedure results in a MAGIC population of advanced RILs that do not have population structure because it was eliminated in the random mating step of the population development (86). The random mating also increases the number of recombination events which can potentially reduce the size of linkage blocks and increase the resolution for detecting sequence variants that influence a given phenotype. MAGIC populations have been developed in a number of crops including rice (87), wheat (88), maize (89), barley (90), cowpea (91) and sorghum (92). Aside from the elimination of population structure, MAGIC populations also have a significant advantage over association panels in that they produce a large breeding pool of novel variation that can be selected upon in breeding programs. However, the main disadvantage of MAGIC populations is that they take several generations to produce which can be laborious and cost prohibitive. This limitation can be further exacerbated if the species of interest does not exhibit a high degree of outcrossing or does not have an established male sterility system. Lastly, as with the NAM populations, the allelic diversity in a MAGIC population may be considerably less than an association panel if the diversity in the founder lines in genetically narrow. Despite these limitations, MAGIC populations have provided a greater understanding of the genomic regions influencing plant architecture, abiotic stress tolerance, disease resistance and nutritional composition in a number of crops and model species (93).

1.4.4 Mapping for *Striga hermonthica* Resistance in Sorghum

As discussed, linkage mapping in sorghum facilitated mapping a QTL that spanned 5.8 cM that explained a significant amount of the variation in *Striga* germination stimulant activity that was further fine-mapped to a 400 kb region (50). This study successfully delivered two of the main goals of linkage mapping; (1) selection of *Striga* resistant sorghum accessions using marker-

assisted selection and also, (2) aided in the gene discovery of *LGS1*. As mentioned by characterizing the strigolactones present in root exudates from a variety of natural sorghum variants *LGS1* was identified as *Sobic.005G213600* encoding an uncharacterized protein with a sulfotransferase domain (51). The mutant phenotype associated with loss of function causes a shift in the stereochemistry of the exuded SLs from 5-deoxystrigol to orobanchol in the root exudates of sorghum which confers resistance to *S. hermonthica*. To date *LGS1* is the only gene conferring a component of quantitative resistance to *S. hermonthica* that has been characterized. Additional studies have utilized linkage mapping and identified 29 QTL that are associated with *S. hermonthica* resistance in two bi-parental populations using field-collected *S. hermonthica* count data (59).

This *in situ* linkage mapping study illustrates that *S. hermonthica* resistance is indeed quantitative. There have been no studies in sorghum that assess the natural variation in sorghum for tolerance to *S. hermonthica* parasitism. By leveraging the low cost of whole-genome sequencing further linkage and association mapping studies could be conducted to determine the genomic regions that confer quantitative resistance on tolerance to *S. hermonthica* in sorghum. However, the existing association panels may not be useful for association mapping for this trait because there may not be significant variation for the phenotype of interest to detect associations. For example, the SAP has a total of 377 individuals, yet only three accessions (SAP-167, SAP-345, and SAP-360) have a reported resistant field reaction to *S. hermonthica* (94). Meaning that less than 1% of the population has a potential sequence variation for the phenotype. Unfortunately, these rare variants may be lost before the association analysis occurs because minor alleles with a frequency of less than 5% are filtered out because they cannot be distinguished from sequencing errors when low-depth sequencing methods such as genotyping-by-sequencing are used (95). Therefore, careful attention must be given when selecting the population that will be utilized for association mapping. MAGIC, NAM and bi-parental populations are structured and require substantial resources to develop, however with careful selection of the founder lines they could be utilized to identify the genomic regions that confer quantitative resistance and tolerance to *S. hermonthica.* Furthermore, these findings could facilitate the use of marker-assisted and genomic selection in expediting the development of elite sorghum accessions that resistant and potentially tolerate *S. hermonthica* parasitism.

The research activities that will be discussed in further chapters sought to leverage the advances in sequencing technology to conduct a GWAS for *S. hermonthica* resistance and tolerance in sorghum. To date, no GWAS has been conducted in any species for *S. hermonthica* resistance. A MAGIC population was used to conduct the GWAS. The PP37 MAGIC population was generated using 25 founder lines, of which 23 have been reported as resistant or tolerant to *Striga*. The populations were phenotyped for resistance and tolerance to *S. hermonthica* in the Northwestern corner of Ethiopia where the parasite is endemic during the 2016 and 2017 sorghum growing season. To account for spatial variation in *S. hermonthica* parasitism across the field an incomplete block design was used, with dense allocation of susceptible checks. Overall this research reveals that despite advances in our ability to genotype large populations segregating for putative *S. hermonthica* resistance mechanism, the primary challenge remains in accurately phenotyping these individuals.

Table 1.1: The previously detected QTL potentially conferring field resistance to *S. hermonthica.* The location of the QTL is based off of version 3 of the sorghum reference genome. Genes refers to the number of genes inside the support interval of the QTL.

QTL	Population	Chr.	Start (bp)	End (bp)	Interval (kb)	LOD	R^2	Genes
QSTRG1.1	N13xE36-1	1	7159896	7455090	295	6.7	24.2	49
QSTRG1.2	IS9830xE36-1	1	8957572	11628163	2671	2.9	10.7	325
QSTRG1.3	N13xE36-1	1	20162320	23989400	3827	4	14.6	201
QSTRG1.4	N13xE36-1	1	53840102	55365887	1526	7.4	26.5	57
QSTRG1.5	N13xE36-1	1	77289787	78396426	1107	3.3	12.5	143
QSTRG2.1	N13xE36-1	2	1533903	2267139	733	2.9	11.2	77
QSTRG2.2	IS9830xE36-1	2	2001997	2879255	877	3	11.6	95
QSTRG2.3	N13xE36-1	2	6058726	6080498	22	6	22.1	2
QSTRG2.4	IS9830xE36-1	2	56518961	57022492	504	5.8	21.9	33
QSTRG2.5	IS9830xE36-1	2	71851092	73416602	1566	5	18.9	224
QSTRG3.1	IS9830xE36-1	3	6900269	9483430	2583	3	11.2	257
QSTRG3.2	IS9830xE36-1	3	58225457	58692191	467	2.9	11.3	56
QSTRG3.3	N13xE36-1	3	68335734	69607180	1271	2.5	9.5	168
QSTRG3.4	IS9830xE36-1	3	71514043	72298881	785	2.8	11.1	102
QSTRG4.2	IS9830xE36-1	4	61565833	61917025	351	2.7	10.2	41
QSTRG4.1	N13xE36-1	4	61665535	61732660	67	12.4	40.5	7
QSTRG5.1	N13xE36-1	5	1994079	2867047	873	5.3	18.9	104
QSTRG5.2	N13xE36-1	5	12102612	13417632	1315	8	28.5	46
QSTRG6.1	IS9830xE36-1	6	3401736	4241078	839	4.4	16	39
QSTRG6.2	IS9830xE36-1	6	47664738	48915576	1251	4.2	16.2	169
QSTRG6.3	N13xE36-1	6	52528921	52773566	245	9	30	35
QSTRG6.4	IS9830xE36-1	6	60079492	60226099	147	12.7	41.5	24
QSTRG7.1	N13xE36-1	7	2486294	2915151	429	4.5	17.3	59
QSTRG7.2	IS9830xE36-1	7	54379522	57687844	3308	2.8	11.1	199
QSTRG7.3	IS9830xE36-1	7	63194655	65460255	2266	3.6	15.7	280
QSTRG8.1	N13xE36-1	8	1121704	1561135	439	2.6	9.9	62
QSTRG9.1	N13xE36-1	9	6107804	7507895	1400	3.8	4	104
QSTRG9.2	IS9830xE36-1	9	52990009	55552544	2563	3.1	11.7	355
QSTRG10.1	IS9830xE36-1	10	56335690	56921445	586	2.9	10.9	65

Figure 1.1: The parasite *in situ*.

A *Striga hermonthica* (purple flowers) infested sorghum field in Humera, Tigray, Ethiopia in 2017. The parasitized sorghum plants are stunted, chlorotic and also displaying whole plant necrosis. This level of infestation resulted in a 100% grain yield loss.

CHAPTER 2. INVESTIGATING A CEREAL KILLER: EXPLORING QUANTITATIVE RESISTANCE TO THE PARASITE *STRIGA HERMONTHICA* IN SORGHUM

2.1 Abstract

Sorghum (*Sorghum bicolor* [L.] Moench) is a staple food crop in some of the most food-insecure places in the world throughout Asia and Africa. In these regions where sorghum is grown, it is afflicted by a parasitic weed known as *Striga hermonthica*. *Striga hermonthica* is an obligate hemi-parasite that attaches to the roots of sorghum and can result in a total crop loss. Previously only one gene, low germination stimulant (*LGS1*) has been identified in sorghum that when deleted confers partial resistance to parasite by reducing the germination of *S. hermonthica*. The research presented here is the first to demonstrate that *lgs1* results in a significant reduction of *S. hermonthica* parasitism in a segregating population, the PP37 multi-parent advanced generation intercross (MAGIC). In order to conduct a genome-wide association study the PP37 MAGIC was phenotyped for two seasons in Northwestern Ethiopia, where *S. hermonthica* infestation is pervasive. In the first season a sub-threshold association with an FRD-corrected p-value of 0.07 was detected on chromosome three that co-localizes with a QTL that was previously mapped by the only other study that has attempted to map genomic regions conferring resistance to *S. hermonthica* field resistance. This association was not detected when the experiment was repeated in the following season, demonstrating the highly qualitative nature of *S. hermonthica*. In order to further understand the genetic architecture of *S. hermonthica* resistance in sorghum, effectively phenotyping remains the primary challenge.

2.2 Introduction

Approximately one percent of all angiosperms utilize parasitism in their lifecycle (96). Parasitic plants can feed directly from the host plant via an organ known as the haustorium (1). Perhaps the most notable members of the parasitic Orobanchaceae family are the *Striga* species as they parasitize agronomically important crop species resulting in significant negative impact on crop productivity worldwide (3). The value of the estimated annual yield loss from *Striga* worldwide is approximately $7 billion USD (3,4). *Striga* species have specific host preferences

and *S. hermonthica* preferentially parasitizes host plants in the Poaceae family, notably maize (*Zea mays*), rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*) (5).

For decades sorghum has remained the world's fifth most important crop after maize, wheat, rice, and barely (97). Sorghum is known as an abiotic stress tolerant member of the Poaceae family that utilizes the C4 photosynthetic process and is a staple food crop of the world's poor. One particular biotic stress, the parasitic plant *S. hermonthica*, greatly hinders sorghum production in areas of Asia and sub-Saharan Africa. *Striga hermonthica* taps the xylem of the host roots to extract water and nutrients, and contributes to the sorghum becoming stunted, chlorotic and wilted (Figure 2.1). In areas where *S. hermonthica* is endemic, parasitism can result in complete crop loss (4) (Figure 2.2). Sorghum possesses natural genetic variation for *S. hermonthica* resistance, although the understanding of the genetic mechanisms conferring resistance remains limited.

Generally in plants, disease resistance can be distinguished as either being qualitative or quantitative. The negative influence of *Striga* on its host is often likened to a disease. Qualitative plant resistance is a complete resistance conferred by a single gene, whereas quantitative resistance is an incomplete resistance conferred by multiple genes of small effect (6,98). Qualitative disease resistance is usually conferred by a single resistance (R) gene, whereas quantitative disease resistance is conditioned by numerous genes of small effect (6). To date there has only been one R-gene identified in the *Striga*-host pathosystem. *Striga gesnerioides* parasitizes members of the Fabaceae, Convolvulaceae, Euphorbiaceae, and Solanaceae families, including the agronomically important cowpea (*Vigna unguiculata*) (99). In cowpea, resistance to *S. gesenerioides* is conferred by the R-gene *RSG-301* which is predicted to encode a protein with a coiled-coil (CC) protein-protein interaction domain at the N terminus, a nucleotide binding site (NBS) and a leucine-rich repeat (LRR) at the C terminus (100). The CC-NBS-LRR, *RSG3-301* elicits a hypersensitive response (HR) in cowpea at the *S. gesenerioides* haustorial attachment site (100,101). Typically biotrophic pathogens, or a parasite in this case, elicit R-gene mediated HR in which the host plant tissue adjacent to the site of infection displays rapid programmed cell death (102). In plants, the majority of R-genes encode for NGS-LRR proteins (103). The sorghum genome contains 24 putative CC-NBS-LRR encoding genes (104). However, there has been no qualitative R-gene mediated *S. hermonthica* resistance reported in sorghum, or any other grass species. Signatures of the hypersensitive response, including necrosis at the haustorium attachment site, have been observed in some cultivars and wild accessions of sorghum parasitized with *Striga asciatica* (54).

Linkage mapping in a bi-parental backcross population identified loci on chromosome two and five that are associated with HR, but further refinement of these genomic regions have not been reported (52,105).

Aside from the potential existence of R-gene mediated HR, resistance to *Striga* species in sorghum appears to be quantitative as is generally demonstrated with grass host species (106,107). Our current understanding of quantitative resistance to *Striga* was greatly facilitated by a research program at Purdue University which sought to deconstruct the life-cycle and interaction between the parasite and host into key events that could be analyzed in detail with unique laboratory assays (4,14,15) . The first committed and non-reversible step in the *Striga* life cycle is the perception of the strigolactone germination stimulant from the host plant. The *in vitro* agar gel assay was specifically developed to examine the variation in *Striga* germination stimulant activity in host plants (29). The agar gel assay was used to phenotype sorghum a bi-parental population and linkage mapping was used to associate the locus resulting in low *Striga* germination stimulant activity to a 30-gene region on chromosome five, known as *LOW GERMINATION STIMULANT 1* (*LGS1*) (50). The agar gel assay together with chemical characterization of the strigolactones present in root exudates from a variety of natural variants at *LGS1* was used to identify *LGS1* as *Sobic.005G213600* encoding an uncharacterized protein with a sulfotransferase domain (51). *Striga hermonthica* preferentially germinates when it perceives 5-deoxystrigol, a strigolactone that is exuded by the roots of the sorghum host plant (33). In *lgs* mutants, where *LGS1* has been deleted, the sorghum plants exude an alternate strigolactone, orobanchol that results in a reduction of *S. hermonthica* germination. To date, the deletion of *LGS1* is the only known mechanism of *S. hermonthica* quantitative resistance that has been characterized.

Additional linkage mapping studies in two bi-parental populations identified 29 QTL across the sorghum that was associated with *S. hermonthica* resistance using phenotypic data that assessed parasitism in infested field locations (59). Linkage mapping provides limited resolution of the genomic regions influencing a trait and is limited to the allelic diversity found in the parental lines (72). Association mapping is routinely used in sorghum to determine the sequence variants responsible for several quantitative traits such as plant architecture, flowering time, kernel composition, yield and stalk rot resistance (Adeyanju, Little, Yu, & Tesso, 2015; Boyles et al., 2016; Chopra, Burow, Burke, Gladman, & Xin, 2017; Higgins, Thurber, Assaranurak, & Brown, 2014; Morris et al., 2013; Rhodes et al., 2017). These studies relied on assessing phenotypes in

natural populations known as association panels or diversity sets which lack mating designs and are simply germplasm collections selected to represent the diversity of a species, such as the sorghum association panel (SAP) (94). Such populations may not be suitable for mapping quantitative resistance to *S. hermonthica* because the number of accessions with resistance to the parasite is very low. For example, only three accessions (SAP-167, SAP-345, and SAP-360) of the 377 lines in the SAP have reported resistance to *S. hermonthica.* In this case, the sequence variants conferring resistance to the parasite would potentially have allelic frequencies so low that are indistinguishable from genotyping errors when low depth sequencing methods such as genotyping-by-sequencing (GBS) are utilized (95,108). Therefore, it is necessary to conduct such mapping efforts in populations that have been specifically designed to evaluate *S. hermonthica* resistance.

The sorghum PP37 multi-parent advanced generation inter-cross (MAGIC) population was established using 25 founder lines, 23 of which are resistant or tolerant to one or more *Striga* species (Table 2.1). CSH1 and Swarna the only two accessions that are reported as susceptible to *Striga* (109). All of the SAR accessions have been reported to be resistant to *S. asiatica* (110). SRN39, Tetron and 555 were utilized to determine the location of *LGS1*, which when deleted confers the low *Striga* germination phenotype (111). The PP37 founders were genotyped for mutations in *LGS1*, and it was determined that 14 of the founders have *lgs1* alleles, meaning they have at least partial quantitative resistance to *S. hermonthica* because they do not stimulate the germination of the parasite to wild type levels. Framida and Dobbs have a reported hypersensitive response reaction to *Striga* in which necrotic lesions form at the haustorium attachment site of the sorghum roots (54). Framida also partially inhibits the formation of a functional *S. hermonthica* haustorium which limits resource diversion from its host (112). N13 and IS4202 exhibit increased lignification at the point where the haustorium attempts to penetrate the endodermis effectively creating a mechanical barrier against the parasite (109). P67083 produces a low density of roots near the soil surface therefore potentially avoiding *S. hermonthica* parasitism (58). Overall the PP37 MAGIC population was constructed using a diverse collection of founder accessions with the majority of them exhibiting *S. hermonthica* resistance from several different mechanisms.

The PP37 MAGIC was phenotyped for *S. hermonthica* resistance in Northwestern Ethiopia where the parasite is endemic during the 2016 and 2017 growing seasons. The number of *S. hermonthica* plants per experimental unit was quantified at three-time points across the growing season and was used to calculate the area under the *Striga* number progress curve (ASNPC) which

provides a measure of the proliferation of the infestation (113). The experimental units were inoculated with conditioned *S. hermonthica* seeds. However the significant spatial variation in the natural parasite populations introduced experimental error. To control for this experimental error, the population was planted in an incomplete block design with dense allocations of the susceptible check, BTx623 in each block (114). The PP37 population consists of 1,000 individuals that were genotyped using genotyping-by-sequencing (GBS) (108). These phenotypic and genotypic data sets were used to conduct a genome-wide association study to determine the genomic regions associated with *S. hermonthica* resistance in sorghum.

2.3 Materials and Methods

2.3.1 Germplasm

The PP37 MAGIC population was developed at the Purdue University Center for Research and Education, Purdue University beginning in 1986. The population was originally developed as a recurrent selection population for the development of elite sorghum cultivars with potentially multiple mechanisms of *S. hermonthica* resistance. The PP37 MAGIC population was constructed using 25 founder sorghum accessions, of which 23 are known to be resistant to one of more species of *Striga*. The founders were randomly mated for six generations using the ms_3 genetic male sterility system (115). The genetic male sterility system involved crossing the founders to a stock of 10 random sorghum plants that carried the ms_3 allele. During the random mating, only male-sterile heads are harvested which ensures that all seeds produced on the head are the result of cross-pollination. After random mating, a total of 1,000 inbred lines were randomly selected and further self-pollinated by single seed descent to the S_6 generation.

2.3.2 Genotypic data

Tissue was collected from one seven-day old sorghum seedling of each accession of the PP37 MAGIC grown under greenhouse conditions at Purdue University. Approximately 40-50 mg of lyophilized leaf tissue was used for DNA extraction. DNA was extracted using the CTAB (cetyl trimethylammonium bromide) method with minimal modifications (116). After elution, the DNA was cleaned using a 1.5:1 by volume ratio of Axygen Clean-Seq beads (Corning Life Sciences, Corning, NY, USA) to remove any remaining inhibitory compounds in the sample. DNA was quantified using Quant-IT PicoGreen fluorescent dye (Thermo Fisher, Waltham, MA, USA).

The concentration of the DNA was quantified using the Quant-iT™ PicoGreen® dsDNA kit (Life Technologies, Grand Island, NY). Library preparation was done using a genotyping-by-sequencing (GBS) protocol with the following modifications; 50 ng of DNA was digested with a 5 basepair cutter ApeKI (New England Biolabs, Ipswich, MA) followed by ligation of barcoded adapters with T4 ligase (New England Biolabs, Ipswich, MA) (108). The adapter-ligated samples are pooled, amplified, and adapter dimers were removed with SPRI bead purification. The quality and quantity of the finished libraries were assessed using the Agilent Bioanalyzer High Sensitivity Chip (Agilent Technologies, Inc., Santa Clara, CA) and Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY). The libraries were standardized to a 2nM concentration, and the cluster was generated using the HiSeq SR Cluster Kit v3 (Illumina Inc, San Diego, CA, USA). Each flow cell was sequenced using single-end, 100 base pair reads on the Illumina HiSeq2500 sequencer. The resulting images were analyzed using the Illumina Pipeline, version 1.8.2. To analyze the GBS data and the TASSEL-GBS pipeline was utilized (95)The original genotypic data set consisted of 449,208 single nucleotide polymorphisms (SNPs) that were filtered using TASSEL (117). A total of 71,448 SNPs with minor allele frequency (MAF) < 5% and missing data >80% were used in this study. The physical position of the SNPs was determined using sorghum genome v3.1.

2.3.3 Genotyping for *LGS1*

The entire PP37 MAGIC was grown in sand benches under greenhouse conditions at Purdue University. The accessions SRN39 and Shanqui Red were also grown to serve as the *lgs1* and *LGS1* check respectively. Leaf tissue was collected from one fourteen-day old seedling using a 0.3 cm Integra Militex standard biopsy punch (Fisher Scientific, Hampton, NH, USA) for each accession in the population and placed in a 96-well plate. The genomic DNA was extracted using a high-throughput method as previously described (118). The *LGS1* marker targets a genomic region where the mutant allele is a deletion, and therefore the target sequence is missing in *lgs1*. Therefore, a forward and reverse primer for a positive control needs to be used in addition to the forward and reverse primer for *LGS1* (Table 2.2). The gene *Sobic005G213200* that encodes for a kinase was used as a positive control because it resides upstream of *LGS1* and is present in all *lgs1* mutants. The primer set used for *LGS1* can distinguish between the two wild-type alleles (*LGS1-1* and *LGS1-2).* The *LGS1-1* allele has a larger amplicon size and is found in Shaqui Red, whereas

the *LGS1-2* allele is found in BTx623. Amplification was performed in a 20 µl reaction volume consisting of 10 µl of MyTaq Red (Bioline, London, UK), 1 µl of *LGS1* forward and reverse primer mix, 1 µl of the positive control forward and reverse primer mix, 1 µl of 20% BSA, 10 µl of 20% PVP and 1 µl of the DNA template. The thermal cycler was programmed for 2.5 minutes at 95 °C, for initial denaturation, 30 seconds at 65 °C for annealing, 60 seconds at 72 °C for extension for 35 cycles (but with 30 seconds at the denaturation temperature) and 5 minutes at 72 °C for the final extension (BioRad, Hercules, CA, USA). Upon completion of the PCR, 2-5 µl of each PCR product was loaded onto a 3% high-resolution agarose gel and amplicons were separated by size by electrophoresis in 0.5% Tris/Borate/EDTA buffer at 45V for 8 hrs. The lines carrying the *lgs1* allele (low *Striga* germination stimulant activity) have only one band which is the amplicon from the positive control primer pair (333bp). Those with the *LGS1* allele (high *Striga* germination stimulant activity) have two gel bands, one from the amplicon primed by the *LGS1* primer pair (208-220 bp), and one from the positive control primer pair (333bp).

2.3.4 *Striga hermonthica* field resistance phenotypes

The PP37 MAGIC population was grown during the 2016 and 2017 growing seasons in the Humera district in the state of Tigray located in Northwestern Ethiopia. In 2016 two replications were grown solely on research fields within the Humera Agricultural Research Center (HARC) of the Tigray Agricultural Research Institute. In 2017 two replications of the experiment were planted in *S. hermonthica*-free conditions and two replications were planted under *S. hermonthica* infestation at HARC in addition to two replications under infestation in a farmer's field in Adabay which is also in the Humera district approximately 20 kilometers from HARC. In 2016 and 2017 all replications consisted of plots that were 3 meters in length with 1.5 meters between the plots and 1-meter allies to prevent germination stimulant from influencing *S. hermonthica* infestations in the neighboring plot. Each replication was weeded three times during the growing such that *S. hermonthica* was the only weed growing in the experimental unit to prevent other Poaceae species from spuriously germinating the parasite. In 2016 each replication was grown in a randomized incomplete augmented block design, where each replication consisted of 100 blocks and each block had the susceptible check Dabar randomized among the entries. In 2016 phenotypic data was collected on the number of *S. hermonthica* emerged per plot (at 50 and 75 days after planting), *S. hermonthica* height, *S. hermonthica* branch number, *S. hermonthica*

emergence distance from the sorghum, for each plot in the replications. In 2017 each replication was grown in a randomized incomplete augmented block design, where each replication consisted of 200 blocks and each block had the susceptible check BTx623 randomized among the entries. Dabar displayed insufficient susceptibility in 2016 and a significantly longer duration than the PP37 population at large, potentially escaping parasitism temporally. The reference genome BTx623 was observed in 2016 at HARC and demonstrated to be highly susceptible to *S. hermonthica* and of similar maturity to the populations at large, making it a better susceptible check than Dabar. In 2017 phenotypic data was collected on the number of *S. hermonthica* emerged per plot (at 40, 50 and 75 days after planting), *S. hermonthica* height, *S. hermonthica* branch number, *S. hermonthica* emergence distance from the sorghum, total sorghum plant height, panicle length, flag leaf height, pre-flag leaf height and chlorophyll meter values (SPAD) for each plot. Successive *S. hermonthica* counts were used to calculate the area under the *Striga* number progress curve (ASNPC) using the following formula with slight modifications (57):

$$\text{ASNPC} = \sum_{i=0}^{n} = \left[\frac{(Y_i + Yj)}{2}\right] \ (T_i \ - T_j)$$

Where:

Y_i is the ith *Striga* count

Y_j is the jth *Striga* count

T_i is the ith *Striga* count date

T_j is the ith *Striga* count date

It was necessary to remove blocks where natural *S. hermonthica* infestations were insufficient. The *S. hermonthica* infestation in the complete second replication of 2016 was insufficient and thus was excluded from the subsequent association analysis, but was used to refine the experimental design in the following season. Blocks were excluded from the data set when the respective susceptible check had less than 30 *S. hermonthica* emerged plants per plot in 2016, and less than 150 *S. hermonthica* emerged plants per plot in 2017 to remove experimental units where infestations were insufficient. Best linear unbiased predictors (BLUP) were calculated for each trait using the filtered phenotypic data for one replication in 2016 and four replications in 2017. The BLUPs from 2016 represent a single replication whereas the experimental unit had to be

present under sufficient infestation in two out of the four replications grown in 2017 for the derivation of BLUPs and to be used in subsequent association analysis. The BLUPs were calculated as $y_{ij}= \mu+\tau_i+ \beta_{j(\rho)} + \varepsilon_{ijk}$ where y_{ij} is the ijth observation, μ is the grand mean, τ_i is ith fixed genotypic effect and $\beta_{j(\rho)}$ is the block effect nested within replication ρ.

2.3.5 *Striga hermonthica* germination stimulant activity

The *S. hermonthica* germination stimulant activity of a random sub-sample (n =87) of the PP37 MAGIC population was determined using the agar gel assay as previously described (29) with minimal modifications. Briefly, ten sorghum seeds from each sorghum accession were surface-sterilized for 30 minutes in a 50% bleach solution (2.6% sodium hypochlorite) containing 0.2% Tween-20 (polyethylene glycol sorbitan monolaurate; Bio-Rad Corp.) and then imbibed overnight in a 5% aqueous slurry of Captan fungicide (active ingredient: N-trichloromethylthio-4-cyclohexene-1,2-dicarboximide, 48.9%; Arysta LifeScience North America LLC) in an incubator at 30°C to germinate. One germinated sorghum was transplanted to each of three 100mm petri dishes containing 0.7% agar into which conditioned *S. hermonthica* seeds were embedded at an average density of 40 seeds/cm^2. The plates were incubated in the dark for three days at 29°C. The *S. hermonthica* germination stimulant activity was inferred from a measure of the maximum germination distance (MGD). The MGD was determined by averaging the distance (mm) of the three furthest germinated *S. hermonthica* seeds from the sorghum root on three plates per accession. The *S. hermonthica* germination stimulant activity is scored as high if MGD was ≥10 mm and low if the MGD was below 10 mm. The *S. hermonthica* utilized for the agar gel assay was collected from Pawe, Ethiopia and showed a germination rate of 80% in agar treated with a 10^{-7} M solution of GR24. The agar gel assay was conducted at the Purdue University Parasitic Weed Containment Room.

2.3.6 Association analysis

The GWAS studies were conducted using the statistical genetics package Genome Association and Prediction Integrated Tool (GAPIT) (119,120). The GBS produced SNPs with a MAF of less than 0.05 and more than 20% were removed from the genotypic data set, and minor SNP states, or minor alleles were also removed, resulting in 71,448 SNPs. Associations were made

using a unified mixed linear (MLM) approach with kinship to account for relatedness among individuals (85). The LGS1 marker status of each individual of the PP37 population were used as a covariate in the MLM. The type I errors rate associated with multiple testing were controlled with a 5% false discovery rate (FDR) that were calculated with the Benjamini and Hochberg procedure (121) in GAPIT. The pairwise correlation between SNPs was calculated in TASSEL 5.2 to estimate linkage disequilibrium (LD) (117). The narrow sense heritability (h^2) was estimated using a compressed mixed linear model to estimate the additive genetics effects using marker-based kinship in GAPIT (122). There were eight and seven traits used in 2016 and 2017 respectively to conduct the GWAS for *S. hermonthica* resistance in the PP37 MAGIC (Table 2.3). Likewise, there were 513 and 916 individuals from the PP37 MAGIC that were grown under sufficient infestation to be represented in the 2016 and 2017 respectively included in the GWAS.

2.4 Results

2.4.1 Structure of the PP37 population

The PP37 MAGIC population was initially established to generate a recurrent selection population by recombining several elite sorghum accessions with accessions known to by field resistance to *S. hermonthica* by different putative modes of action. The PP37 population was established with 25 founder accessions of which 23 have reported field resistance to *S. hermonthica* (9–11,110). To assess the putative resistance mechanisms, the PP37 population founders were genotyped with the *LGS1* marker. In total, 22 of the PP37 founders were genotyped as the seed was not available for the remaining three founder accessions. Out of the 22 founders that were genotyped 14 had *lgs1* alleles and eight had wild-type *LGS1* alleles (Table 2.1). At least 56% of the founders have some extent of *S. hermonthica* resistance conferred by the low germination stimulant activity attributed to *lgs1*. The exact *S. hermonthica* resistance mechanism of the other founders reported to be resistant is unknown. Therefore, the PP37 MAGIC population derived from these founders was further analyzed to better understand *S. hermonthica* resistance in sorghum.

In generating the PP37 MAGIC population, the founders were random mated for six generations using the ms_3 genetic male sterility system. The phylogeny of the PP37 population founders was analyzed using neighbor-joining trees and the 23 founders that were analyzed

clustered into seven distinct groups (Figure 2.3). Similarly, the phylogeny of the entire PP37 MAGIC population was analyzed and only clustered into one group and the lines are distributed at equal distances from the root of the tree (Figure 2.4A), suggesting that the population structure present in the founders were eliminated during the successive random mating that occurred during the development of the population. The lack of population structure on the PP37 MAGIC is further demonstrated with principal component analysis, in which no significant relationship was observed (Figure 2.4B).

2.4.2 Genome-wide association of *S. hermonthica* germination stimulant activity

The *S. hermonthica* maximum germination distance was determined for a random subset (n=85) of the PP37 MAGIC population. Of the subset that was analyzed, 54% of the individuals scored as having high *S. hermonthica* germination stimulant activity, with the remaining tested individuals being scored with low germination stimulant activity. All individuals of the PP37 population were genotyped for the *LGS1* marker. The average maximum germination distance of those with the wild type *LGS1* is significantly higher than those with *lgs1* (Figure 2.5A). There is a perfect correlation ($r^2 = 1.0$) between the MGD phenotype from the agar gel assay and the *LGS1* genotype (Figure 2.5B) on the subset of individuals that were tested. A GWAS of MGD on this subset mapped significantly to one and only one region of the genome, on chromosome five (Figure 2.6). This GWAS significantly associated 42 SNPs with the maximum germination distance score using $\alpha = 0.05$ FDR adjusted p-value threshold. The significant SNPs spanned from positions 69,042,547 to 69,943,558bp, a 900kb region. The previously characterized *S. hermonthica* resistance gene *LGS1* (*Sobic005G213600*) resides within this interval. The peak SNP (*S5_69539513*) is 444,927bp from the start site of *LGS1*. Linkage disequilibrium analysis indicated there were two linked regions contributing to this genomic association. An additional SNP (*S5_69943558*) of on equal significance to the peak SNP is only 40,882 bp from the start site of *LGS1*. The *LGS1* marker status of the individuals was used to infer the germination stimulant activity of the entire PP37 MAGIC. Genotyping for the *LGS1* marker revealed that 57% of the PP37 MAGIC individuals have the wild type *LGS1* allele while 43% have the *lgs1* allele (Figure 2.7). From this, it is inferred that 43% of the population have the low *S. hermonthica* germination phenotype which conveys partial resistance to the parasite. Here we demonstrate that the *lgs1* allele

results in a significant reduction in the number of *S. hermonthica* parasitizing the host plant at 40, 50 and 75 days after planting across the PP37 MAGIC (Figure 2.8).

2.4.3 Genome-wide association of *S. hermonthica* field resistance

A GWAS was conducted for all *S. hermonthica* resistance phenotypes (Table 2.3) in 2016 and 2017. The Manhattan plots for these studies indicate that there are no statistically significant associations when using an FDR adjusted *p*-value of 0.05. All association studies that utilize the *S. hermonthica* count or the derived ASNPC as a phenotype detect a subthreshold peak at *LGS1,* for example the 2017 GWAS for ANSPC in the PP37 MAGIC (Figure 2.9A). When the *LGS1* marker is used as a covariate the subthreshold peak for *LGS1* is no longer present, validating that this peak is *LGS1*. The most significant SNP of the sub-threshold *LGS1* peak $-\log_{10}$ *p*-value of 3.53 which is an FDR corrected *p*-value of 0.34. Given that we demonstrated that the *lgs1* allele results in a statistically significant reduction in *S. hermonthica* and that a sub-threshold peak was detected for *LGS1*, we propose investigating sub-threshold peaks with SNPs in the peak at or exceeding $-\log_{10}$ p-value of 3.0, further referred to as the LGS-level. There are 1,943 SNPs that cross the LGS-level for the *S. hermonthica* count data and ASNPC combined from the 2016 and 2017 data sets. To provide further evidence that these peaks are not spurious, comparisons were made to previously mapped QTL.

A notable sub-threshold peak was detected on chromosome three using the 2016 *S. hermonthica* count data at 75 days after planting, further referred to as ST-SH3.1. The peak SNP (*S3_8564738*) of ST-SH3.1 had an FDR-corrected p-value of 0.07. The associated Manhattan plot reveals that the ST-SH3.1 peak is significantly higher than the LGS-threshold (Figure 2.10). To provide further evidence that these peaks are not spurious, a comparison was made to previously mapped QTL that is associated with *S. hermonthica* count data (59,123). The ST-SH3.1 peak is located within the support interval of one such previously mapped QTL, known as QSTRG3.1. To map QSTRG3.1, a bi-parental mapping population was utilized that IS9830 as the resistant parent, which is as a founder line of the PP37 MAGIC population and carries a *lgs1* allele. There are 257 putative candidate genes within QSTRG3.1. The 100,000 base pair linkage block around the peak SNP of ST-SH3.1 reduces this candidate gene list to 20 candidate genes (Table 2.4). No peaks were detected with the 2017 data that co-localized with ST-SH3.1. Furthermore, no other peaks

were detected in the GWAS that co-localized across years when considering all SNPs that cross the LGS-level.

2.5 Discussion

Despite decades of research and effort, limited advancements have been made in understanding the genetic architecture of *S. hermonthica* resistance in sorghum. The seminal studies regarding *S. hermonthica* quantitative resistance in sorghum were hindered by phenotypic and genotypic limitations (50,59). Resistance to *S. hermonthica* is a highly quantitative trait. There is profound variation in *S. hermonthica* infestations even under inoculated well-designed experiments. Also, the germination and subsequent parasitic lifecycle of *S. hermonthica* is highly influenced by the environment, to an extent that is only partially understood. These points highlight some of the main hindrances that confound phenotyping for *S. hermonthica*. Pioneering studies decomposed the quantitative nature of *S. hermonthica* resistance into qualitative laboratory-based assays that led to the detection of the only gene associated with a reduction in the germination of *S. hermonthica, LGS1*. The detection of *LGS1* was facilitated by the rapid advancement in sequencing technology. Undoubtedly, effectively phenotyping for *S. hermonthica* field resistance remains the primary hindrance in furthering our understanding of the genetic architecture conferring resistance to this parasite. To date, only one other study has evaluated a sorghum mapping population for *S. hermonthica* resistance in the field, and this study was limited by low genetic marker coverage (59). Advances in sequencing technology have facilitated larger populations to be examined with genetic resolution. We hypothesized that a purposefully designed large MAGIC population phenotyped with the careful experimental design would allow for greater resolution of the genetic architecture underlying *S. hermonthica* in sorghum. Overall the results of this GWAS are inconclusive, inviting a critical examination of the population, experimental design, and the results to inform future research on this elusive parasite.

2.5.1 An improved understanding of LGS1 and mapping structural variants

In sorghum, *LGS1* is the first, and the only gene reported that alters the primary exudate for the sorghum root from 5-deoxystrigol to orobanchol and thus reduces the germination rate of *S. hermonthica* (124). *LGS1* was characterized using the *in vitro* agar gel assay which resolves sorghum genotypes as either being high or low stimulators of *S. hermonthica* germination (29).

The PP37 population was originally designed as a recurrent selection population to develop elite *S. hermonthica* resistant sorghum inbred lines for hybrid development. The majority of the PP37 founders had been previously reported as resistant to *Striga* by varying putative modes of resistance. In total 14 of the PP37 founders have a *lgs1* mutant allele in which all or part of *LGS1* is deleted. In total 43% of the PP37 population had a homozygous *lgs1* genotype. This research is the first to demonstrate that individuals having only *lgs1* alleles have less severe *S. hermonthica* field infestation than those carrying wildtype *LGS1* alleles.

Furthermore, the agar gel assay was performed on a small subset consisting of 85 individuals of the PP37 population and was used to perform at GWAS to determine the genomic regions that influence *S. hermonthica* germination. One and only one significant peak was detected with the LD block of *LGS1*. Detection of this peak with such a low sample size demonstrates the power of the PP37 population to detect genomic regions associated with a phenotype of interest even when the population size is small if the allele frequency permits. The two peak SNPs of equal significance were respectively 444 kb and 44 kb from *LGS1*. We validate that this peak is, in fact, *LGS1* signal because when the *LGS1* marker was used as a co-variate, the peak is no longer detected. The *lgs1* alleles are structural variants involving deletion of all or part of *LGS1*, often extending to the surrounding genes. In the sorghum reference genome BTx623 that was used for SNP calling in this study, *LGS1* is present. Therefore, all homozygous *lgs1* accessions in the PP37 MAGIC population would have missing data in this region. Overall this finding reinforces the importance of considering candidate genes in long-range LD as the GWAS peak may be linked to a structural genome rearrangement such as large-scale insertions or deletions that confer the phenotype of interest.

2.5.2 The challenge remains phenotyping *Striga hermonthica* resistance

We demonstrated that lines in the PP37 MAGIC population fixed for *lgs1* alleles have significantly fewer events of successful *S. hermonthica* parasitism as measured in this study than those with wildtype alleles at this locus. However, when we used the same *S. hermonthica* count phenotypes and the associated derived phenotypes to conduct a GWAS an insignificant subthreshold peak with an FDR-corrected p-value of 0.34 appeared on chromosome five near the genomic position of *LGS1* but was likely only noticeable because of our *a priori* knowledge of the gene. Therefore, although *lgs1* significantly reduces *S. hermonthica* parasitism, it is a small effect

gene on the overall highly quantitative resistance. In the first year of this experiment a peak, ST-SH3.1, associated with *S. hermonthica* count and progression was detected on chromosome three with an FDR-corrected p-value of 0.07. The ST-SH3.1 peak resides within a previously mapped QTL, QSTRG3.1. The bi-parental mapping population used to locate QSTRG3.1 utilized the sorghum accession IS9830 as the *S. hermonthica* the resistant parent. IS89830 carries the *lgs1lgs1* genotype but appears to have a region on chromosome three with potentially large effect gene action at reducing *S. hermonthica* parasitism. Candidate genes within the LD block of ST-SH3.1 include a resistance related receptor-like kinase (*Sobic.003G096900*). These preliminary results were the primary driver in the decision to repeat and expand the experiment in the following 2017 season.

During the 2016 and 2017 growing seasons the field experiment was designed such that the plots were planted at twice the normal 75 centimeters spacing such that *S. hermonthica* germination stimulant released from one genotype would not affect the rate of parasite germination in the neighboring plot. Plants in the experimental check plots were densely spaced such that each block had the susceptible check to serve as an indicator of the presence of *S. hermonthica.* In 2016, Dabar, a commonly used susceptible sorghum accession was used as a susceptible check. However, in northwestern Ethiopia we found that Dabar is not highly susceptible to the resident *S. hermonthica* population. Dabar also has a duration and flowering time that far exceeded those of members of the PP37 MAGIC population. Therefore, in 2017 the experiment was repeated using double the number of replications and double the number of blocks within each replication. Furthermore, the accession BTx623 was used as the susceptible check because it is highly susceptible to *S. hermonthica*, yet highly productive when there is no parasite in this environment and had a similar duration to the population at large. These adjustments in 2017 permitted a greater resolution of *S. hermonthica* infestations and allowed blocks with insufficient parasite infestations to be removed from further analysis.

In addition to the challenges imposed by environmental variation in *S. hermonthica* infestation is the fact that this research was conducted in a developing country where variation was introduced into the experiment that could otherwise be controlled by technology applications. For example, mechanical planters have not been introduced in this region of the country. The nearly 8,200 plots that encompassed this experiment were demarcated, planted and inoculated by hand, all tasks that could easily be done with one machine over the course of a couple days. However,

in this study, these tasks were all done by hand and with daily monsoon rainfall and scorching temperatures, only 600 plots could be planted per day, even with a crew of 15 incredibly strong people. These hand activities introduced significant temporal variation that may have confounded our ability to resolve the genetic underpinnings of *S. hermonthica* resistance in the GWAS presented here. This is completely anecdotal but farmers in this region plant sorghum late in the available planting season, nearly a month after the onset of monsoon rains because they have observed delayed planting reduces *S. hermonthica* infestation levels. We still do not have a complete understanding how variation in planting time due to our technical limitations impacted the severity and uniformity of *S. hermonthica* parasitism and our ability to resolve genotype by environment interactions.

Regardless, the field-testing location at Humera in the state of Tigray in the northwestern corner of Ethiopia is potentially ideal for large quantitative genetics-based studies on *Striga*. The soil type is uniform, with limited changes in topography over large expanses. *S. hermonthica* is endemic and severe in Humera. Furthermore, there is an active and flourishing research station, HARC, that is supported by the federal and state governmental agricultural agencies in Ethiopia. The limitations that we experienced in this study might be mitigated by investments in technology in this region and further human capacity efforts to improve the quality of plant science research at this hub of *S. hermonthica* activity. Technological investments worth considering are mechanical planters, biomass dryers, and high-throughput phenotyping systems, perhaps even aerial imaging to quantify *S. hermonthica* parasitism. These efforts will greatly assist in reducing the environmental variation in the experiments so that greater resolution of the genetic architecture of *S. hermonthica* resistance can be obtained.

Further GWAS on the PP37 MAGIC population are also warranted. The PP37 MAGIC is a powerful population that displays visual phenotypic variation on the order of an association panel. Moreover, it is potentially packed with alleles impacting response to *S. hermonthica* since it was constructed by recombination among *S. hermonthica* resistant sorghum accessions with different putative mechanisms of resistance. The frequency of *S. hermonthica* resistance may be too low in other populations that would otherwise be suitable for GWAS, like the Sorghum Association Panel, where the frequency of resistant individuals is so low that it does not permit genetic resolution. Above all, we have learned that the MAGIC population is extremely powerful, when combined with precise phenotyping. With the aforementioned considerations in mind further research can be

conducted to better understand the highly quantitative nature of *S. hermonthica* resistance in sorghum and even potentially in other afflicted crops such as maize, rice and sugarcane. *Striga hermonthica* remains a major constraint to production by developing country farmers and swift action towards fixing this situation is as pertinent as ever.

Table 2.1: The founding accessions of the PP37 MAGIC population

Pedigree information is not available (-) for all accessions. The reported field reaction is taken from the associated reference that was available prior to the conception of the population. The S. hermonthica counts are reported on a per plant basis collected over three replicates in Northeastern Ethiopia in 2017. The country of origin information was collected from the U.S. National Plant Germplasm System.

Accession	Pedigree	Reported *Striga* Field Reaction	*Striga hermonthica* Count	Std Dev	LGS Marker Status	Country of Origin	Reference
555	-	Resistant	4	2	*lgs*	Yemen	(Rao, 1985)
CSH1	-	Suceptible	.	.	.	India	(Maiti, 1984)
Dobbs	-	Resistant -Tolerant	9	9	LGS	Uganda	(Doggett, 1965)
Framida	-	Resistant	17	5	*lgs*	Kenya	(Doggett, 1965)
IS9830	-	Resistant	9	12	*lgs*	Sudan	(Ramaiah, 1983)
IS4202	-	Resistant	.	.	.	India	(Maiti, 1984)
M90360	-	Tolerant	17	15	LGS	.	(Jain & Ejeta, 1981)
M90362	-	Tolerant	10	9	LGS	.	(Jain & Ejeta, 1981)
N13	-	Resistant	9	8	LGS	India	(Maiti, 1984)
P967059	-	Tolerant	10	2	*lgs*	.	(Jain & Ejeta, 1981)
P967083	-	Resistant	9	1	LGS	USA	(Jain & Ejeta, 1981)
SAR 1	(555 x 168)-1-1	Resistant	7	4	*lgs*	India	(Rao, 1983)
SAR 2	(555 x 168)-16	Resistant	11	6	*lgs*	India	(Rao, 1983)
SAR 9	[SRN 4841 x (WABC x P 3)-3]-7-3	Resistant	9	11	*lgs*	India	(Rao, 1983)
SAR 10	[555 x (PD 3-1-11 x CSV 4)-29-3]-5-2-1-1	Resistant	11	5	*lgs*	India	(Rao, 1983)
SAR 16	(555 x 168)-19-2-7	Resistant	8	2	*lgs*	India	(Rao, 1983)
SAR 19	(555 x GPR 168)-23-1-2	Resistant	13	12	*lgs*	India	n.a.
SAR 25	(555 x Awash 1050)-2-2-1	Resistant	11	5	LGS	India	n.a.
SAR 29	(SRN 4841 x SPV 104)-17	Resistant	7	3	LGS	India	n.a.
SAR 34	(148 x Framida)-39-2-4-1-2-1	Resistant	20	12	*lgs*	India	n.a.
SAR 35	(GPR148 x 555)-29-3-2-1-1	Resistant	15	8	*lgs*	India	n.a.
Serena	-	Resistant	13	6	LGS	Zimbabwe	(El Hiweris, 1987)
SRN39	-	Resistant - Tolerant	4	4	*lgs*	.	(Hess, 1992)
Swarna	-	Suceptible	.	.	.	India	(Maiti, 1984)
Tetron	-	Resistant	6	4	*lgs*	Sudan	(Ramaiah, 1983)

Table 2.2: The primers used in the LGS1 marker screening of the PP37 MAGIC population.

Forward (F) and reverse (R) primers were designed for LGS1 and a positive control.

Primer	Gene Target	Primer Sequence	GC Content	Expected Amplicon Size (bp)
Sb05G2136.0617-F	*Sobic005G213600 (LGS1)*	5'-TTGCCCAAGTCCGGCACCACCTGC-3'	67%	208 (*LGS1-1)* or 220 (*LGS1-2)*; no amplicon (*lgs1)*
Sb05G2136.0837-R		5'-GTGGCGATCAGCCGTGGAGCTC-3'	68%	
Sb05G2132.0792-F	*Sobic005G213200 (positive control)*	5'-TCGCTGCCCGGATTGCCGAAACC-3'	65%	333
Sb05G2132.1125-R		R 5'-TTCGCCTGCTCCGGCACGAGCTT-3'	65%	

Table 2.3: The phenotypic traits collected in 2016 and 2017.

The traits used to conduct the GWAS for S. hermonthica resistance in the PP37 MAGIC population. BLUPs were calculated for each trait and the ASNPC was constructed from the respective BLUPs.

Year	Trait	Unit	Description
2016	*Striga*_Count_50_DAP	count/plot	*S. hermonthica* count per plot at 50 days after planting per plot
	*Striga*_Count_75_DAP	count/plot	*S. hermonthica* count per plot at 70 days after planting per plot
	Max_*Striga*_Height	cm	Maximum *S. hermonthica* plant height in a plot
	Max_*Striga*_Branch	cm	*S. hermonthica* branch number of plant with maximum height
	*Striga*_Emerg_Dist_R	cm	Maximum *S. hermonthica* emergence distance to the right side of the plot
	*Striga*_Emerg_Dist_L	cm	Maximum *S. hermonthica* emergence distance to the left side of the plot
	*Striga*_Fresh_Weight	g/plot	Fresh weight of *S. hermonthica* at 100 days after planting
	ASNPC	n.a.	Area under the *Striga* number progress curve from 50 to 75 days after planting
2017	*Striga*_Count_40_DAP	count/plant	Mean *S. hermonthica* count per plot at 40 days after planting per plant
	*Striga*_Count_50_DAP	count/plant	Mean *S. hermonthica* count per plot at 50 days after planting per plant
	*Striga*_Count_75_DAP	count/plant	Mean *S. hermonthica* count per plot at 75 days after planting per plant
	*Striga*_Fresh_Weight		Fresh weight of *S. hermonthica* at 100 days after planting
	ASNPC_1	n.a.	Area under the *Striga* number progress curve from 40 to 50 days after planting
	ASNPC_2	n.a.	Area under the *Striga* number progress curve from 50 to 75 days after planting
	ASNPC_3	n.a.	Area under the *Striga* number progress curve from 40 to 75 days after planting

Table 2.4: Candidate gene list for ST-SH3.1

The candidate genes within a 100 kb window of the peak SNP of ST-SH3.1 on chromosome three in the 2016 GWAS for *S. hermonthica* count at 75 days after planting.

Gene Name	Chr	Start	End	Annotation	Distance to Peak SNP
Sobic.003G096000	3	8474356	8478421	fructose-bisphospate aldolase isozyme, putative, expressed	90382
Sobic.003G096100	3	8499492	8501691	auxin-induced protein 5NG4, putative, expressed	65246
Sobic.003G096300	3	8516801	8523922	homeobox and START domains containing protein, putative, expressed	47937
Sobic.003G096400	3	8533586	8537784	protein kinase domain containing protein, expressed	31152
Sobic.003G096500	3	8532845	8541635	protein kinase domain containing protein, expressed	31893
Sobic.003G096600	3	8554330	8556966		10408
Sobic.003G096750	3	8562424	8564918	protein kinase domain containing protein, expressed	2314
Sobic.003G096900	3	8573857	8577124	resistance-related receptor-like kinase, putative, expressed	-9119
Sobic.003G096950	3	8578937	8579697	lectin protein kinase family protein, putative, expressed	-14199
Sobic.003G097000	3	8580681	8582674	expressed protein	-15943
Sobic.003G097100	3	8588949	8592440	receptor kinase ORK10, putative, expressed	-24211
Sobic.003G097200	3	8585283	8588728	Ser/Thr receptor-like kinase, putative, expressed	-20545
Sobic.003G097300	3	8593915	8595519		-29177
Sobic.003G097400	3	8601835	8602722		-37097
Sobic.003G097500	3	8604843	8607611		-40105
Sobic.003G097600	3	8607613	8608841		-42875
Sobic.003G097701	3	8611635	8612475		-46897
Sobic.003G097800	3	8617949	8619145		-53211
Sobic.003G097900	3	8626250	8627845	seven in absentia protein family domain containing protein, expressed	-61512
Sobic.003G098000	3	8664262	8668413	armadillo/beta-catenin repeat family protein, putative, expressed	-99524

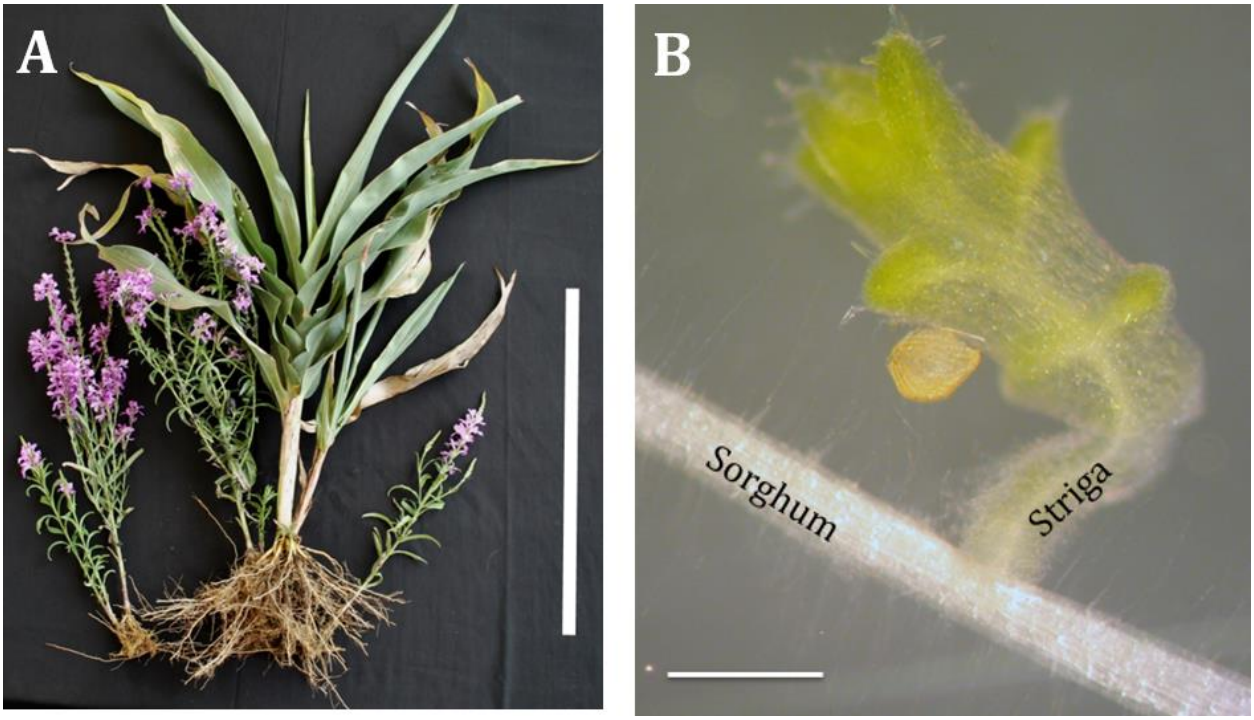

Figure 2.1: *Striga hermonthica* attachment to the host root

(A) *Striga hermonthica* parasitizes the roots of the susceptible sorghum accession BTx623. The white bar is 50 centimeters. (B) *Striga* will make a xylem to xylem connection with the host root system. A *Striga asciatica* seed is also embedded in the agar next to the established parasite. The white bar is 200 μm. (Photo B Credit: Dr. Patrick Rich)

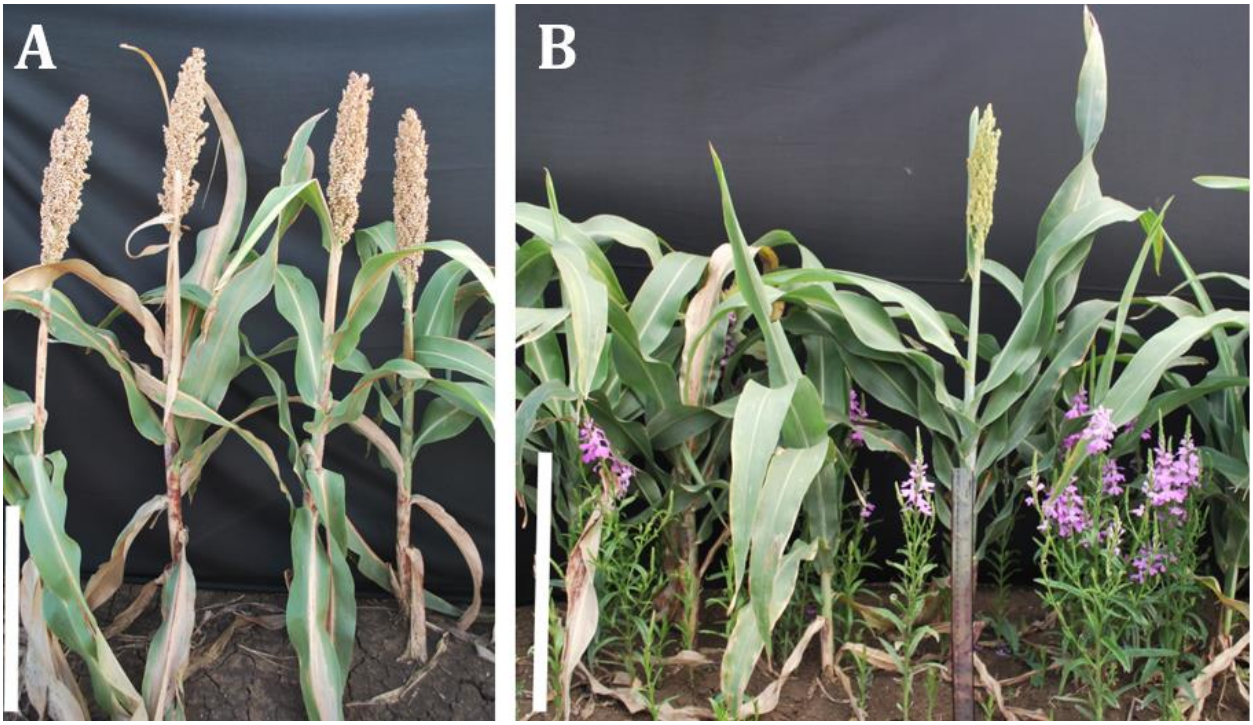

Figure 2.2: The effect of *S. hermonthica* on BTx623.

The susceptible sorghum accession BTx623 grown in *S. hermonthica* free soil (A) and *S. hermonthica* infested conditions (B) in Humera, Ethiopia. The sorghum in the infested condition was planted 12 days after the uninfested condition. The white bar is 0.5 meters.

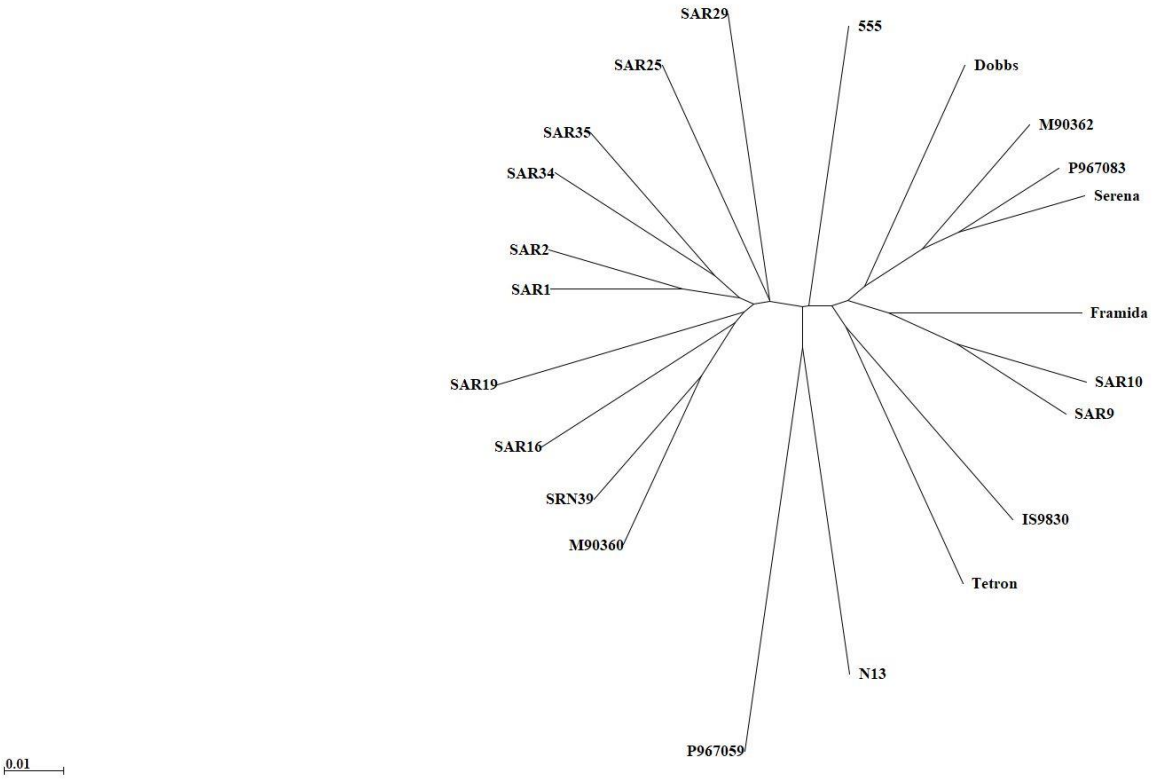

Figure 2.3: Population structure of the PP37 MAGIC founders.

Phylogenetic tree of the PP37 MAGIC population founders.

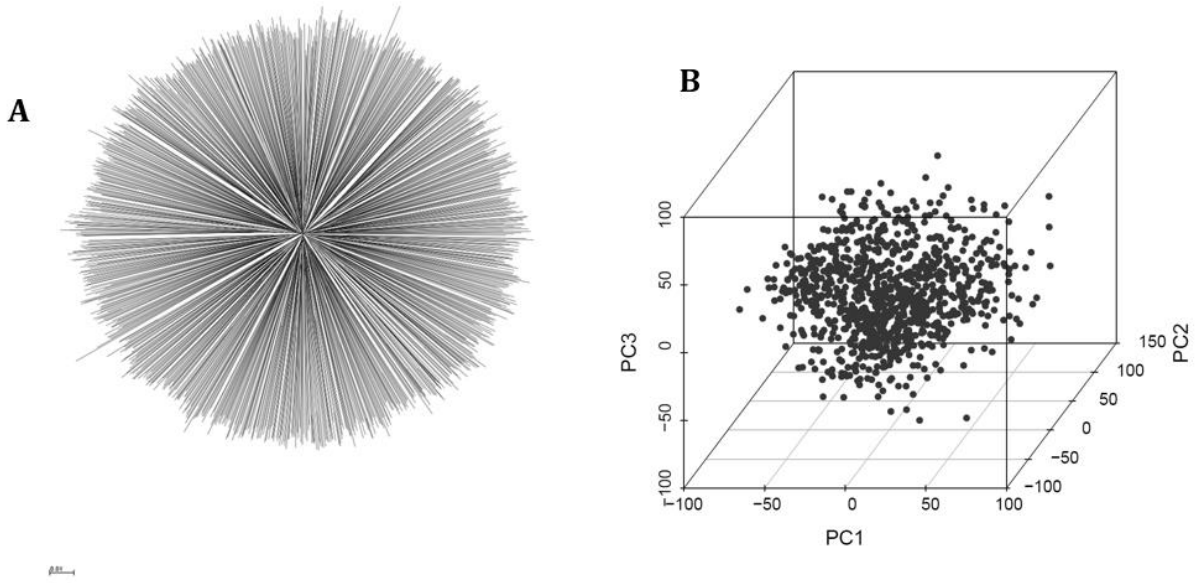

Figure 2.4: Population structure of the PP37 MAGIC population.

The (A) phylogenetic tree and (B) plot of principal components of the PP37 MAGIC population, illustrating a lack of population structure.

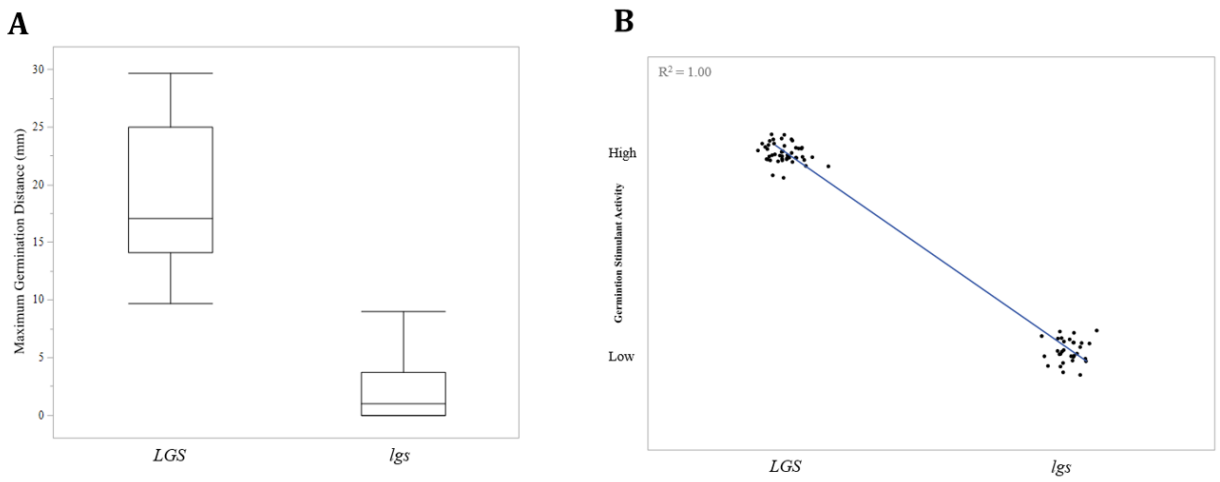

Figure 2.5: Germination stimulant activity of the PP37 MAGIC population.

(A) The average maximum germination distance (mm) of a subset (n=85) of the PP37 MAGIC population. (B) The germination stimulant activity score of the same subset. In both figures the x-axis is the *LGS1* marker status with *LGS* being the wild type and *lgs* being the natural mutant genotype that confers low *S. hermonthica* germination stimulant activity.

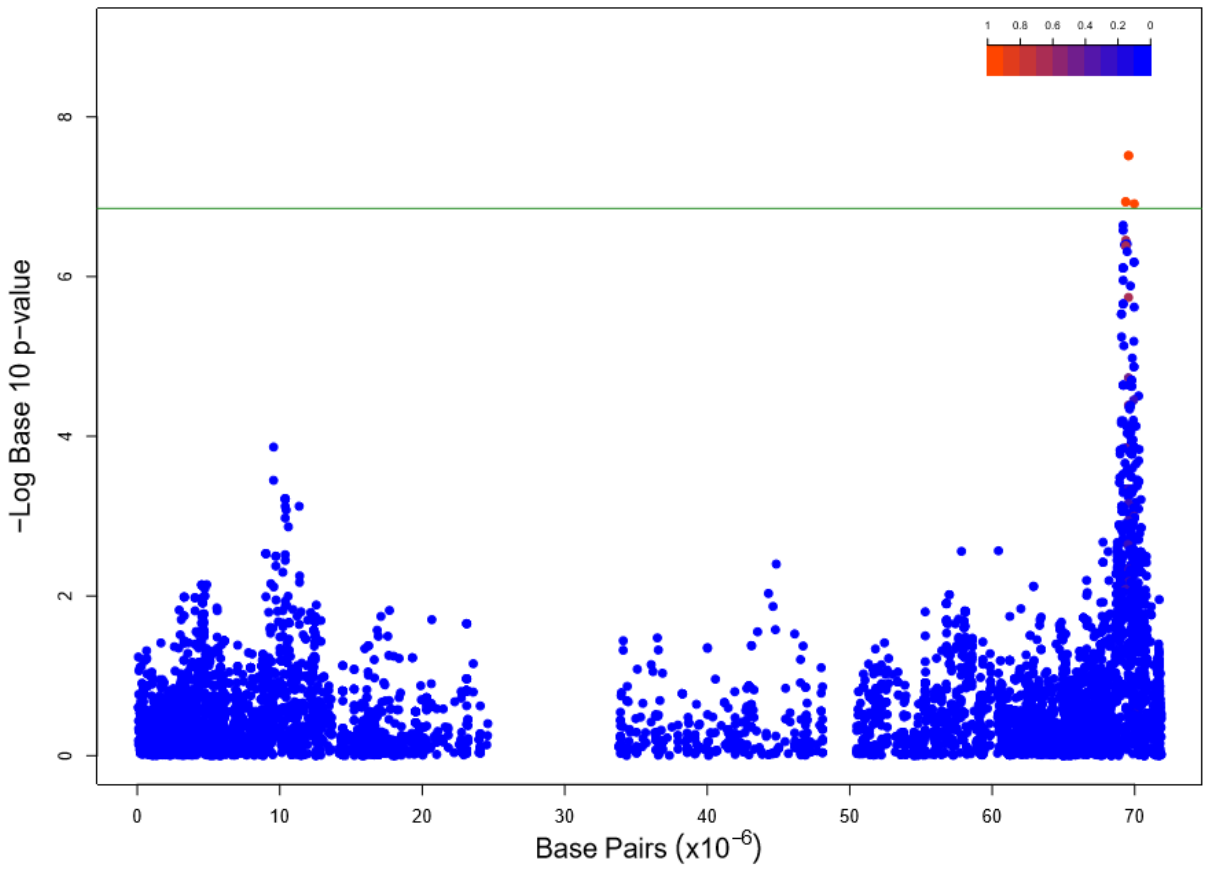

Figure 2.6: A Manhattan plot of the maximum germination distance score in a subset of the PP37 MAGIC population (n=85).

The x-axis is the genomic position of the SNPs on chromosome five, and the y-axis is the negative log base 10 of the p-values. The green line indicates the Bonferroni threshold for statistical significance. The SNP color indicates the correlation of the respective markers.

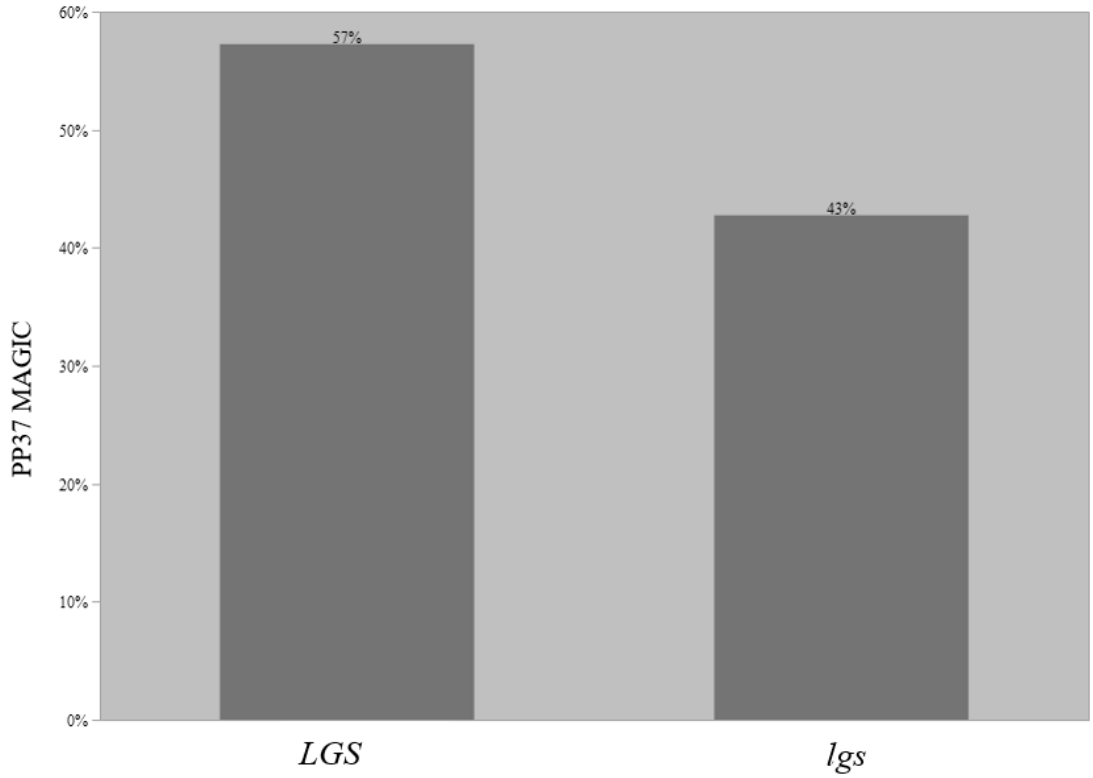

Figure 2.7: Allelic distribution of *LGS1* in the PP37 MAGIC population.

The y-axis represents the percentage of individuals within the whole PP37 population. The x-axis is the *LGS1* marker status with *LGS* being the wild type and *lgs* being the natural mutant genotype that confers low *S. hermonthica* germination stimulant activity.

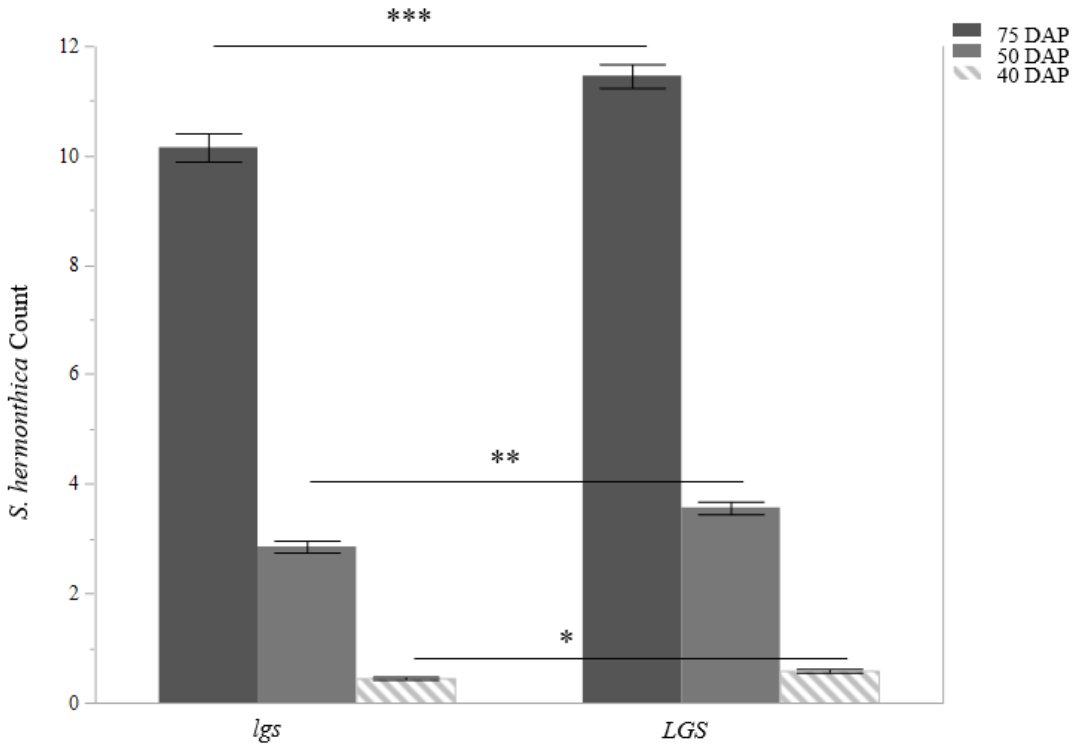

Figure 2.8: The effect of *LGS1* on the *S. hermonthica* count.

The mean *S. hermonthica* count per sorghum plant at 40, 50 and 75 days after planting across the entire PP37 MAGIC population (n=916). The low and high stimulant grouping is determined by genotyping respectively for the absence or presence of the *LGS1* molecular marker. The *S. hermonthica* count at 40, 50 and 75 days after planting is significantly different between the high and low stimulant genotypes at a significance level of < 0.01, <0.001 and <0.0001, respectively. The error bars were constructed using one standard error of the mean.

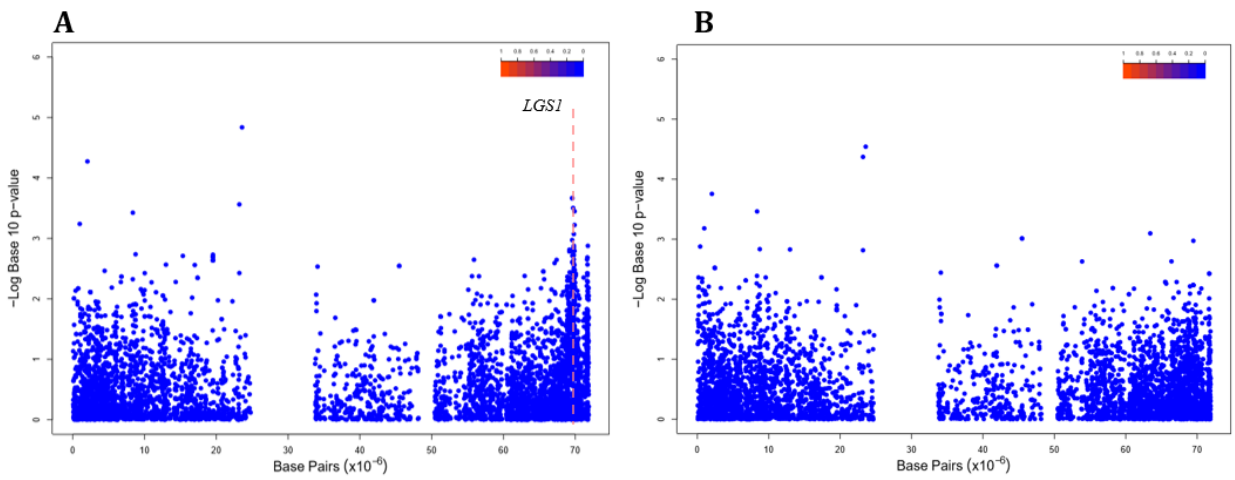

Figure 2.9: Detecting *LGS1* in the GWAS

Manhattan plot of the 2017 GWAS for the ANSPC phenotype in the PP37 MAGIC using the K-model (A) and the K-model with the *LGS1* marker as a covariate (B). The red vertical line indicates the position of *LGS1*. The x-axis is the –log10 p-value and the x-axis is the base pair position on chromosome five.

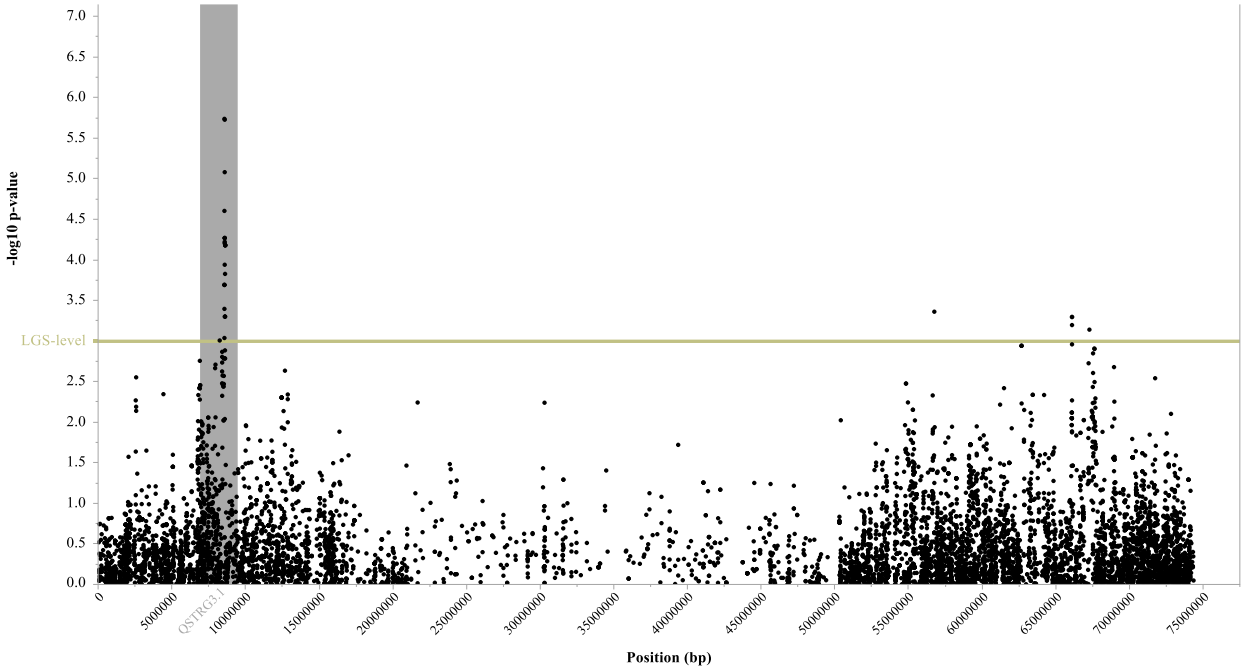

Figure 2.10: A novel GWAS peak for *S. hermonthica* resistance

Manhattan plot of the 2016 GWAS for the ANSPC phenotype in the PP37 MAGIC with *LGS1* marker as a covariate. The gold horizontal line indicates the position of significant level of the peak SNP linked to *LGS1* detected without the covariate. The SNPs crossing the *LGS1*-level are known as ST-SH3.1 and it co-localizes with the previously detected QTL, QSTRG3.1 that is indicated by the grey bar. The x-axis is the $-\log_{10}$ p-value and the x-axis is the base pair position on chromosome three.

CHAPTER 3. EXPLORING THE GENETIC ARCHITECTURE OF PLANT ARCHITECTURE IN SORGHUM

3.1 Abstract

Sorghum is the fifth most important cereal crop in the world. Depending on the end user the plant architecture needs of sorghum are quite diverse. In most cases where the sorghum grain is the component of interest there is a global demand for sorghum that is short in stature to facilitate mechanical harvest. Whereas when sorghum biomass is preferred for fodder or ethanol, often sorghum that is tall in stature is preferred. The genetic mechanisms that influence sorghum plant architecture are only partially understood. The objective of this research was to improve our understanding of plant architecture in sorghum. To that end, the PP37 multi-parent advanced generation inter-cross (MAGIC) population was developed, originally as a recurrent selection population that was developed to recombine sorghum accessions with different putative resistance mechanisms to *S. hermonthica*. The plant architecture of the PP37 MAGIC was also assessed at a location in Northwestern Ethiopia that is free of the parasite, as it significantly reduces plant height. To asses plant architecture the total plant height, the height of the panicle base, flag leaf height, and pre-flag leaf height were collected using a relatively high-throughput barcoded measurement system. Sorghum head exertion and panicle length were derived from this data. The measures of plant architecture and the derived traits were used to conduct a genome-wide association study. Highly significant peaks were detected that resolved the *dwarf-3* locus and an uncharacterized locus, qHT7.1 that had only been previously resolved using near-isogenic lines. Furthermore, a novel significant locus was associated with exertion on chromosome 1. The random mating that was utilized to develop the PP37 MAGIC has broken all population structure and subsequently broke ancestral linkages that limited resolution previously. As a result, novel candidate gene lists have been developed as an outcome of this research that refines the potential genes that need to be explored to validate qHT7.1 and the novel association on chromosome 1.

3.2 Introduction

Sorghum is a unique agronomic crop because it has a wide range of end uses. Traditionally we know sorghum as the fifth most important cereal crop globally, as it is a staple food crop in a number of developing countries (FAO 2019). Within those developing countries sorghum has a number of competing uses, such as for the brewing industry, animal feed both as grain and fodder and is used in the construction of traditional homes. In the developed countries sorghum may be utilized in a similar manner, aside from construction but is increasingly being regarded as a potential biomass crop for ethanol production. In most cases where the sorghum grain is the component of interest, there is a global demand for sorghum that is short in stature to facilitate mechanical harvest. Whereas when the sorghum biomass is preferred for fodder or ethanol, often sorghum that is tall in stature is preferred (Regassa and Wortmann 2014). Here plant architecture is defined as the organization and dimensions of the plant structure, here focusing on the above ground structures. In addition to consumer preference, certain aspects of plant architecture such as exertion can have serious impacts on the agronomic performance and quality of sorghum. Exertion is the ability of the sorghum head to emerge completely from the flag leaf. Poor exertion results in the flag leaf partially covering the sorghum head or inflorescence, and this promotes pathogen growth that deteriorates the grain quality. Our current understanding of plant architecture is limited. Therefore, there remains a need to better understand the genetic variation that influences plant architecture in order to facilitate current and future plant breeding needs.

Generally, the sorghum landraces are tall as they were utilized for a fodder, construction, and grain. Selection for dwarfness occurred when the sorghum landraces were introduced into environments where it was advantageous. The dwarfing phenotypes in modern sorghum were thought to be controlled for four dwarfing loci, *Dw1-Dw4* (Quinby and Karper 1954; Quinby 1974). *Dw1* was mapped to a 300 kb region on chromosome 9 (Brown et al. 2008) in an association mapping study. It was later determined that the causative gene in this region was Sobic.009G229800 which encodes a putative membrane protein that regulates internode cell proliferation (J. Hilley et al. 2016; Yamaguchi et al. 2016). The seminal mapping study of *Dw2* indicated that the causative gene was located on chromosome 6 and was linked to a locus the influences maturity, *Ma1* (Lin, Schertz, and Patemon 1995). The causative gene at the *Dw2* locus was determined to be Sobic.006G067700, as a protein kinase whose closest *Arabidopsis thaliana*

homolog is the kinesin-like calmodulin-binding protein interacting protein kinase (J. L. Hilley et al. 2017). *Dw3* was the first dwarfing gene to be cloned and it encodes Sobic.007G163800.1 an unstable phosphoglycoprotein auxin efflux carrier that orthologous to brachytic2 (*br2*) in maize (Multani et al. 2003). An additional locus known as qHT7.1 that is linked to *Dw3* has been resolved and it has an opposite impact on plant height as the neighboring dwarfing locus (Li et al. 2015). To date, the gene underlying qHT7.1 has not been identified but it does reside within a large region that syntenic to maize at the locus surrounding *br2*, the ortholog of *dw3*. Furthermore, an additional locus was detected on chromosome 1 in a genome-wide association study, but the causative gene remains to be determined (Upadhyaya et al. 2013).

The objective of this research was to improve our understanding of plant architecture in sorghum. It is thought that sorghum originated in the Northwestern regions of Ethiopia, spanning into Sudan (Ejeta, 2007). In this study, the PP37 multi-parent advanced generation inter-cross (PP37 MAGIC) population was phenotyped for plant architecture in Northwestern Ethiopia for two consecutive growing seasons starting in 2016. The phenotypic measurements include the total height of the plant, the height of the base of the panicle, the flag leaf height, and the pre-flag leaf height. Measurements of the panicle length and exertion were derived from these measurements. Using these phenotypes, a genome-wide association study (GWAS) was conducted to determine the regions of the genome that impact the various impacts of plant architecture. We hypothesized that the large population size and extensive population size coupled with accurate measures of plant architecture would allow us to further resolve the genomic regions influencing these traits. Here we validate the previously known role of *dw3* and qHT7.1, but further, demonstrate that only qHT7.1 plays a role in exertion. Furthermore, we identify novel regions of chromosome 1 that influence the panicle length and exertion that do not co-localize with any previously characterized dwarfing genes.

3.3 Materials and Methods

3.3.1 Germplasm

The PP37 MAGIC population was developed at the Purdue University Center for Research and Education, Purdue University beginning in 1986. The population was originally developed as a recurrent selection population for the development of elite sorghum cultivars with potentially

multiple mechanisms of *S. hermonthica* resistance. The PP37 MAGIC population was constructed using 25 founder sorghum accessions. The founders were randomly mated for six generations using the *ms*$_3$ genetic male sterility system (Doggett and Eberhart 1968). The genetic male sterility system involved crossing the founders to a stock of 10 random sorghum plants that carried the *ms*$_3$ allele. During the random mating, only male-sterile heads are harvested which ensures that all seeds produced on the head are the result of cross-pollination. After random mating, a total of 1,200 inbred lines were randomly selected and further self-pollinated by single seed descent to the S_6 generation with a final population size of 1,000 individuals.

3.3.2 Genotypic data

Tissue was collected from one seven-day old sorghum seedling of each accession of the PP37 MAGIC grown under greenhouse conditions at Purdue University. Approximately 40-50 mg of lyophilized leaf tissue was used for DNA extraction. DNA was extracted using the CTAB (cetyl trimethylammonium bromide) method with minimal modifications (Saghai-Maroof et al. 1984). After elution, the DNA was cleaned using a 1.5:1 by volume ratio of Axygen Clean-Seq beads (Corning Life Sciences, Corning, NY, USA) to remove any remaining inhibitory compounds in the sample. DNA was quantified using Quant-IT PicoGreen fluorescent dye (Thermo Fisher, Waltham, MA, USA). The concentration of the DNA was quantified using the Quant-iT™ PicoGreen® dsDNA kit (Life Technologies, Grand Island, NY). Library preparation was done using a genotyping-by-sequencing (GBS) protocol with the following modifications; 50 ng of DNA was digested with a 5 basepair cutter ApeKI (New England Biolabs, Ipswich, MA) followed by ligation of barcoded adapters with T4 ligase (New England Biolabs, Ipswich, MA) (Elshire et al. 2011). The adapter-ligated samples are pooled, amplified, and adapter dimers were removed with SPRI bead purification. The quality and quantity of the finished libraries were assessed using the Agilent Bioanalyzer High Sensitivity Chip (Agilent Technologies, Inc., Santa Clara, CA) and Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY). The libraries were standardized to a 2nM concentration, and the cluster was generated using the HiSeq SR Cluster Kit v3 (Illumina Inc, San Diego, CA, USA). Each flow cell was sequenced using single-end, 100 base pair reads on the Illumina HiSeq2500 sequencer. The resulting images were analyzed using the Illumina Pipeline, version 1.8.2. To analyze the GBS data and the TASSEL-GBS pipeline was utilized (Glaubitz et al. 2014). The original genotypic data set consisted of 449,208 single

nucleotide polymorphisms (SNPs) that were filtered using TASSEL (Bradbury et al. 2007). A total of 71,448 SNPs with minor allele frequency (MAF) > 5% and missing data < 80% were used in this study. The physical position of the SNPs was determined using sorghum genome v3.1.

3.3.3 Field experimental design

The PP37 MAGIC population was grown during the 2016 and 2017 growing seasons in the Humera district in the state of Tigray located in Northwestern Ethiopia. *Striga hermonthica* is endemic in this area and parasitism can result in a significant change in plant architecture therefore, steps to mitigate parasitism were implemented. In 2016 two replications were grown solely on research fields within the Humera Agricultural Research Center (HARC) of the Tigray Agricultural Research Institute and only one replication was in the *S. hermonthica* free location and thus was used in this study. In 2017 two replications of the experiment were planted in *S. hermonthica*-free conditions at HARC. All replications consisted of plots that were 3 meters in length with 1.5 meters between the plots and 1-meter allies to prevent germination stimulant from influencing *S. hermonthica* infestations in the neighboring plot. Each replication was weeded three times during the growing such that *S. hermonthica* was the only weed growing in the experimental unit so that those plots could be identified and excluded. In 2016 each replication was grown in a randomized incomplete augmented block design, where each replication consisted of 100 blocks and each block had the susceptible check Dabar randomized among the entries. In 2017 each replication was grown in a randomized incomplete augmented block design, where each replication consisted of 200 blocks and each block had the susceptible check BTx623 randomized among the entries. The check was replaced across seasons because BTx623 is more sensitive to *S. hermonthica* and there was a better indicator of the pest.

3.3.4 Phenotyping for plant architecture

Phenotypic data were collected on the total sorghum plant height, panicle length, flag leaf height, pre-flag leaf. Exertion was derived by subtracting the height of the base of the panicle from the height of the panicle. The plant height measurements were collected using a barcoding system in which a strip of barcodes was printed to represent 1 cm intervals. The data were recorded electronically with a barcode reader that sent a data to a tablet. One randomly selected plant per plot was utilized to collect the plant height measurements. Best linear unbiased predictors (BLUP)

were calculated for each trait using the filtered phenotypic data for one replication in 2016 and two replications in 2017. The BLUPs from 2016 represents a single replication whereas the experimental unit had to be free of *S. hermonthica* parasitism. BLUPs were obtained using all the data collected over the two years. The BLUPs were calculated from the following mixed linear model for each trait

$$y_{ij} = \mu + \tau_i + \beta_{j(\rho)} + \varepsilon_{ijk}$$

in which y_{ij} is the ijth observation, μ is the grand mean, τ_i is ith fixed genotypic effect and $\beta_{j(\rho)}$ is the block effect nested within replication ρ.

3.3.5 Population structure and association analysis

The phylogeny of the PP37 parents and population was assessed using the neighbor joining-tree method TASSEL (Bradbury et al. 2007). This was done to determine the clustering of the 23 founders in which genotypic data were available for compared to the whole population. The population structure of PP37 was further inferred using a principal component analysis (PCA) to determine if there was any genetic clustering. The PCA and GWAS were conducted using the statistical genetics package Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al. 2012; Tang et al. 2016). The GBS produced SNPs with a MAF of less than 0.05 and more than 20% were removed from the genotypic data set, and minor SNP states or minor alleles were also removed, resulting in 71,448 SNPs. Associations were made using a unified mixed linear (MLM) approach with kinship to account for the relatedness among individuals (Yu et al. 2006), The PCA analysis did not show any genotypic clustering therefore population structure was not controlled for in the subsequent GWAS model. Multiple testing was controlled using the Benjamini and Hocheberg (1995) procedure to control the false discovery rate at 5%. The squared Pearson's correlation coefficient between SNPs was calculated in TASSEL 5.2 to estimate linkage disequilibrium (LD) (Bradbury et al. 2007) and local linkage black was determined when $r^2 < 0.2$. There were 403 and 916 individuals from the PP37 MAGIC that were grown under sufficient infestation to be represented in the 2016 and 2017 respectively included in the GWAS. To further determine if the peak were inside of syntenic

3.4 Results and Discussion

3.4.1 Structure of the PP37 population

In generating the PP37 MAGIC population, the founders were randomly mated for six generations using the ms3 genetic male sterility system. The phylogeny of the PP37 population founders was analyzed using neighbor-joining trees and the 23 founders that were analyzed clustered into seven distinct groups (Figure 1). Similarly, the phylogeny of the entire PP37 MAGIC population was analyzed and only clustered into one group and the lines are distributed at equal distances from the root of the tree (Figure 2A), suggesting that the population structure present in the founders were eliminated during the successive random mating that occurred during the development of the population. The lack of population structure on the PP37 MAGIC is further demonstrated with principal component analysis, in which the individuals did not subdivide into distinct clusters (Figure 2B). Spurious genotype to phenotype associations can arise when population structure is present and not controlled for. However, controlling for population structure reduces power to detect genotype to phenotype associations. The lack of clustering in the PP37 population was inferred as a lack of population structure in the PP37 MAGIC population, it was not necessary to control for it in any of the models used for the GWAS.

3.4.2 Exertion is not associated with *dw3*

A GWAS was conducted for the plant architecture data collected on the PP37 MAGIC. The known sorghum gene *dw3* resides on chromosome 7 and has been previously mapped and characterized. A bi-parental mapping study was able to resolve *dw3* from a previously unmapped locus known as qHT7.1 (Li et al. 2015). To date the causative gene associated with qHT7.1 is unknown. The biparental mapping effort reported that *dw3* is associated with the changes in plant architecture below the flag-leaf and qHT7.1 only controlling plant architecture above the flag-leaf (Li et al. 2015). In the research presented here, we detected two highly significant associations that resolved the linkage blocks between qHT7.1. It was demonstrated that in the PP37 MAGIC population qHT7.1 had significant associations with plant height, exertion, flag leaf height and the pre-flag leaf height (Figure 3.3A-D). Furthermore, in the PP37 MAGIC population, it was found that *dw3* has highly significant associations to plant height (Figure 3.3A), flag leaf height (Figure 3.3C) and pre-flag leaf height (Figure 3.3D). It was also demonstrated that *dw3* does not influence sorghum head exertion but qHT7.1 does. This finding may facilitate the genetic improvement of

sorghum by producing sorghum hybrids that have panicles consistently emerging from the flag leaf, which will produce higher quality grain. This evidence indicates that qHT7.1 and *dw3* may have a more complex effect on plant architecture than previously illustrated in the bi-parental mapping population. It is possible that there are additional alleles of qHT7.1 and *dw3* segregating from one or more of the 25 founds of the PP37 population. Further studies that sequence the full-length *dw3* in the population founders, and the yet to be determined causative gene for qHT7.1would facilitate testing the proposed hypothesis.

Given that the causative gene for qHT7.1 has not been determined to date it was justified to identify the candidate genes within the linkage block surrounding the peak SNP at this locus (Table 3.1). As previously demonstrated *dw3* is an ortholog of *br2* in maize (Multani et al. 2003). In fact, comparative genomic analysis reveals that qHT7.1 and *dw3* reside within a large syntetic region that is shared between chromosome 7 of sorghum and chromosome 1 of maize (Figure 3.4). The peak SNP of qHT7.1 resides within the Sobic.007G137101 (SbMYB), which is a MYB family transcription factor orthologous to Zm00008a003767 (ZmMYB). There are several biologically intuitive candidate genes within the 100kb linkage block surrounding the peak SNP at qHT7.1, including other MYBs. However, because of the strength of the association and that the SNP resides within the gene it may be worth prioritizing this first within sorghum and maize given the syntenic relationship (Figure 3.5).

3.4.3 Novel locus associated with exertion and panicle length

In addition to validating and expanding our knowledge on previously detected associations, this GWAS on plant architecture in the PP37 population also mapped novel associations. A novel locus was significantly associated with panicle length (Figure 3.6A) and exertion (Figure 3.6B) on chromosome one. This locus is not linked to a locus that influences the maturity or flowering time. A candidate gene list has been produced for the 200kb region that surrounds this novel locus on chromosome 1. The panicle length and exertion are both derived traits and the data used to generate them is not independent. Therefore, it is not altogether surprising that they map to the same location. However, as demonstrated with *dw3* in the PP37 MAGIC population, it is possible for traits that are not entirely independent to associate with a different trait.

As with most GWA studies, the results of this study do not paint a complete picture of the genetic mechanisms that influence architecture. However, we have demonstrated that the qHT7.1

and *dw3* locus could be resolved at high resolution in the PP37 MAGIC. We have evidence to suggest that qHT7.1 and *dw3* have a more complex role than previously demonstrated. We have refined the candidate gene list at qHT7.1 and further information on the syntenic relationship between sorghum and maize in this region that has remained highly conserved over evolutionary time scales. We have also provided evidence of a novel region influence the length of the panicle and the ability of that panicle to emerge from the flag leaf, which is an important trait that may be able to improve sorghum grain quantity but also quality. The desired plant architecture needs of the future remain largely unknown as globalization continues. However, we do know that variation for this trait will be utilized for different breeding targets and furthering our understanding of plant architecture remains an important research endeavor to facilitate these future selection efforts.

Table 3.1: Candidate genes for qHT7.1 associated with all plant architecture phenotypes.

The candidate genes with in the 100 kb linkage block of the peak SNP associated with plant height on chromosome 7.

Gene(v3.1.1)	Transcript Start (bp)	Transcript End (bp)	Annotation	Distance to Peak SNP (bp)	Maize Homolog (B73 v4)
Sobic.007G136300	56367592	56369617	no apical meristem protein, putative, expressed	99055	Zm00008a003774
Sobic.007G136400	56372628	56381009		94019	.
Sobic.007G136500	56380444	56382321	MYB family transcription factor, putative, expressed	86203	Zm00008a003773
Sobic.007G136600	56396557	56397894	SET domain-containing protein, putative, expressed	70090	Zm00008a003772
Sobic.007G136700	56399351	56400674	YGL010w, putative, expressed	67296	Zm00008a003771
Sobic.007G136800	56430217	56437440	polypyrimidine tract-binding protein, putative, expressed	36430	Zm00008a016215
Sobic.007G136900	56439996	56442238	chlorophyll A-B binding protein, putative, expressed	26651	Zm00008a003768
Sobic.007G137000	56445832	56450594	lipoate protein ligase, putative, expressed	20815	Zm00008a019058
Sobic.007G137100	56444373	56446237		22274	.
Sobic.007G137101	56463393	56469232	MYB family transcription factor, putative, expressed	0	Zm00008a003767
Sobic.007G137201	56494648	56495884		-28001	.
Sobic.007G137300	56497770	56501149	myb-like DNA-binding domain containing protein, expressed	-31123	Zm00008a016217
Sobic.007G137350	56508283	56511067	WD domain, G-beta repeat domain containing protein, expressed	-41636	Zm00008a027111
Sobic.007G137400	56516168	56526441	CSLA11 - cellulose synthase-like family A, expressed	-49521	Zm00008a040471
Sobic.007G137501	56528079	56529349		-61432	.
Sobic.007G137600	56567941	56570679	lactate/malate dehydrogenase, putative, expressed	-101294	Zm00008a003762

Table 3.2: Candidate genes for association novel association on chromosome 1

The candidate genes with in the 200 kb linkage block of the peak SNP association with panicle length and exertion.

Gene(v3.1.1)	Annotation	Transcript Start (bp)	Transcript End (bp)	Distance to Peak SNP (bp)	Expression in the Sorghum Inflorescence (FPKM)	Maize Homolog (B73 v4)	Maize Description
Sobic.001G222100	putative uncharacterized protein	21125541	21125846	174106	1.85532	-	-
Sobic.001G222200	similar to Os10g0472400 protein	21126170	21130232	173477	10.2568	Zm00001d014043	PPPDE peptidase domain-containing protein 2
Sobic.001G222300	similar to Cyclin-dependent kinase inhibitor 1	21176895	21180687	122752	9.8558	Zm00001d014046	Cyclin-dependent kinase inhibitor 5
Sobic.001G222400	similar to TPR Domain containing protein	21198140	21211324	101507	13.141	Zm00001d032734	no pollen germination related 2
Sobic.001G222600	similar to Cyanate hydratase	21233680	21236680	65967	23.2314	Zm00001d032736	Cyanate hydratase
Sobic.001G222700	similar to CER1 protein	21300007	21306449	-360	30.4006	Zm00001d014055	Protein ECERIFERUM 1
Sobic.001G222900	putative uncharacterized protein	21314169	21320655	-14522	110.925	Zm00001d032739	Heterogeneous nuclear ribonucleoprotein 1
						Zm00001d014058	Heterogeneous nuclear ribonucleoprotein 1
Sobic.001G223100	similar to POT family protein	21334060	21340341	-34413	1.97271	Zm00001d014060	Protein NRT1/ PTR FAMILY 5.2
						Zm00001d032740	Protein NRT1/ PTR FAMILY 5.2
Sobic.001G223200	similar to POT family protein	21379629	21384632	79982	22.5505	Zm00001d032744	Protein NRT1/ PTR FAMILY 5.2
Sobic.001G223300	similar to OJ000223_09.12 protein	21408273	21410994	-108626	0	-	-

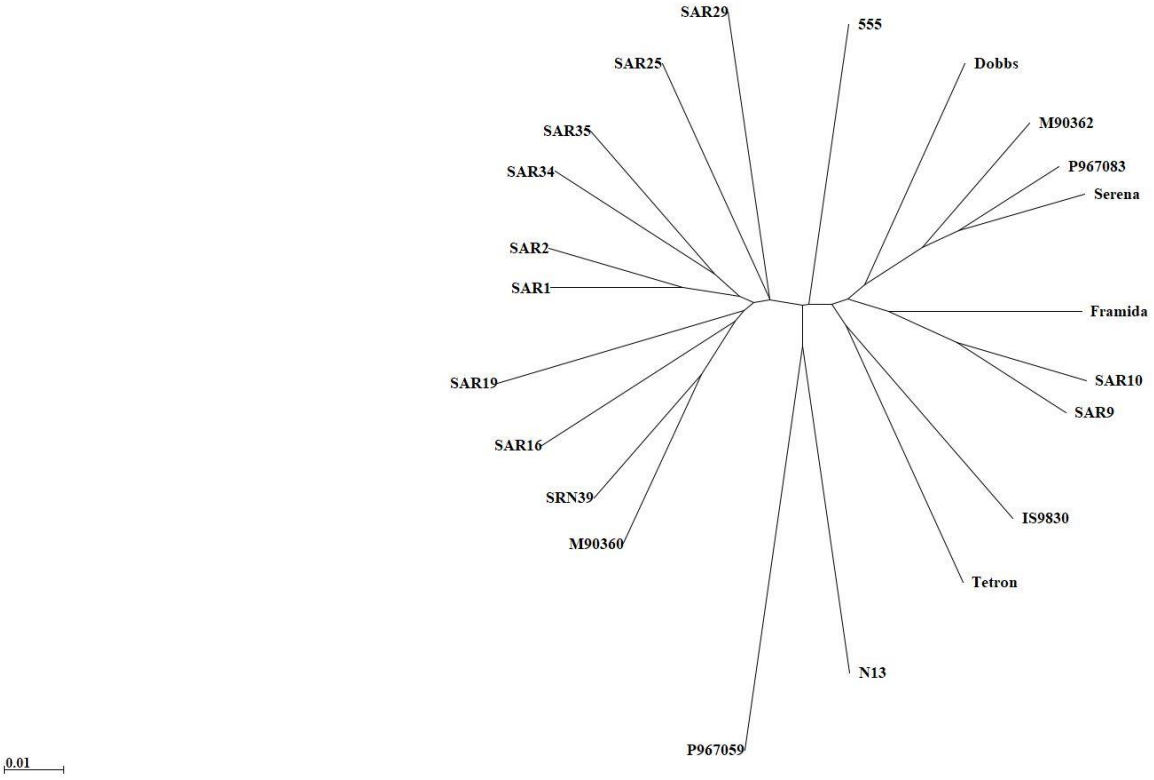

Figure 3.1: Phylogenetic tree of the PP37 MAGIC population founders.

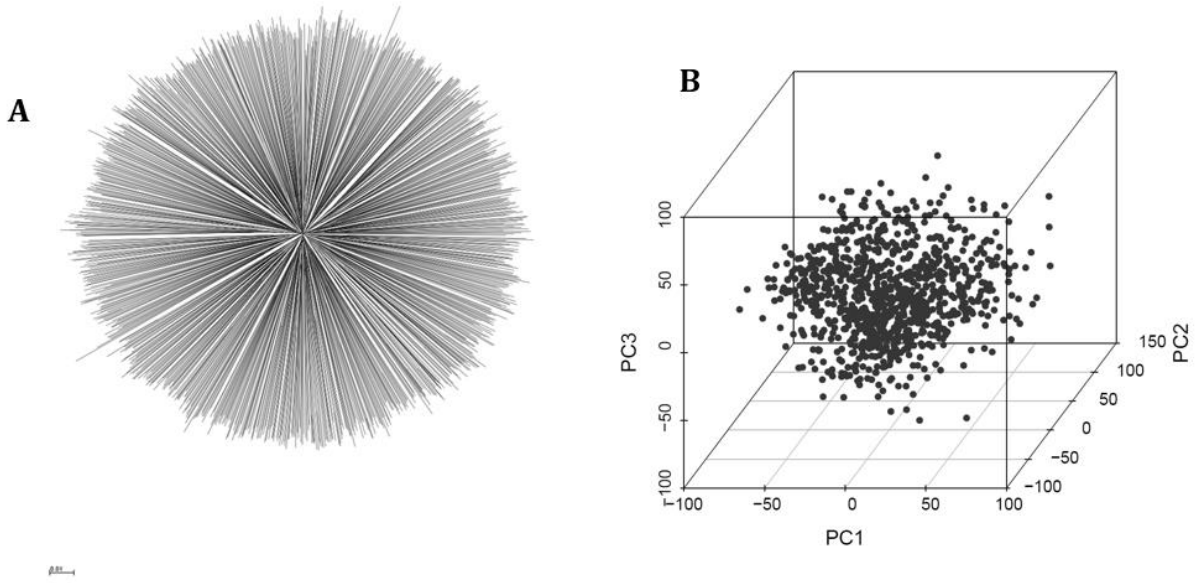

Figure 3.2: Phylogenetic tree of the PP37 MAGIC population

The (A) phylogenetic tree and (B) plot of principal components of the PP37 MAGIC population, illustrating a lack of population structure.

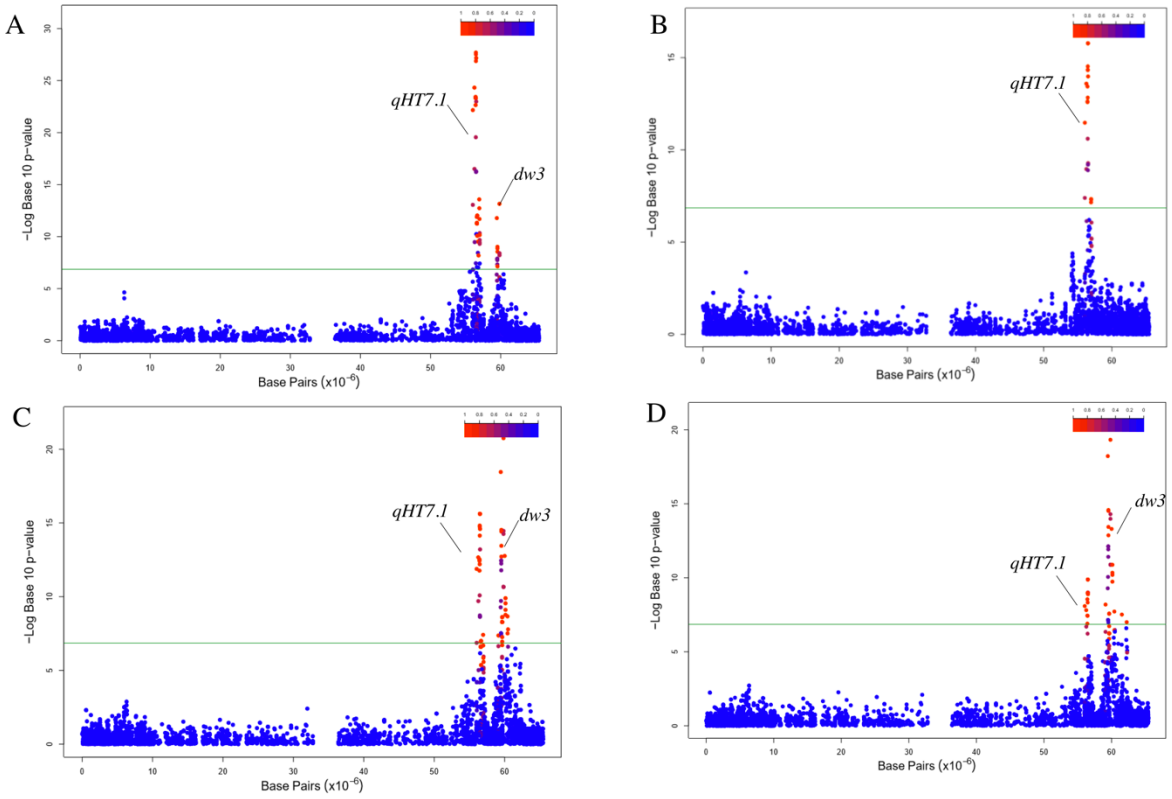

Figure 3.3: Manhattan plots for significant associations with plant architecture on chromosome 7. Significant associations on Chromosome 7 (A) Plant Height (B) Exertion (C) Flag Leaf Height (D) Pre-flag Leaf Height

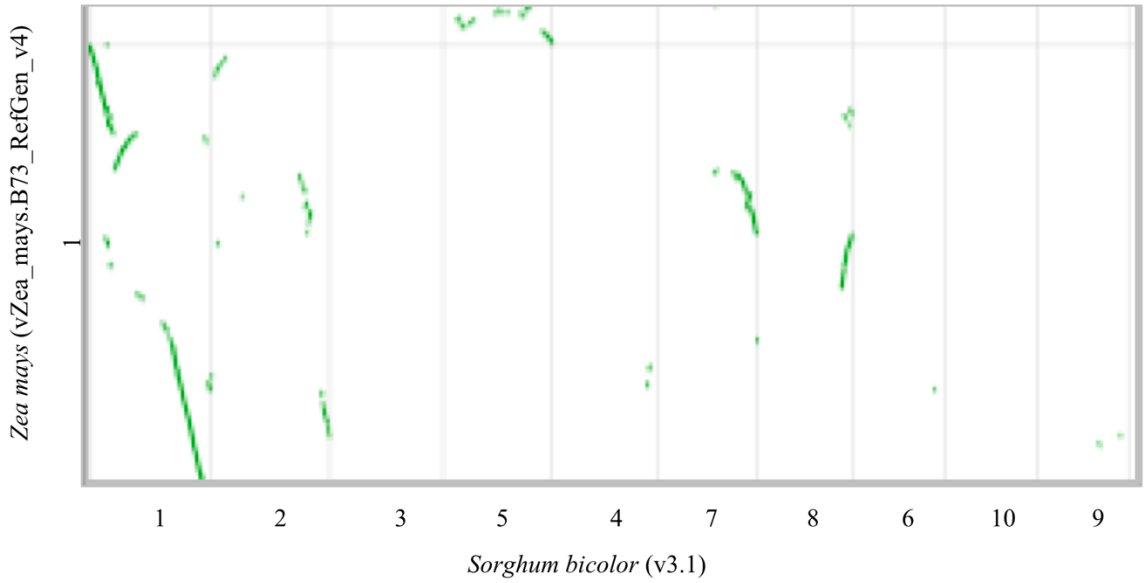

Figure 3.4: A syntenic dotplot between maize and the sorghum.

The syntenic region between maize chromosome 1 and chromosome 7 of maize contains *dw3* and the maize ortholog *br2*, and the putative candidate gene for qHT7.1

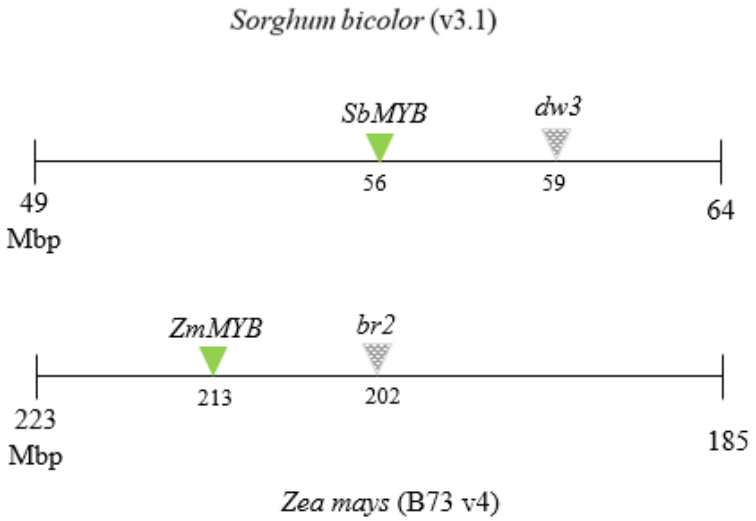

Figure 3.5: A schematic of the putative syntenic region in sorghum and maize influencing plant height.

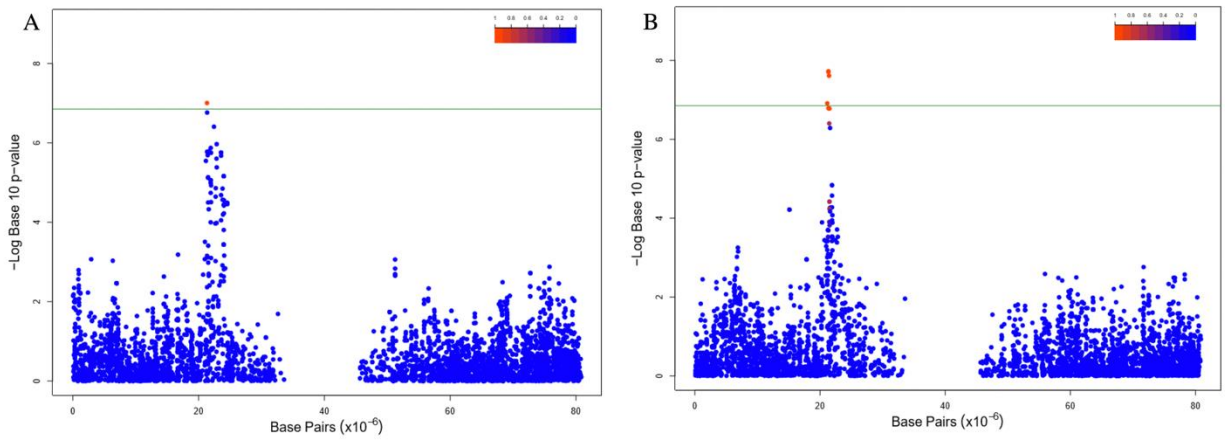

Figure 3.6: Manhattan plots for significant associations with plant architecture on chromosome 1. Significant associations on Chromosome 1 (A) Panicle Length (B) Exertion

CHAPTER 4. ACHIEVING GLOBAL FOOD SECURITY: THE BROADER IMPACTS OF PLANT SCIENCE RESEARCH

4.1 Abstract

Too often a dissertation will solely detail the research related processes and findings of our scientific endeavors. This is counter-intuitive to what those with a Ph.D. will be expected to do when they enter the academic workforce. Every grant awarded by the National Science Foundation (NSF) is judged equally on the potential of the research to possess intellectual merit and broader impacts. The intellectual merit or research is what we are predominately trained to do during our graduate education. Broader impacts are the larger societal improvements such as benefit to the public and human capacity building of our research. The NSF, and increasingly other funders require that broader impacts be designed with intent, implemented succinctly with institutional support and be well reported. Here I report the major broader impacts of my Ph.D. research that was funded by the Bill and Melinda Gates Foundation. The hope is that by sharing these broader impacts we can improve the design and implementation of these endeavors to deliver the education, resources and human capacity that is necessary to implement the findings of our research in order to achieve global food security.

4.2 Introduction

In the plant science literature, it is common for the author to introduce the need for their research by describing the cataclysmic state of global hunger and malnutrition that is occurring amid increasing global populations and climate change. It is indeed true that approximately 821 million people across the globe experienced chronic food deprivation in 2017, up from 804 million in 2016 (97). The number of malnourished people in the world has been on the rise since 2014 (Figure 4.1). Plant scientists have correctly identified global food security as a needed outcome of their research. However, it has been argued that basic plant science does not directly contribute to global food security. It is commonly accepted that basic science is necessary to facilitate the applied plant science research such as plant breeding or food science that directly result in improved food quantity and quality. Here we will demonstrate that the broader impacts of basic and applied research not only improve the quantity and quality of the food supply but also deliver

vital broader impacts that are necessary in order to achieve sustainable global food security in perpetuity. Evidence of broader impacts are often qualitative, thus to demonstrate their role in potentially achieving global food security an anecdote collected from the dissertation research conducted in the previous chapters will be presented for each broader impact.

4.3 What are the broader impacts?

The NSF is an independent federal agency in the United States of America that supports fundamental research in science, engineering, and education with an annual budget of $7.5 billion USD in 2017 (125). Of the approximately 50,000 proposals that the NSF receives each year, 22% to 24% will be funded (125). The proposals are considered for funding based off of (1) the intellectual merit and (2) the broader impacts of the proposed project. Intellectual merit refers to the potential of the project to achieve the NSF's first strategic goal which is to advance or expand knowledge in science, engineering and learning, whereas broader impacts are defined as "the potential to benefit society and contribute to the achievement of specific, desired social outcomes (125,126). The broader impacts align with the NSF's second strategic goal of "advancing the capability of the Nation to meet current and future challenges (125). The *America COMPETES Reauthorization Act of 2010* is the piece of legislation that mandates that the NSF evaluates research proposals for their potential to make broader impacts(127).

4.4 Translational research

Translational research generally means research that drives the translation from discovery to releasing an innovative product to an end user. Generally, plant genetics is considered basic science and plant breeding is considered applied science with translational research outcomes. Accelerating the translation from discovery to innovation is one of the strategic objectives of the NSF under the societal impacts category (125). The experiments that encompassed my research were not funded by the NSF, however, did strive to deliver these translation research outcomes. My research was funded by the generous support of the Bill and Melinda Gates Foundation (BMGF) utilizes the Sustainable Development Goals (SDGs) (Table 4.1) implemented by the United Nations to align the research they towards having broader societal impact. There are 17

SDGs that seek to build on the former Millennium Development Goals and strive to end global poverty and hunger (128).

In plant genetics and breeding the translation research outcomes usually are improved plant varieties that farmers can utilize to improve their livelihoods. Delivering improved plant varieties can be considered a broader impact and can potentially address several SDGs including but not limited to goal one, two, three, eight, twelve and fifteen. My Ph.D. research utilized the PP37 multi-parent advanced generation intercross population (MAGIC). The PP37 MAGIC population was developed as a recurrent selection population by Dr. Gebisa Ejeta, under the advisement of Dr. Rex Bernardo in 1986. The PP37 MAGIC population was at that time known as a recurrent selection population in which 25 sorghum accessions, each with their own unique desirable agronomic property, were intermated for six generations to develop a population of novel phenotypic variation. The purpose of developing the PP37 MAGIC population at that time was to identify one or potentially more unique sorghum varieties through recombination combined several of the desired agronomic properties. The breeding objective was to develop a sorghum accession that was high yielding, high quality, resistant to abiotic stress and above all resistant by more than one mode of action to the parasitic plant species *Striga*.

When I began my research 1,200 of these intermated individuals were randomly selected and self-pollinated to develop the PP37 MAGIC population. To date, there are 1,008 individuals of the PP37 population that are not in a genetically homozygous state after six generations of self-pollination. These 1,008 unique sorghum accessions have been evaluated for their *Striga hermonthica* resistance and could be immediately used in the development of hybrid sorghum or released as varieties as inbred sorghum seed does not experience extensive inbreeding depression and inbred lines are commonly used in developing countries. Therefore my Ph.D. project directly possessed a translational research component and with additional evaluation of these unique PP37 MAGIC sorghum accession, these lines could be utilized by farmers across the globe, especially in Sub-Saharan Africa where the parasite *Striga hermonthica* is a direct hindrance to the achievement of the SDGs.

4.5 Germplasm collections

The PP37 MAGIC population is unique because it is a breeding population as discussed, but as presented in earlier chapters this population is extremely powerful for in the realm of identifying the genomic regions associated to a quantitative trait using relatively simple statistical genetic tools. There are several publicly available MAGIC populations in other species such as cowpea, maize, rice and, sorghum (86,87,89,91,92). The currently available sorghum MAGIC population is also very powerful statistically however it does not have extensive phenotypic diversity for certain traits such as seed quality traits and *Striga* resistance (66). The currently available sorghum MAGIC also of 1,000 unique accessions, however only 200 of them have been sequenced to date (92). The PP37 MAGIC population consists of 1,008 unique accessions and we generated low coverage genome sequences for all of them. The PP37 MAGIC was also developed using a different and arguably more diverse set of founding parental sorghum accessions than the seminal sorghum MAGIC. This is purely anecdote however when walking the PP37 fields with Dr. Gebisa Ejeta and Dr. Tesfaye Mengiste they were shocked to see the extensive diversity present in the population, especially when compared to the seminal sorghum MAGIC which is also known as PP34.

As an aside, the PP stands for Purdue Population and the numbering indicates that there are potentially 36 populations developed my Dr. Ejeta that predicate those that are currently being researched and the numbers continue. Therefore, the genetic potential sitting in the Dr. Ejeta's germplasm collections is astounding and awe-inspiring to say the very least. However, when focusing on specifically on the PP37 MAGIC, this population represents 1,008 recombinant inbred lines, derived from 25 diverse parental lines that can be used to test an infinite number hypothesis. This is corroborated by the fact that there are genome sequences for all 1,008 inbred and 20 of the 25 founding parental lines. By making the PP37 MAGIC population public or open access, scores of other basic and applied scientists can continue to unravel the genetic architecture of sorghum and expedite translation research of this staple food crop (129). The delivery of well-curated sequenced MAGIC populations could strongly be considered a broader impact of any PGRP funded research and would contribute towards the SDGs numbers one, two, three, four, eight, nine, thirteen and fifteen.

4.6 Building human capacity in plant breeding and genetics

Around the globe, agriculture is often the primary economic driver. The global seed market was valued at $58 Billion USD in 2016 (130). Access and utilization to improved seeds help farmer become more productive and profitable. To support this robust industry there needs to be a well-trained global workforce trained in plant breeding and genetics. Plant breeding is a unique career that is inherently interdisciplinary as indicated by the following definition of a plant breeder:

> "A plant breeder possesses strong knowledge of genetics and information management and draws upon elements of many other disciplines, including pathology, entomology, crop physiology, experimental design and analysis, remote sensing, and soil and environmental sciences. Although a plant breeder must be comfortable with laboratory techniques, she or he is committed first and foremost to high quality, high throughput fieldwork, observing plant response in the target environment. It is in the field that putative yield enhancing mechanisms are properly tested and where they must demonstrate improved or stabilized crop yields under realistic production conditions. A breeder must be able to understand, critically evaluate, and use new methods emerging from many diverse fields of investigation and different disciplines, and to evaluate their potential for improved rates of genetic gain. This calls for flexibility, openness, a commitment to staying abreast of the professional literature, and a willingness to retrain." (131)

Despite the strong market, training in plant breeding and the associated applied genetic techniques in limited globally especially across Sub-Saharan Africa (132,133). In a survey of 137 Eastern and Southern Africa plant breeders it was found that only 26% of those surveyed had a Ph.D. and that only 60% of all plant breeders in Eastern and Southern Africa were using any molecular breeding techniques as of 2013 (134). In Ethiopia more than half the plant breeders have only been trained at the BSc level (134). In addition to the undereducated workforce, there is also a gender disparity given that only 15% of the Ethiopian plant breeders surveyed were women which is the second lowest in all of the countries in East and South Africa (134).

While in Ethiopia conducting this research, I managed up to 25 Ethiopian women and three men, 13 of which worked with me for two consecutive years. The employees in Ethiopia were between the ages of 17 and 25 years old and all of them had some high school education or had completed high school. All of the 25 Ethiopian employees had attended high school at least until the 10^{th} grade and have received formal instruction in English. Only two of the 25 Ethiopian employees could speak basic English, however, a translator was required for all technical discussions especially regarding the instruction of the daily activities. The daily activities for the

Ethiopian staff were similar to field staff in the United States such as field preparation, field experimental design layout, planting, weeding and extensive phenotypic data collection.

In Ethiopia, these workers are referred to as "daily labor" as they are usually paid daily at the Tigray Agricultural Research Institute rate which is approximately $2.00 USD a day. The weather of Humera however, is dangerously hot, there are no public bathroom facilities and potable water is often in short supply. At the Humera Agricultural Research Center I was the first to implement policies on work breaks to facilitate the use of the limited bathroom facilities, policy to mandate that researchers provide an ample supply of potable water and food, especially when traveling to offsite location. I was also the catalyst in developing policy mandating that any and all accidents of impairments of human health be dealt with at the local hospital despite the local preference of taking any ill person to a place of worship. I also implemented safety protocols to ensure that each employee had an emergency contact on file so that in case of an accident the contact would be notified as we were on route to the hospital so that decisions could be made with respect to their cultural, ethical and religious preference. It is important to remember that all of these accommodations that we almost take for granted in the United States are not ensured in developing countries and it is your moral and ethical imperative to ensure these basic human and worker rights wherever you go. Also, providing a young woman with plenty of water, food, access to education and safe sanitation facilities is the purest form of women empowerment.

In hindsight, I regret not collecting more information about my Ethiopian field staff, aside from this basic demographic information. However, such a research activity would have required the permission of the Institution Review Board and since it was not a funded research question this was not pursued. In the future, such research questions and data collection will be incorporated into my research endeavors. I ca however say anecdotally that the women who worked with me were extremely shy when we started the work and would barely lift their gaze from the ground to tell me their name when we started work. By the end of the research experience, these young women would confidently correct my mistakes, ask questions and take leadership roles with ease. These young men and women were not just trained but also empowered to implement field research projects that were larger than all of the other research projects of the 20 other researchers at the center combined. These young people learned how to carefully collect data and had the confidence to tell me when I was wrong, or something wasn't working well. They took ownership and pride in regard to work. I can proudly state that three of the young women are pursuing

Agricultural Science at the undergraduate level at Ethiopian universities and one of the men is pursuing a certificate program in agriculture at a technical college. I continue to support them in an advisory role and the experience has been mutually beneficial.

4.7 Responsibly implementing international research for sustainable impact

After spending a significant portion of my graduate work and the last decade abroad I have developed a certain ethos about conducting international research. Although this ethos is ever evolving, anecdotal, and not founded in a direct scientific observation I feel that it is appropriate to share this ethos as a conclusion to this brief discussion of broader impacts. As discussed, broader impacts are essentially the social impacts of a given research endeavor. The social impacts can be designed with intent as instructed by the NSF and have wonderful impacts. Broader impacts can empower people to change the world, they can put data in the hands of scientists across the globe, inspire a young person to consider a career in science and potentially influence our society in ways we don't even yet understand. However, in the words of the fictional Ben Parker, the uncle of Spiderman, "with great power comes great responsibility". Despite the entertainment value one of America's favorite heroes, this is a statement that should not be taken lightly because if we are not careful, the best-intended researchers can do the opposite of what they set out to achieve and actually do more harm than good with their broader impacts. On October 4th, 2016 Dr. Sharon Gray was returning to the Addis Ababa airport in Ethiopia and she was struck by a rock being thrown by anti-government protestors as she was traveling in transportation provided by the Ethiopian government. Dr. Gray passed away before she could reach the airport and it is a day I will never forget for the rest of my life. To solidify that Dr. Gray did not die in vain I will discuss my observations opinions and potential solutions to ensure our broader impacts are implemented responsibly for safe and sustainable positive impact.

Above all, the most important objective is to keep you and your team members safe as you execute your research and broader impacts. It is quite easy to become overzealous about the exciting potential outcomes of our research and ignore seemingly innocuous warning signs of danger or maligned impacts. Please always use institutional support in the form of a research ethics office, travel clinics, register with the United States Department of State before embarking on international travel, use the Center for Disease Control website for health information and above

all educate yourself on the religion, culture, language and other relevant aspects of your target population.

Personally, while traveling abroad I have always found the in print hard copies (you never know when the network will go out) of the Lonely Planet travel guides to be well written, painfully honest, candid and this information literally saved my life and reputation a time or two. Did you know that if you eat all of the food on your plate in some parts of Ethiopia, and don't leave a small bite or two that your hosts will think you are inviting famine into their home? I will assume the reader doesn't know this, and it is something I learned from the travel guide and made sure to observe as a broke bread across the country. Similarly, did you know that the word "no" does not always mean "no" when a man makes an advance on a woman in Ethiopia and culturally woman say "no" to imply their sexual inexperience? Lonely Planet suggests never being alone with a man in any capacity because it can be a mistaken invitation. As a woman working in a male-dominated occupation, it is vital to know this kind of cultural information and respect it. These are all examples of the importance of educating yourself to avoid offending your hosts or entering into a potentially life-threatening situation. Even while traveling in your home country it isn't safe to assume you understand the cultural, religious and social mindset of those around you and your research will always be better if you take time to understand these things. Our home Universities are filled with wonderful experts from ethicists, rural sociologists, human health professionals, risk managers and economists just to name a few that are available to provide institutional support to make sure we stay safe and sound.

When personal safety is ensured with intent, our intent must be focused on the broader impacts of our science. Whether we are in the metropolitan areas of America, a kindergarten classroom or simply having a cup of coffee outside a grass house in rural Ethiopia we have a responsibility to be stewards and diplomats of science. I have seen egregious behavior from members of the scientific community domestically and abroad that truly alienates the public and fosters the populist views of the general public. Most of our research is executed with either tax dollars or tuition dollars and we should respect those who fund our endeavors. Furthermore, it is our duty to effectively communicate our science in a respectful manner. I have heard scientists on countless occasions saying that they have to "dumb down" their research to talk to the public. As a previous agricultural education major and extension specialist at Cornell University, I can say confidently that making a document the appropriate reading level and vernacular for the general

audience is not trivial. When implementing broader impacts, we need to consider the audience, how to educate our audience appropriate and respectfully and how to use institutional support such as extension services and the libraries to support our endeavors.

In the increasingly digital age as we take to social media, we also need to be stewards of this powerful tool to and use it to respectfully provide information. As any extension educator worth their salt will tell you, it is not our role as public sector educators to tell people what to think, we just provide them with information to think about and generate their own opinions that align with their ethical, cultural and social values. It is high time we reinvigorate the mission of the land grant universities by planning the broader impacts with research, education and extension represented for the planning and execution of our work.

As a 2015 graduate of the U.S. Borlaug Global Food Security Summer Institute I have been trained to approach research in an interdisciplinary manner. This means inviting ethicists, economists, nutritionists, food scientists, rural sociologists and even policy makers to my research table, that conventionally would one have plant breeders and genetics around it. I have never regretted this mode of action and it has always improved my research and broader impacts. This may seem daunting; however, your home institutions are brimming with the human capacity to do this and therefore all you need to do is extend a kind hand of collaboration and design a mutually beneficial research endeavor. Interdisciplinary endeavors will shape the next generation research that has the potential to truly achieve global food security and end the poverty pandemic.

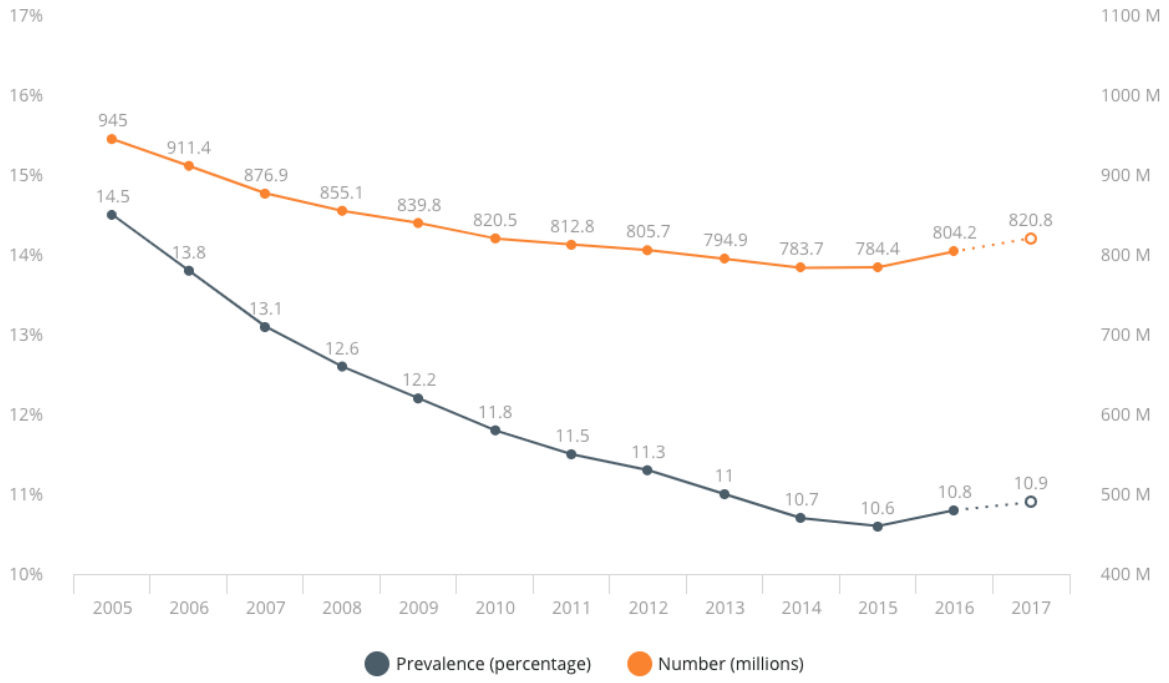

Figure 4.1: The number of malnourished people in the world is increasing.

The number (orange) and prevalence (blue) of malnourished people globally over time. Source: Food and Agriculture Organization. The y-axis on the left is the percentage of the global population and the y-axis on the right is the number of people in millions (M).

Table 4:1 The Sustainable Development Goals, developed by the United Nations.

Sustainable Development Goals

Goal 1. End poverty in all its forms everywhere

Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture

Goal 3. Ensure healthy lives and promote well-being for all at all ages

Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

Goal 5. Achieve gender equality and empower all women and girls

Goal 6. Ensure availability and sustainable management of water and sanitation for all

Goal 7 Ensure access to affordable, reliable, sustainable and modern energy for all

Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

Goal 10. Reduce inequality within and among countries

Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable

Goal 12. Ensure sustainable consumption and production patterns

Goal 13. Take urgent action to combat climate change and its impacts*

Goal 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development

Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss

Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

Goal 17. Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

* Acknowledging that the United Nations Framework Convention on Climate Change is the primary international, intergovernmental forum for negotiating the global response to climate change.

REFRENCES

1. Kuijt J. The Biology of Parasitic Flowering Plants. University of California Press; 1969.
2. Irving LJ, Cameron DD. You are What You Eat. Interactions Between Root Parasitic Plants and Their Hosts. In: Advances in Botanical Research [Internet]. 1st ed. Elsevier Ltd; 2009. p. 87–138. Available from: http://dx.doi.org/10.1016/S0065-2296(08)00803-3
3. Parker C. Observations on the current status of *Orobanche* and <i>Striga<i> problems worldwide. Pest Manag Sci. 2009;65(5):453–9.
4. Ejeta G. Breeding for *Striga* resistance in sorghum: Exploitation of an intricate host-parasite biology. Crop Sci. 2007;47:S216–27.
5. Spallek T, Mutuku M, Shirasu K. Pathogen profile The genus *Striga* : a witch profile. 2013;14:861–9.
6. Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 2009;14(1):21–9.
7. Caldwell RM, Schafer JF, Compton LE, Patterson FL. Tolerance to Cereal Leaf Rusts. Science (80-). 1958;128.
8. Watt WL. Control of Striga Weed in Nyanza Province , Kenya. East African Agric J. 1936;1(4):320–2.
9. Doggett H. *Striga hermonthica* on sorghum in East Africa. J Agric Sci. 1965;65(1965):183–194.
10. Jain RP, Ejeta G. Annual Report of the Sudan Cooperative Sorghum and Millet Crop Improvement Program. 1981.
11. Rao MJV. Techniques for screening sorghums for resistance to *Striga*. Vol. 20, Information Bulletin. Patancheru, India; 1985.
12. Gilliver B, Rao MJV, Venkateswarlu P. A design and methods of analysis to monitor crop growth conditions illustrated with sorghum screening trials for resistance to Striga. Explor Agric. 1986;21(3):233–40.
13. Van Delft GJ, Graves JD, Fitter AH, Pruiksma MA. Spatial distribution and population dynamics of *Striga hermonthica* seeds in naturally infested farm soils. Plant Soil. 1997;195(1):1–15.
14. Ejeta G, Butler LG. Host plant resistance to Striga. Int Crop Sci. 1993;I:561–9.

15. Ejeta G, Butler LG. Host-parasite interactions throughout the *Striga* life cycle, and their contributions to *Striga* resistance. Vol. 1, African Crop Science Journal. 1993. p. 75–80.
16. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet [Internet]. 2011 Jun [cited 2014 Jul 9];12(6):443–51. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3593722&tool=pmcentrez&rendertype=abstract
17. Ejeta G. The Striga Scourge in Africa: A Growing Pandemic. In: Integrating New Technologies for Striga Control : Towards Ending the Witch-Hunt. 2007. p. 3–16.
18. Bennett JR, Mathews S. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. Am J Bot. 2006;93(7):1039–51.
19. Yoshida S, Ishida JK, Kamal NM, Ali AM, Namba S, Shirasu K. A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, Striga hermonthica. BMC Plant Biol. 2010;10:55.
20. Mohamed KI, Musselman LJ, Riches CR. The Genus *Striga*(Scrophulariaceae) in Africa. Missouri Bot Gard. 2001;88(1):60–103.
21. Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, et al. Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol. 2015;32(3):767–90.
22. Gbéhounou G, Pieterse AH, Verkleij JA. Endogenously induced secondary dormancy in seeds of Striga hermonthica. Weed Sci. 2000;48(5):561–6.
23. Brown R, Edwards M. The germination of the seed of *Striga lutea*. Ann Bot. 1944;8(2–3):131–48.
24. Vallance KB. Studies on the germination of the seeds of *Striga hermonthica*. Ann Bot. 1950;XIV(55):347–63.
25. Saunders AR. Studies in Phanerogamic Parasitism. Sci Bull. 1933;128.
26. Brown R, Johnson AW, Robinson E, Todd AR, F.R.S. The stimulant involved in the germination of *Striga hermonthica*. R Soc Publ. 1949;136:1–12.

27. Cook CE, Whichard LP, Wall ME, Egley GH, Coggon P, Luhan PA, et al. Germination Stimulants. II. The Structure of Strigol–A Potent Seed Germination Stimulant for Witchweed (*Striga lutea* Lour.)1,2. J Am Chem Soc. 1972;94(17):6198–9.
28. Chang M, Lynr DG, Netzly DH, Butler LG. Chemical Regulation of Distance: Characterization of the First Natural Host Germination Stimulant for *Striga asiatica*. J Am Chem Soc. 1986;108(24):7858–60.
29. Hess DE, Ejeta G, Butler LG. Selecting sorghum genotypes expressing quantitative biosynthetic trait that confers resistance to Striga. Phytochemistry. 1992;31(2):493–7.
30. Hauck C, Müller S, Schildknecht H. A Germination Stimulant for Parasitic Flowering Plants from Sorghum bicolor, a Genuine Host Plant. J Plant Physiol. 1992;139(4):474–8.
31. Siame BA, Weerasuriya Y, Wood K, Ejeta G, Butler LG. Isolation of Strigol, a Germination Stimulant for *Striga asiatica*, from Host Plants. J Agric Food Chem. 1993;41(9):1486–91.
32. Butler LG. Chemical Communication Between the Parasitic Weed Striga and Its Crop Host. In: Allelopathy. 1994. p. 158–68.
33. Mohemed N, Charnikhova T, Bakker EJ, Ast A Van, Gt A, Bouwmeester HJ. Evaluation of field resistance to Striga hermonthica (Del.) Benth. in Sorghum bicolor (L.) Moench. The relationship with strigolactones. Pest Manag Sci. 2016;72:2082–90.
34. Yoneyama K, Arakawa R, Ishimoto K, Kim H Il, Kisugi T, Xie X, et al. Difference in *Striga* -susceptibility is reflected in strigolactone secretion profile, but not in compatibility and host preference in arbuscular mycorrhizal symbiosis in two maize cultivars. New Phytol. 2015;
35. Arite T, Umehara M, Ishikawa S, Hanada A, Maekawa M, Yamaguchi S, et al. *d14*, a Strigolactone-Insensitive Mutant of Rice, Shows an Accelerated Outgrowth of Tillers. Plant Cell Physiol. 2009;50(8):1416–24.
36. Hamiaux C, Drummond RSM, Janssen BJ, Ledger SE, Cooney JM, Newcomb RD, et al. DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. Curr Biol. 2012;22(21):2032–6.
37. Toh S, Holbrook-Smith D, Stokes ME, Tsuchiya Y, McCourt P. Detection of parasitic plant suicide germination compounds using a high-throughput Arabidopsis HTL/KAI2 strigolactone perception system. Chem Biol. 2014;21(8):988–98.

38. Chevalier F, Nieminen K, Sanchez-Ferrero JC, Rodriguez ML, Chagoyen M, Hardtke CS, et al. Strigolactone Promotes Degradation of DWARF14, an α/β - Hydrolase Essential for Strigolactone Signaling in *Arabidopsis*. Plant Cell. 2014;26(3):1134–50.
39. Waters MT, Nelson DC, Scaffidi A, Flematti GR, Sun YK, Dixon KW, et al. Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in Arabidopsis. Development. 2012;139(7):1285–95.
40. Tsuchiya Y, Yoshimura M, Sato Y, Kuwata K, Toh S, Holbrook-Smith D, et al. Probing strigolactone receptors in Striga hermonthica with fluorescence. Science (80-). 2015;349(6250):864–8.
41. Toh S, Holbrook-Smith D, Stogios PJ, Onopriyenko O, Lumba S, Tsuchiya Y, et al. Structure-function analysis identifies highly sensitive strigolactone receptors in Striga. Science (80-). 2015;350(6257):203–7.
42. Sugimoto Y, Mukhtar Ali A, Yabuta S, Kinoshita H, Inanaga S, Itai A. Germination strategy of Striga hermonthica involves regulation of ethylene biosynthesis. Physiol Plant. 2003;119(1):137–45.
43. Chang M, Lynn DG. The haustorium and the chemistry of host recognition in parasitic angiosperms. J Chem Ecol. 1986;12(2):561–79.
44. Kim D, Kocz R, Boone L, Keyes WJ, Lynn DG. On becoming a parasite: Evaluating the role of wall oxidases in parasitic plant development. Chem Biol. 1998;5(2):103–17.
45. Dörr I. How *Striga* parasitizes its host: A TEM and SEM study. Ann Bot. 1997;79(5):463–72.
46. Yoshida S, Cui S, Ichihashi Y, Shirasu K. The Haustorium, a Specialized Invasive Organ in Parasitic Plants. Annu Rev Plant Biol. 2016;67:643–67.
47. Berner D, Kling G, Singh B. *Striga* research and control. A Perspective from Africa. Plant Dis. 1994;79(7):652–60.
48. Graves JD, Press MC, Stewart GR. A carbon balance model of the sorghum-Striga hermonthica host-parasite association. Plant Cell Environ. 1989;12(1):101–7.
49. Bebawi FF, Eplee RE, Harris CE, Norris RS. Longevity of Witchweed (*Striga asiatica*) Seed. Weed Sci. 1984;32(4):494–7.

50. Satish K, Gutema Z, Grenier C, Rich PJ, Ejeta G. Molecular tagging and validation of microsatellite markers linked to the low germination stimulant gene (lgs) for *Striga* resistance in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet. 2012;124(6):989–1003.
51. Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, et al. Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes *Striga* resistance. Proc Natl Acad Sci. 2017;114(17):4471–6.
52. Ejeta G. Integrating biotechnology, breeding, and agronomy in the control of the parasitic weed Striga spp in sorghum. In: In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution Avenue Media, Bologna. 2005. p. 239–251.
53. Mace ES, Rami J-F, Bouchet S, Klein PE, Klein RR, Kilian A, et al. A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol. 2009;9:1–14.
54. Mohamed A, Ellicott A, Housley TL, Ejeta G. Hypersensitive Response to *Striga* Infection in *Sorghum*. Crop Sci. 2003;43:1320–4.
55. Doggett H. Sorghum, 2nd Edition. Essex England: Longman Scientific & Technical; 1988.
56. Kim S-K. Genetics of maize tolerance of *Striga hermonthica*. Crop Sci. 1994;34(4):900–7.
57. Haussmann BIG, Hess DE, Welz H-G, Geiger HH. Improved methodologies for breeding Striga-resistant sorghums. F Crop Res. 2000;66:195–211.
58. Cherif-Ari O, Housley TL, Ejeta G. Sorghum root length density and the potential for avoiding *Striga* parasitism. Plant Soil. 1990;121:67–72.
59. Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Geiger HH. Genomic regions influencing resistance to the parasitic weed Striga hermonthica in two recombinant inbred populations of sorghum. Theor Appl Genet. 2004;109:1005–16.
60. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The *Sorghum bicolor* genome and the diversification of grasses. Nature. 2009 Jan 29;457:551–6.
61. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 2013 Jan;4:2320.

62. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J. 2018;93(2):338–54.
63. Dugas D V., Monaco MK, Olsen A, Klein RR, Kumari S, Ware D, et al. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics. 2011;12(1):514.
64. Yazawa T, Kawahigashi H, Matsumoto T, Mizuno H. Simultaneous Transcriptome Analysis of Sorghum and *Bipolaris sorghicola* by Using RNA-seq in Combination with De Novo Transcriptome Assembly. PLoS One. 2013;8(4).
65. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci. 2013;110(2):453–8.
66. Ongom PO, Volenec JJ, Ejeta G. Selection for drought tolerance in sorghum using desiccants to simulate post-anthesis drought stress. F Crop Res [Internet]. 2016;198:312–21. Available from: http://dx.doi.org/10.1016/j.fcr.2016.03.015
67. Chopra R, Burow G, Burke JJ, Gladman N, Xin Z. Genome-wide association analysis of seedling traits in diverse Sorghum germplasm under thermal stress. BMC Plant Biol. 2017;17(1):1–15.
68. Rhodes DH, Hoffmann L, Rooney WL, Herald TJ, Bean S, Boyles R, et al. Genetic architecture of kernel composition in global sorghum germplasm. BMC Genomics [Internet]. 2017;18(1):1–8. Available from: http://dx.doi.org/10.1186/s12864-016-3403-x
69. Higgins RH, Thurber CS, Assaranurak I, Brown PJ. Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. G3: Genes|Genomes|Genetics [Internet]. 2014;4(9):1593–602. Available from: http://g3journal.org/cgi/doi/10.1534/g3.114.013318
70. Boyles RE, Cooper EA, Myers MT, Brenton Z, Rauh BL, Morris GP, et al. Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. Plant Genome. 2016;9(2).
71. Adeyanju A, Little C, Yu J, Tesso T. Genome-Wide Association Study on Resistance to Stalk Rot Diseases in Grain Sorghum. G3 Genes|Genomes|Genetics. 2015;5:1165–75.

72. Doerge RW. Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet. 2002 Jan;3(1):43–52.
73. Salvi S, Tuberosa R. To clone or not to clone plant QTLs: Present and future challenges. Trends Plant Sci. 2005;10(6):297–304.
74. Borevitz JO, Nordborg M. Update on genomics and natural variation in Arabidopsis: The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol. 2003;132:718–25.
75. Liu D, Ma C, Hong W, Huang L, Liu M, Liu H, et al. Construction and analysis of high-density linkage map using high-throughput sequencing data. PLoS One. 2014;9(6).
76. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, et al. Genetic properties of the maize nested association mapping population. Science. 2009;325(5941):737–40.
77. Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, et al. Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. Genetics. 2017;206(2):573–85.
78. Li H, Bradbury P, Ersoz E, Buckler ES, Wang J. Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS One [Internet]. 2011 Jan [cited 2014 May 27];6(3). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3057965&tool=pmcentrez&rendertype=abstract
79. Brown PJ, Rooney WL, Franks C, Kresovich S. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics. 2008;180(1):629–37.
80. Holland JB. Genetic architecture of complex traits in plants. Curr Opin Plant Biol. 2007;10(2):156–61.
81. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. Curr Opin Biotechnol. 2006;17(2):155–60.
82. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Disease. Science (80-). 1996;273:1516–7.
83. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of Linkage Disequilibrium in Plants. Annu Rev Plant Biol. 2003;54(1):357–74.

84. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9.
85. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38(2):203–8.
86. Holland JB. MAGIC maize: A new resource for plant genetics. Genome Biol. 2015;16(1):15–7.
87. Bandillo N, Raghavan C, Muyco PA, Anna M, Sevilla L, Lobina IT, et al. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice. 2013;6(11):1–15.
88. Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, et al. An Eight-Parent Multiparent Advanced Generation Inter-Cross Population for Winter-Sown Wheat: Creation, Properties, and Validation. G3 Genes|Genomes|Genetics. 2014;4(9):1603–10.
89. Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, et al. Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. Genome Biol. 2015;16(1):1–23.
90. Sannemann W, Huang BE, Mathew B, Léon J. Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol Breed. 2015;35(3).
91. Huynh BL, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JRP, et al. A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). Plant J. 2018;93(6):1129–42.
92. Ongom PO, Ejeta G. Mating Design and Genetic Structure of a Multi-parent Advanced Generation Inter-cross (MAGIC) Population of Sorghum (*Sorghum bicolor* (L.) Moench). G3 Genes|Genomes|Genetics. 2017;8(January):g3.300248.2017.
93. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, et al. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet. 2009;5(7).
94. Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, et al. Community resources and strategies for association mapping in Sorghum. Crop Sci. 2008;48(1):30–40.

95. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One. 2014;9(2).
96. Westwood JH, Yoder JI, Timko MP, dePamphilis CW. The evolution of parasitism in plants. Trends Plant Sci [Internet]. 2010;15(4):227–35. Available from: http://dx.doi.org/10.1016/j.tplants.2010.01.004
97. FAO. FAOSTAT. Food and Agricultural Organization of the United Nations. 2019.
98. Van Der Plank J. Diease resistance in plants. Acad Press. 1968;1–203.
99. Musselman LJ, Parker C. Studies on Indigo Witchweed, the American Strain of Striga gesnerioides (Scrophulariaceae). Weed Sci. 1981;29(5):594–6.
100. Li J, Timko MP. Gene-for-gene resistance in striga-cowpea associations. Science (80-). 2009;325(5944):1094.
101. Lane JA, Bailey JA, Butler RC, Terry PJ. Resistance of cowpea to Striga gesnerioides Vatke, a parasitic angiosperm. New Phytol [Internet]. 1993;125:405–12. Available from: http://www.jstor.org/view/0028646x/di008775/00p0244l/0
102. Lam E, Kato N, Lawton M. Programmed cell death, mitochondria and the hypersensitive response. Nature. 2001;411.
103. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: Adaptable guards. Genome Biol. 2006;7(4).
104. Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, et al. The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. BMC Plant Biol. 2014;14(1):1–14.
105. Grenier C, Ibrahim Y, Haussmann BIG, Ejeta G. Marker-assisted selection for Striga resistance in sorghum. In: Ejeta G, Gressel J, editors. Integrating New Technologies for Striga Control: Towards Ending the Witchhunt. World Scientific; 2007. p. 159–72.
106. Scholes JD, Press MC. Striga infestation of cereal crops – an unsolved problem in resource limited agriculture. Curr Opin Plant Biol. 2008;11:180–6.
107. Timko MP, Huang K, Lis KE. Host Resistance and Parasite Virulence in Striga–Host Plant Interactions: A Shifting Balance of Power. Weed Sci [Internet]. 2012;60(02):307–15. Available from: https://www.cambridge.org/core/product/identifier/S0043174500021342/type/journal_article

108. Elshire RJ, Glaubitz JC, Sun Q, Poland J a, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One [Internet]. 2011 Jan [cited 2014 Jul 9];6(5):e19379. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3087801&tool=pmcentrez&rendertype=abstract
109. Maiti RK, Ramaiah K V, Bisen SS, Chidley VL. A comparative study of the haustorial development of Striga asiatica (L.) kuntze on Sorghum cultivars. Ann Bot. 1984;54(4):447–57.
110. Rao MJV. Striga: Biology and Control. In: Ayensu ES, Doggett H, Keynes RD, Marton-Lefèvre J, Musselman LJ, Parker C, et al., editors. Workshop on the Biology and Control of Striga. ICSU Press; 1983.
111. Gobena D, Shimels M, Rich PJ, Ruyter-spira C, Bouwmeester H, Kanuganti S, et al. Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. PNAS. 2017;114(17):4471–6.
112. Arnaud M-C, Veronesi C, Thalouarn P. Physiology and histology of resistance to Striga hermonthica in sorghum biocolor var. Framida. Aust J Plant Phyriology. 1999;26:63–70.
113. Haussmann BIG, Hess DE, Reddy BVS, Welz HG, Geiger HH. Analysis of resistance to Striga hermonthica in diallel crosses of sorghum. Euphytica. 2000;116(1):33–40.
114. Federer W, Raghavarao D. On Augmented Designs. Biometrics. 1975;31(1):29–35.
115. Doggett H, Eberhart SA. Recurrent Selection in Sorghum. Crop Sci. 1968;8:119–21.
116. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci [Internet]. 1984;81(24):8014–8. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.81.24.8014
117. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics [Internet]. 2007 Oct 1 [cited 2014 Jul 12];23(19):2633–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17586829
118. Xin Z, Velten JP, Oliver MJ, Burke JJ. High-throughput DNA extraction method suitable for PCR. Biotechniques. 2003;34(4):820–6.

119. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics [Internet]. 2012 Sep 15 [cited 2014 Dec 15];28(18):2397–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22796960
120. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. Plant Genome [Internet]. 2016;9(2):0. Available from: https://dl.sciencesocieties.org/publications/tpg/abstracts/9/2/plantgenome2015.11.0120
121. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing [Internet]. Vol. 57, Journal of the Royal Statistical Society B. 1995. p. 289–300. Available from: http://www.stat.purdue.edu/~doerge/BIOINFORM.D/FALL06/Benjamini and Y FDR.pdf%5Cnhttp://engr.case.edu/ray_soumya/mlrg/controlling_fdr_benjamini95.pdf
122. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet [Internet]. 2010;42(4):355–60. Available from: http://dx.doi.org/10.1038/ng.546
123. Mace E, Innes D, Hunt C, Wang X, Tao Y, Baxter J, et al. The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. Theor Appl Genet [Internet]. 2018; Available from: https://doi.org/10.1007/s00122-018-3212-5
124. Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, et al. Mutation in sorghum LGS1 alters strigolactones and causes Striga resistance. Proc Natl Acad Sci. 2017;
125. National Science Foundation. Building the Future: Investing in Discovery and Innovation, NSF Strategic Plan for Fiscal Years (FY) 2018-2022. 2018; Available from: https://www.nsf.gov/pubs/2018/nsf18045/nsf18045.pdf
126. Perspectives on Broader Impacts [Internet]. National Science Foundation; Available from: https://www.nsf.gov/od/oia/publications/Broader_Impacts.pdf
127. America Competes Reauthorization Act of 2010. Public Law 111th Congress; 2011 p. 1–69.

128. Transforming our world: the 2030 Agenda for Sustainable Development [Internet]. United Nations General Assembly 2015. Available from: http://www.un.org/en/development/desa/news/sustainable/un-adopts-new-global-goals.html#more-15178
129. Langridge P, Waugh R. Harnessing the potential of germplasm. Nat Genet. 2019;51(February):200–1.
130. Seeds Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2018-2023. 2017.
131. Morris M, Edmeades G, Pehu E. The global need for plant breeding capacity: What roles for the public and private sectors? HortScience. 2006;41(1):30–9.
132. Guimarães EP, Kueneman E, Carena MJ. Assessment of national plant breeding and biotechnology capacity in Africa and recommendations for future capacity building. HortScience. 2006;41(1):50–2.
133. Guimarães EP, Kueneman E, Paganini M. Assessment of the national plant breeding and associated biotechnology capacity around the world. Crop Sci. 2007;47(SUPPL. DEC.).
134. Diop NN, Okono F, Ribaut J-M. Evaluating Human Resource Capacity for Crop Breeding in National Programs in Africa and South and Southeast Asia. Creat Educ. 2013;04(10):72–81.