TRANSCRIPTIONAL PROFILING OF SORGHUM MIDGE DEFENSE RESPONSES IN SORGHUM USING cDNA-AFLP

MBINDA WILTON MWEMA (B.Sc)
Reg. No. 156/11080/2004

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (BIOTECHNOLOGY) IN THE SCHOOL OF PURE AND APPLIED SCIENCES, KENYATTA UNIVERSITY.

NOVEMBER, 2010
DECLARATION

I declare that the work presented in this thesis is my original work and has not been presented for a degree in any other university or any other award.

Mbinda Wilton Mwema
Signature ........................................ Date ........................................

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Prof. Eliud N.M. Njagi
Department of Biochemistry and Biotechnology
Kenyatta University
Nairobi, Kenya
Signature ........................................ Date ........................................

Prof. Christina Dixelius
Department of Plant Biology and Forest Genetics
Swedish University of Agricultural Sciences
Uppsala, Sweden
Signature ........................................ Date ........................................
Nelson Mandela said "Education is the great engine to personal development. It is through education that the daughter of a peasant can become a doctor, that the son of a mine worker can become the head of a mine that the child of a farm worker can become the president of a great nation ".

This thesis is dedicated to my mother; Agnes Munyiva Mbinda, for her sacrifice which enabled me reach this heights. Thank you sweet mum.
ACKNOWLEDGMENTS

The road to success is long and has a lot of ambushes but with courage and perseverance, it can be made easier with the willingness and the support of good people. This is a good opportunity for me to thank all those who have been involved in the success of this work.

I would like to express my sincere gratitude to my SLU supervisor, Prof. Christina Dixelius, for accepting me to work on this project in her lab. Her supervision, great advice, valuable guidance, assistance and personal support helped me conduct this research and complete my studies. She was always ready to help, instruct and teach anytime, although she is always busy and hardworking. She gave me the freedom to perform my work in my own directions and steered me to the right directions when I desperately needed them. To Dr. Dionysious Kiambi of ICRISAT, Nairobi, Kenya, for choosing me among many candidates to perform my MSc. research project in Prof. Christina’s group through their joint research collaboration. Dr. Kiambi devoted a great deal of his valuable time to the planning and the entire development of this work. Prof. Eliud Njagi deserves credit for his expert advice and encouragement from the start of the studies to the end.

To Prof. Richard Hopkins, thank you for providing your chamber to culture the sorghum midge.

Many thanks and deepest gratitude Thomas Martin for his valuable time, ideas, productive discussions and great teachings, instructive guidance, advice, generous personal encouragement and for providing me with the opportunity to learn from his vast experience, but more importantly, the honor of his friendship. For this, I will be always grateful.

To Ramesh, Na, Shinichi, Jonas, Tina, Vera and Hanneke, I had the great fortune of sharing ideas and receiving wise counsel from every one of you. This work would not have been possible without your help and support. Thanks to office mates; Magnus, Izabella, and Veronica for providing an ample environment for academic work.

To my mother, Agnes Munyiva, you sacrificed a lot, for me to be where I am today. I am forever indebted to your love and sacrifice. And that is why I dedicate this thesis to you, mama.

I cannot forget you, Munyiva Mailu, for keeping in touch with me and arguing me to go on even when the winter was too harsh for me.

To Thedi and Chris, follow my example.

This work was carried out at the Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences and was financially supported by the Swedish Research Council (FORMAS), in collaboration with BIO-EARN.
# TABLE OF CONTENTS

DECLARATION .................................................................................................................. II

DEDICATION .................................................................................................................... III

ACKNOWLEDGMENTS ..................................................................................................... IV

TABLE OF CONTENTS .................................................................................................. V

LIST OF TABLES ............................................................................................................. VII

LIST OF FIGURES ......................................................................................................... VIII

ABSTRACT ....................................................................................................................... XII

CHAPTER ONE .............................................................................................................. 1

INTRODUCTION .............................................................................................................. 1

1.1 Background information ....................................................................................... 1

1.2 Statement of problem and justification .................................................................. 5

1.3 Research question .................................................................................................. 6

1.4 Hypothesis .............................................................................................................. 6

1.5 Objectives .............................................................................................................. 6

1.5.1 General objective .............................................................................................. 6

1.6 Output and significance ......................................................................................... 7

CHAPTER TWO ............................................................................................................... 8

LITERATURE REVIEW .................................................................................................. 8

2.1 Origin and distribution of sorghum ........................................................................ 8

2.2 Biology of sorghum ................................................................................................ 8

2.3 Production of sorghum .......................................................................................... 10

2.4 Constraints to production of sorghum in Eastern Africa ....................................... 11

2.5 Sorghum midge biology ......................................................................................... 11

2.6 Integrated pest management of sorghum midge .................................................... 16

2.7 Plant-insect herbivore interactions ....................................................................... 19

2.8 Signaling pathways in plant defense ...................................................................... 20

2.9 Global transcriptome changes in response to insect herbivory ............................. 21

2.10 Differential defense responses to mechanical wounding and insect herbivory feeding .................................................................................................................. 22

2.11 Insect response to plant defense ......................................................................... 23

2.12 Genomics techniques for the study of gene expression ...................................... 24
LIST OF TABLES

Table 1: Primer combinations for selective amplification ........................................................................ 41
Table 2: Primers used for both RT-PCR and qRT-PCR ....................................................................... 44
Table 3: Summary of differentially expressed fragments generated by cDNA-AFLP between AF28 and Seredo upon sorghum midge infestation and homology of differentially expressed fragments with nucleotide or protein sequences in the NCBI database ............................................. 47
LIST OF FIGURES
Figure 1: (a) A flowering sorghum panicle, (b) diagrammatic part of a sorghum floret............10
Figure 2: An adult midge probing inside a flowering sorghum spikelet.................................12
Figure 3: The life cycle of sorghum midge ........................................................................14
Figure 4: A model of recognition of herbivore feeding by resistant and susceptible plants.....19
Figure 5: Schematic description of cDNA-AFLP ..................................................................33
Figure 6: The experimental design .......................................................................................36
Figure 7: A typical cDNA-AFLP gel ...................................................................................46
Figure 8: Functional categories of the sorghum genes responsive to sorghum midge floret feeding ........................................................................................................................52
Figure 9: RT-PCR analysis of three selected differentially expressed cDNA-AFLP transcripts at three different time points after infesting with sorghum midge...........................................53
Figure 10: Relative expression of the three target gene transcripts ........................................55
LIST OF APPENDICES

Appendix 1: Sequences of differentially expressed fragments generated by cDNA-AFLP between AF28 and Seredo upon sorghum midge infestation ................................................................. 85
Appendix 2: qRT-PCR quantification of transcripts for the selected cDNA-AFLP fragments and α-tubulin in infested and non-infested floret tissues of AF28 sorghum at various time-points after infestation with sorghum midge larvae ........................................................................................................... 92
## ABBREVIATIONS AND ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO-EARN-</td>
<td>East African Regional Programme and Research Network for Biotechnology, Biosafety and Biotechnology Policy Development</td>
</tr>
<tr>
<td>Blast-</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td><strong>BstYI-F</strong></td>
<td><strong>BstYI</strong> Forward</td>
</tr>
<tr>
<td><strong>BstYI-R</strong></td>
<td><strong>BstYI</strong> Reverse</td>
</tr>
<tr>
<td>cDNA-</td>
<td>Complimentary Deoxyribonucleic acid</td>
</tr>
<tr>
<td>cDNA-AFLP</td>
<td>Complimentary Deoxyribonucleic acid Amplified Fragment Length Polymorphism</td>
</tr>
<tr>
<td>C&lt;sub&gt;t&lt;/sub&gt;</td>
<td>Critical Threshold</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethylprocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic aAcid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-Stranded Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithioerythritol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FORMAS</td>
<td>Swedish Research Council</td>
</tr>
<tr>
<td>ICRISAT</td>
<td>International Crops Research Institute for the Semi-Arid Tropics</td>
</tr>
<tr>
<td>IMP</td>
<td>Integrated Pest Management</td>
</tr>
<tr>
<td><strong>MseI-F</strong></td>
<td><strong>MseI</strong> Forward</td>
</tr>
<tr>
<td><strong>MseI-R</strong></td>
<td><strong>MseI</strong> Reverse</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative Reverse Transcription Polymerase Chain Reaction</td>
</tr>
</tbody>
</table>
RACE  Random Amplification of the cDNA Ends
RL   Restriction Ligation
RNA  Ribonucleic Acid
rpm  Revolutions per minute
RT   PCR-Reverse Transcription Polymerase Chain Reaction
SIDA Swedish International Development Cooperation Agency
TAE  Tris-Acetate-Ethylenediaminetetraacetic acid
Tris Tris(hydroxymethyl)aminomethane
ABSTRACT

Sorghum midge (Stenodiplosis sorghicola (Coquillett), is the most ubiquitous damaging pest of grain sorghum [Sorghum bicolor (L.) Moench] worldwide. In spite of its economic importance, little is known about the molecular basis of compatible interaction between the pest and sorghum host. Monitoring the transcriptome changes in sorghum in response to developing embryo-feeding sorghum midge will provide a solid foundation to understand the genetic mechanisms of the plant defense against a sorghum midge attack. To identify genes that are differentially expressed in sorghum upon sorghum midge attack, a transcriptome analysis using cDNA-AFLP was profiled in the sorghum midge resistant and susceptible genotypes at time intervals 0, 3, 6 and 12 hours. In total, 12 different primer combinations were used, generating cDNA fragments ranging from 50 bp to 500 bp in size. Of approximately 6,720 cDNA fragments generated, 118 were novel, 17 up-regulated and 23 were down regulated genes. These identified cDNAs were cloned, sequenced, and aligned against GenBank databases using National Centre for Biotechnology Information BlastX and BlastN tools. Database analysis predicted that the proteins encoded by these cDNAs are involved in signal transduction, secondary metabolism, oxidative response, photosynthesis, cell maintenance, and other cellular activities. There are also a large number of identified genes with unknown function. Of the 53 reliable cDNA-AFLP transcripts, 4 homologous to genes known to play a role in defense genes were further selected for validation of cDNA-AFLP expression patterns using qRT-PCR analyses. Overall, the profiles of sorghum genes responsive to sorghum midge feeding indicated that defense genes in sorghum were co-ordinately regulated. Collectively, the results suggest that the transcripts differentially expressed during sorghum midge feeding represent candidate genes of sorghum defense response to sorghum midge stress. Maize protease inhibitor, subtilisin/chymotrypsin inhibitor CI-1B, and cinnamoyl-CoA reductase 2 proteins were hypothesized to be vital for enabling the plant to overcome the stresses inflicted by sorghum midge feeding. The results show that cDNA-AFLP is a reliable technique for studying expression patterns of genes involved in the sorghum-sorghum midge interactions. This study provides the first global catalogue of genes expressed during sorghum and sorghum midge infestation, together with their functional annotations. The results will help to elucidate the molecular basis of the attack process and identify genes and chemicals that could help to inhibit the growth of the pest.
CHAPTER ONE
INTRODUCTION

1.1 Background information

Sorghum, *Sorghum bicolor* (L.) Moench, the fifth most important cereal crop in the world after wheat, maize, rice and barley is one of the main staples for the world's poorest and most food insecure people. Grain sorghum is ranked second among these five important cereals, forming an essential component in traditional farming systems and diets in millions of people in Sub-Saharan Africa (Idris et al., 2008). The crop is a C4 grass and is genetically suited to hot and dry agro-ecologies where it is difficult to grow other food crops. Developing countries account for roughly 90% of the world's sorghum area and 77% of the total output (FAOSTAT, 2010). The lion's share of the crop in these developing countries is grown by small scale farming households operating at the margins of subsistence, and more than 70% of the sorghum is cultivated in the dry and low lands, resulting into low yields. This low production levels per unit area is attributed to deprived agronomic practices and a number of unfavorable factors both abiotic and biotic.

Among the most important biotic factors are insects pest. Of the 150 insect pests that damage the sorghum crop, sorghum midge, *Stenodiplosis sorghicola* (Coquillett), is the most ubiquitous damaging pest of grain sorghum worldwide (Young & Teetes, 1977). Globally, losses due to sorghum midge are known to vary over seasons and locations, but are thought to approximate 10-15% of the world sorghum crop (Sharma, 2004). At flowering, female sorghum midges
oviposit into spikelets, and the larvae feed on the ovary during the following 2 weeks, which results in the failure of kernel formation and seed yield losses (Henzell et al., 1994). The main methods used to control sorghum midge are cultural practices, chemical control and/or development and growth of resistant varieties in an integrated pest management strategy. However, cultural practices are not always effective and the cost of chemical control is prohibitive to most small-scale farmers in addition to the negative impact of insecticides to the environment. Therefore, the most economic, efficient and promising method for controlling the sorghum midge is to exploit host resistance to attack. Several sorghum hybrids which have different levels of sorghum midge resistance have been developed (Mutaliano, 2006).

Genetic resistance of sorghum to sorghum midge is attributed to tolerance, antixenosis and antibiosis. To date, no study has been made to profile sorghum responses to sorghum midge feeding and therefore the genes of sorghum defense to sorghum midge attack still remain largely unknown.

In nature, plants are constantly surrounded by enormous diversity of herbivorous insects that negatively influence plant fitness. However, long-term co-evolution of herbivorous insects and plants has led to the development of an array of constitutive and induced defenses that allow plants to effectively defend themselves from herbivore attack (Becerra, 2007; Rausher, 2001). Constitutive defense mechanisms include physical barriers (such as cuticles, thorns, trichomes, and cell walls) and pre-existing metabolites that are harmful or even toxic to insects (Mithöfer et
al., 2005). In contrast to constitutive defense mechanisms, induced plant defenses involve the activation of mechanisms that directly or indirectly deter herbivores (Maffei et al., 2007). Some secondary metabolites and induced proteins including the release of volatiles which signal the location of insects on infested plants to parasitoids and predators (Kempema et al., 2007; Schnee et al., 2006; De Moraes et al., 2001). When attacked by insects, plants can also produce endogenous signal molecules, such as stress hormones, including jasmonic acid (JA), ethylene, abscisic acid (ABA), and salicylic acid (SA) that regulate signal transduction cascades in plant cells, leading to activation and modulation of defense-related genes (Schmelz et al., 2006; Li et al., 2002).

Plants show varied responses to herbivores that are strongly correlated with the mode of herbivore feeding. Chewing insects and cell-content feeders, such as caterpillars and beetles, cause extensive tissue damage and activate wound-signaling pathways in which JA plays a central role (Kandoth et al., 2007). In contrast, attacks by phloem feeders, such as aphids, plant hoppers, and whiteflies, elicit only weak wound responses, but induce the transcription of SA and pathogen signaling pathways, which may mirror responses to virus vectors or insect-associated bacterial endosymbionts. Alternatively, this transcription signature may be due to similarities between intercellular fungal hyphae growth and stylet penetration, and hence may produce similar responses (Maffei et al., 2007; Li et al., 2006). Limitations associated with JA-mediated defense responses may arise from antagonistic cross talk with SA and ethylene.
One useful approach to the molecular analysis of plant-insect interactions is the determination of changes in steady state mRNA levels occurring in both the host and the insect during attack. Such transcriptomic approaches have been undertaken for different plant-insect/pathogen interactions either by microarray analysis or alternative, open-architecture technologies such as cDNA-amplified fragment length polymorphism (cDNA-AFLP), thus revealing novel information about pathogen genes (Casimiro et al., 2006; Restrepo et al., 2005; Moy et al., 2004; Bittner-Eddy et al., 2003; Maleck et al., 2000; van der Biezen et al., 2000). A few studies have also included proteomic analysis (Bruce et al., 2006).

The PCR-based technique of cDNA-AFLP is widely available at a low cost for various plant species, even if there is little information available at the molecular level. The sensitivity and specificity of this technique allows detecting even the poorly expressed genes and distinguishing between homologous sequences (Breyne et al., 2003). This is an extremely efficient less labour-intensive mRNA fingerprinting method for isolation of those genes which show differential gene expression in stressed conditions. Because of high sensitivity of the technique, it is possible to identify even rare transcripts (Fukumura et al., 2003). As a differential screening method, cDNA-AFLP is more stringent and reproducible than many other methods because it can amplify low-abundance transcripts. The technique is a robust, high-throughput, genome-wide expression tool for gene discovery, where prior knowledge of sequences is not required. The cDNA-AFLP technique has been further improved to avoid the possibility of several transcript derived fragments (TDFs) arising from a single gene/cDNA (Vuylsteke et al. 2006).
1.2 Statement of problem and justification

In East Africa and the entire Sub-Saharan Africa, agriculture is the most important enterprise and the key to economic development. According to Food and Agricultural Organization (FAO, 2009) nearly 59% of the people of Africa earn their livelihood from agriculture. The bulk of African agriculture remains traditional; land holdings are small, crop production is labour intensive and little or no external inputs are used. Pest management practice under this condition is a built-in process in the overall crop production system rather than a separate, well-defined activity.

Arthropod pests are one of the major constraints to agricultural production in Africa. A large number of insect and mite pests attack crops during all stages of growth—seedling to storage. Of these, sorghum midge is the most important pest of grain sorghum worldwide (Harris, 1976). Damage by the sorghum midge can be avoided through early and uniform planting of the same cultivar over a large area in a geographical region. However, it is difficult to plant at times when midge damage can be avoided because of uncertainties of rainfall, inability of the farmers to plant the entire sorghum crop in an area at the same time, and differences in flowering of the sorghum cultivars. Chemical control is costly, ineffective, and beyond the reach of most farmers in the semi arid tropics. Natural enemies for sorghum midge exist, but their populations build up only after damage has been caused. Host plant resistance is therefore the most effective means of keeping midge populations below economic threshold levels (Sharma, 2001). However, the genes which are responsible for midge resistance remain largely unknown.
Full transcriptome analyses using molecular techniques provide important insight into overall changes in gene expression. Such analyses are of special interest as they provide important information on the phenotypic plasticity in the expression of specific genes of interest that can be linked to the dynamics of plant breeding.

This study therefore describes the changes in gene expression in sorghum that occur on infestation by the sorghum midge larvae. Characterization of this set of genes induced provides an insight into the dynamics of the interactions between this midge pest and their host plants.

1.3 Research question

Does sorghum genotype AF28 carry sorghum midge defense response genes?

1.4 Hypothesis

Sorghum genotype AF28 does not carry sorghum midge defense response genes.

1.5 Objectives

1.5.1 General objective

The main goal of the study was to genetically profile the sorghum midge defense responses on sorghum.
1.5.1 Specific objectives

The specific goals of the study were:

i. to apply cDNA-AFLP technique to the sorghum-sorghum midge pest interaction pathosystem

ii. to identify genes that are differentially expressed in sorghum upon sorghum midge infestation using cDNA-AFLP technique,

iii. to validate the expression patterns for some of the regulated genes through reverse transcription polymerase chain reaction (RT-PCR) quantitative real-time-PCR (qRT-PCR).

1.6 Output and significance

This study provides the first global catalogue of sorghum and sorghum midge genes expressed during infestation, together with their functional annotations. This will help to elucidate the molecular basis of the attack process and identify genes and chemicals that could help to inhibit the insect. Additionally, the results could be used for developing molecular markers for sorghum midge defense genes on grain sorghum which would be critical in the development of sorghum midge resistance cultivars through molecular breeding and/or genetic engineering.
CHAPTER TWO
LITERATURE REVIEW

2.1 Origin and distribution of sorghum

Sorghum is a tropical cereal crop that grows in a wide range of environments, playing a significant role as a staple food for many people all over the world. It is believed to have originated from Northeast Africa, where the domestication of the species occurred between 5,000 and 7,000 years ago, and thereafter distributed throughout Africa, Asia and America (De Wet & Huckabay, 1967). Sorghum production is well suited to developing nations of the world because of its sustainability in marginal conditions with limited inputs of water, fertilizers, and pesticide. It is widely grown in a variety of sub-tropical and tropical climates where its versatility and yield is superior to other cereal grains. Sorghum production is greatest in Africa and Asia where it competes with maize and rice as a major food crop, while in countries such as the USA and Australia it is grown almost entirely as a crop for animal feed.

2.2 Biology of sorghum

Sorghum bicolor (L.) Moench (2n=2x=20) is an annual C4 monocot. This African grass which is related to sugarcane and maize has a genome size of about 730 megabases (Paterson et al., 2009). It is predominantly a self-pollinated species, although outcrossing does occur at rates ranging between 3 to 15%, depending on genotype and environment. Genetically, S. bicolor is a functional diploid crop although evidence of tetraploids have been reported (Ghaffari, 2009).
Sorghum plant heights vary from less than a meter to 5 m; inflorescence types vary from open to compact with a wide range of dimensions; basal plant color can be either tan or colored (red or purple). The grain color can vary widely depending on the specific genotype.

The flowering process in grain sorghum has been studied by numerous workers who give variable reports of flowering times (Sharma & Hariprasad, 2002; Diarisso et al., 1997). Blooming occurs mainly in the morning before or after dawn, and is influenced by darkness, temperature and humidity. The flowering process begins with the fanning and opening of the two glumes; followed by the emergence of the stigmas and anthers; and is completed with the closing of the glumes (Figure 1).

The entire flowering period of each spikelet may take from as little as 20 minutes to more than three hours, depending on the sorghum variety and environmental conditions. Pollen dehisces during flowering and germinates on the same day leading to fertilization of the caryopsis and subsequent kernel development.

Depending on growing conditions maximum seed size occurs 10-20 days later, while maximum dry weight of the seed usually 25-50 days after flowering, and grain is harvested when moisture content drops below 15 percent.
When planted under subtropical conditions, even planting densities and optimum soil conditions, sorghum hybrids take between 60-80 days to flower (Doggett, 1988). Flowering proceeds basipetally on each panicle. Yanase et al., (2008) reported that in commercially grown hybrids individual panicles completed flowering in 8 days, while whole fields took 13 days to flower.

2.3 Production of sorghum

Worldwide annual sorghum production is about 60 million tonnes from approximately 40 million hectares (FAOSTAT, 2010; Sasaki & Antonio, 2009). The mean grain yield of sorghum is high, and comparable with crops such as maize and rice. Under optimal field conditions, grain yield can reach 15,000 kg per ha with good yields ranging from 7,000 to 9,000 kg per ha when rain is not a limiting factor. However, because the crop is usually grown in marginal environments, the average sorghum yields are low ranging from 3,400 kg per ha in the Americas to 800 kg per ha.
in Africa (Haussmann et al., 1998). Under average conditions, sorghum yields vary between 3,000 and 4,000 kg per ha and decrease to 300 to 1,000 kg per ha under draught conditions (Ahmed et al., 2000).

2.4 Constraints to production of sorghum in Eastern Africa

In Eastern Africa, more than 70% of sorghum is cultivated in the dry and low lands with low production yield per unit area. Notable factors limiting production of sorghum in tropical Africa include poor climatic conditions, low soil nutrients, weeds, diseases and insect pests (Voortman et al., 2003). Some of these limitations have been solved through farm management and development of suitable plant varieties. Important among the insect pests, the sorghum midge, *Stenodiplosis sorghicola* is the most damaging pest of grain sorghum worldwide (Franzmann et al., 2006, Young & Teetes, 1977).

2.5 Sorghum midge biology

The sorghum midge, *Stenodiplosis sorghicola* (Coquillett), is a fly of the family Cecidomyiidae (Teetes & Pendleton, 2001; Harris, 1976). Adult flies are approximately 2 mm long and have a bright red abdomen making them relatively easy to see despite their small size (Figure. 2).
Figure 2: An adult midge probing inside a flowering sorghum spikelet

In spring midge populations begin to build up slowly from surviving diapaused larvae present in sorghum trash within the soil. Upon increasing soil temperatures and spring rains, the diapaused larvae develop through to adults which fly to infest nearby host plant Johnson grass (*Sorghum halepense* L) spikelets. Johnson grass acts as a primary host crop for the midge population over a couple of generations before flowering grain sorghum and forage sorghum crops begin to flower in summer (Franzmann *et al.*, 2006). Both male and female adult midges are short-lived insects. Emergence of the male imago (adult) from the pupa begins at dawn under favourable temperatures 30-45 minutes before emergence of the female, and peaks one to two hours later (Fisher & Teetes, 1982). It has been established that delayed emergence occurs at temperatures below 23°C and ceases altogether below temperatures of 13-16°C for males and 18-21°C for females (Bowden, 2009). Additionally, it has been showed that rainfall resulted in decreased
adult emergence, and increased mortality of pupae. After emergence, male midge swarm around the panicles from which they have emerged and mate with one or more females before the later-emerging females can expand their wings and leave the spikelet surface (Bowden, 2009).

Females generally mate only once, and each female is capable of laying between 20-150 eggs (Teetes, 1985; Hallman et al., 1984a). Egg laying occurs at temperatures between 20-40°C at 8-11 am on sunny days with a peak period of oviposition usually occurring around 10 am (Modini et al., 1987). Females probe a suitable spikelet before remaining very still for approximately two seconds while an egg is laid. This procedure is repeated until the female is exhausted and the full quota of eggs is seldom deposited before the death of the female. Both males and females seldom live for more than 24 hours. However, this may be extended to a couple of days in field cage conditions when given water or dilute honey. Each female produces unisexual progeny at a ratio of near 1:1 female to male broods (Teetes, 1985).

At mean daily temperatures of approximately 24°C, eggs hatch after 2-3 days followed by larval periods of 10-12 days. The orange sausage shaped eggs hatch into transparent white first instar larvae which begin feeding on the glumes, or lemmas, and later move down to lie lengthwise against the developing caryopsis with their heads towards the base of the caryopsis. Larval feeding on the caryopsis creates a large depression in the normally plump filling caryopsis in which a fully fed larva reverses itself and pupates (Figure 3). One larva feeding on the developing caryopsis is sufficient to prevent seed kernel development. Within each single
seeded spikelet it is common for more than one egg, and subsequently more than one larva to develop. An average of 2.3 larvae would survive to the pupal stage under higher initial egg densities, and therefore that larval competition while feeding on the developing caryopsis is a significant mortality factor. Pupation occurs in the aborted spikelet and is followed in approximately 3-5 days by adult emergence. The total lifecycle of midge varies according to temperature. Under filed temperatures of 18-29°C, and average soil temperatures of 33°C, the midge has a total life cycle of 16-20 days (Franzmann, 1993).

Figure 3: The life cycle of sorghum midge
This short generation time of 2-4 weeks allows many generations of midge to occur each season, which accounts for the buildup of extremely high midge densities where the flowering period of sorghum is extended by successive planting dates. The host plant species for sorghum midge have been reviewed by numerous workers (Baxendale et al., 1984; Passlow, 1965; Doering & Randolph, 1963; Passlow, 1958;). While numerous wild non-sorghum grasses have been reported as hosts of the sorghum midge, both Franzmann (1996) and Harris (1979) concluded that only members of the genus Sorghum are hosts for the sorghum midge. These include Johnson grass (S. halapense), Columbus grass (S. almum), and Sudan grass (S. sudanense).

Johnson grass, Sorghum halepense (L.) Persoon, has been recognized as important in the population dynamics of the sorghum midge. Harris (1961) in Nigeria (Africa), Teetes (1985) in Texas (USA) and Franzmann et al., (2006) in Queensland (Australia) reported that Johnson grass acts as an early season host for midge. Peak adult emergence of over-wintering midge coincides with the widespread flowering of Johnson grass in an area after spring rains. Johnson grass acts as a host for two to three generations, and then midge migrate to infest flowering grain sorghum panicles.

Flowering grain sorghum crops are most susceptible to midge egg lay during anthesis. Female midges are attracted to sorghum crops via a combination of visual and chemical stimuli. Sharma et al., (1990) found that midges were attracted to yellow and white traps compared to darker colors such as blue or green. The chemical and physical properties of flowers play a major role in
attracting and stimulating oviposition by sorghum midge. Sharma & Hariprasad (2002) found in laboratory choice tests that sorghum midge females were more readily attracted to yellow, red and green colors combined with odors than by color, or odour alone. Under natural conditions sorghum midge lay approximately 90% of their eggs in flowering spikelets, about 10% in post flowering spikelets, and less than 0.5% of eggs in pre-flowering spikelets (Hallman et al., 1984a). Certain sorghum varieties may be particularly susceptible to egg lay in pre-flowering spikelets. Female midges have great difficulty laying in panicles after anthesis and under caged conditions. Second generation midge may re-infest post flowered aborted sessile spikelets in small numbers (Sharma, 2004).

2.6 Integrated pest management of sorghum midge

In the 1950’s a new theory ‘integrated control’ was developed to manage insects in response to control problems that arose from the complete reliance on one chemical method of control (Kogan, 1998). The term was later modified and described as Integrated Pest Management (IPM), an ecological approach to crop protection in which different control methods are used to enhance each other, and together contribute to sustainable insect management practices over the long term.

The international lesson from the history of insect management is that no single technology or method of control is likely to be ideal, and sustainable (Binns & Nyrop, 1992). One of the main aims of IPM is to reduce pesticide usage. Integrated pest management in practice relies on
benign tactics such as cultural practices, biological control, and plant resistance to maintain fluctuating pest populations below economic injury levels. Such an integrated management approach may be adopted to control midge damage in grain sorghum, including cultural, chemical, and biological means of control.

Cultural practices include: the destruction of Johnson grass to eliminate the early season host; the destruction of aborted sorghum spikelets within the soil containing diapausing larvae; the planting of uniform varieties with adequate fertilizer and moisture to ensure even flowering; and the early planting of sorghum to escape high midge numbers. Of these practices the early, uniform planting of sorghum crops is the most effective method of control, but lack of soil moisture or excessive rain can often mean that this is not possible (Tao et al., 2003; Teetes, 1985). Where environmental conditions cause delayed and successive planting of crops, the flowering period is usually extended, resulting in large midge infestations (Baxendale & Teetes, 1981). Under such conditions the use of chemical control may be required. Sprays are most effective on adult midge as sorghum midge larvae within sorghum spikelets are protected within the glumes. Insecticide applications are used to control ovipositing females, which emerge in the morning and die in the afternoon. Application over the entire flowering period is expensive for a low-input crop such as sorghum and may be only moderately effective, and has negative effects on the environment.
The presence of chemical sprays may adversely affect any biological control present. Larval parasitoids are known to provide some control of sorghum midge larvae (Baxendale et al., 1983; Young & Teetes, 1977). The parasitism of midge larvae in late maturing crops may be as high as 24.2%. However an average of 14.1% parasitism is of little economic significance, and such parasitoid populations build up only after yield damage has occurred. While other predatory insects such as lacewings, pirate bugs and spiders have been commonly observed controlling midge, their economic value is difficult to measure and thought to be minor.

Plant resistance to insects can be an important component in IPM strategies, and the tactic has wide applicability and function (Kennedy et al., 1987). There has been much effort placed into the search for and development of midge-resistant hybrids over the last three decades. Host plant resistance has many advantages over other methods of insect control. These include simplified management, decreased chemical spray costs, and the added benefit of biological control under decreased chemical usage. The use of genetically midge resistance sorghum hybrids is the most promising control strategy of sorghum midge. The addition of midge resistant varieties simplifies management, decreases costs and provides greater midge control, ultimately resulting in greater profitability (Groot & Dicke, 2002). It had been established by Teetes et al., (1986) that the sorghum midge resistant sorghum hybrids gave higher yields and greater returns than susceptible hybrids under the same insecticidal spray regime. For a range of susceptible sorghum hybrids, 1.4 to 1.5 grams of grain (42-48 kernels) are destroyed by the progeny produced by one female per panicle, an approximate 4% of the total kernels in each panicle. Similarly, each insect
infesting resistant sorghum hybrids destroys 0.32 to 0.4 grams of grain (9 kernels), less than 1% of the total kernels in each panicle (Hallman et al., 1984b).

2.7 Plant-insect herbivore interactions

In spite of the vulnerability of plants to diverse biotic and abiotic stresses, they dominate much of the land surface. To effectively combat these stresses, plants have consequently evolved complex defense responses that are based on constitutive or induced mechanisms (Figure 4).

![Figure 4: A model of recognition of herbivore feeding by resistant and susceptible plants](image)

The modified figure was adapted from Smith & Boyko (2007).
Constitutive mechanisms are features that are present in the plant prior to attack while induced defenses are activated upon attack, and could either be local or systemic (Schoonhoven et al., 2005). Induced expression of resistance traits increases plant fitness in the environments that harbor a variety of plant parasites. The inducibility of plant resistance was first reported in fungal and bacterial pathogens in the early 1900s and much later inducible defenses were shown to exist also against insect herbivores (Cornish-Bowden & Storer, 1986).

The responses of plants to insect infestation have in the recent past been profiled in several plants using molecular methods. In particular, identification of plant genes differentially expressed in response to herbivory attack has been undertaken in *Nicotiana* (Hui et al., 2003; Hummelbrunner & Isman, 2001); *Arabidopsis* (De Vos et al., 2006; Reymond et al., 2000), *Populus* (Peters & Constabel, 2002), *Picea* (Ralph et al., 2006) and *Solanum* (Lawrence & Novak, 2006). From these studies and others, it has been established that insect infestation induces defense mechanisms that are important in direct defenses which may inhibit the growth or development of the insect. The indirect defenses consisting of volatiles emitted from plants serve as airborne signals that deter insect herbivores or attract predators and parasites of the herbivores.

### 2.8 Signaling pathways in plant defense

The induced plant defenses against insects involve phytohormone-mediated signal transduction pathways that link the damage with the phenotypic change in the plant. There are three main signal transduction pathways that underlie induced defenses; the jasmonate pathway, the
shikimate pathway, and the ethylene (ET) pathway, characterized by the phytohormones jasmonic acid (Harding et al., 1997), salicylic acid (SA), and ET, respectively (Dicke & Van Poecke, 2002; Kessler & Baldwin, 2002). Of these pathways, the jasmonate pathway seems to play a leading role in insect-induced and wound-induced plant responses both in terms of direct and indirect defense (Thines et al., 2007; Howe, 2004; Dicke & Van Poecke, 2002; Kessler & Baldwin, 2002). However, not all herbivores induce JA in plants. The silverleaf whitefly, *Bemisia tabaci*, suppresses JA-dependent defenses and induces SA-dependent defenses (Kempema et al., 2007). Individual attackers therefore elicit distinct phytohormone signatures consisting of dynamic phytohormone induction patterns.

### 2.9 Global transcriptome changes in response to insect herbivory

In response to insect herbivory, plants undergo an extensive rearrangement of gene transcription. Hundreds of genes can be up- or down-regulated. Since plants response to insect herbivory is specific, there are considerable differences in the transcriptome response of a plant to different attackers (De Vos et al., 2006; Voelckel et al., 2004), as well as responses of different plant genotypes in response to the same herbivore. For example, two white cabbage (*Brassica oleracea var. capitata*) cultivars differ considerably in the global gene expression patterns induced by *Pieris rapae* caterpillar attack (Broekgaarden et al., 2007).
2.10 Differential defense responses to mechanical wounding and insect herbivory feeding

Despite the commonality between mechanical wounding and insect-feeding damage, plants can distinguish damage done mechanically from that done by insects. Inevitably, insect feeding causes wounding of the plant, but little is known about how plants distinguish and respond to the very different threats posed by mechanical wounding and insect herbivory. Although reports show that some genes or proteins can be activated by both mechanical wounding and insect attack (Sarosh & Meijer, 2007; McKay, 2003), other observations have revealed responses that are induced specifically or activated more rapidly by damage from insects. Differences have been observed in the expression of several wound-induced genes (Korth & Dixon, 1997) and also in the release of volatiles (Pare & Tumlinson, 1999). A DNA microarray study with a set of preselected defense-related Arabidopsis genes showed that caterpillar-elicited transcriptional changes differed from those of wounding by forceps. Insect chewing also accelerates the accumulation of wound-induced transcripts (Sarosh, & Meijer, 2007).

Further, unique to insect feeding, plants are able to initiate indirect defense by synthesizing specific blends of volatiles that attract natural enemies of their aggressors (Pare & Tumlinson, 1999; Alborn et al., 1997), and, by forming special tissues that impede the entry into the plant tissue (Truitt et al., 2004). The bias between mechanical wounding and insect herbivory is thought to be due to recognition of insect derived elicitors by plant cells (Sarosh, & Meijer, 2007).
2.11 Insect response to plant defense

Given the enormous variety of plant species that insects feed on, they have a wide variety of digestive proteases. These proteases break down dietary proteins into simple peptides and amino acids, which are then absorbed into the hemolymph. Proteases can be classified into serine, cysteine, aspartate and metallo proteases, all of which have been found in insect guts (Terra & Ferreira, 1994). The insect digestive tracts can adapt successfully to deal with various toxins and anti-metabolites in their diets. Some insects adapt to the protease inhibitors by overproduction of existing digestive proteases, while others selectively induce inhibitor-insensitive proteases (Ahn et al., 2004; Mazumdar-Leighton & Broadway, 2001; Jongsma & Bolter, 1997).

Insect herbivores are potentially vulnerable to plant secondary metabolites, in addition to plant defense proteins. To evade this, insects hijack plant defense mechanisms, making the plants susceptible to their attack (Hopkins et al., 2009). For instance, in Brassicaceae family, plants synthesize glucosinolates, a class of secondary compounds that shares a core consisting of a β-thioglucose moiety and a sulphonated oxime. Specific side-group structures distinguish one glucosinolate from another. When intact, they have little toxicity. Plant myrosinase, a β-thioglucosidase, can remove the β-glucose from glucosinolates, resulting in the formation of an unstable intermediate product. This product then breaks down to a variety of toxic products, which deter herbivore feeding (Hopkins et al., 2009; Bones & Rossiter, 1996). In undamaged tissue, myrosinase is stored separate from glucosinolates. When plant tissue is damaged, glucosinolates come into contact with plant myrosinase. This glucosinolate-myrosinase complex
is however disarmed by diamondback moth, *Plutella xylostella* (L.). The insect induces glucosinolate sulphatase to desulphate glucosinolates, which are invisible to myrosinase (Hopkins *et al.*, 2009; Ratzka *et al.*, 2002). The action of this enzyme can effectively compete with myrosinase for glucosinolate substrates and prevent the formation of toxic hydrolysis products arising from this plant defense system.

### 2.12 Genomics techniques for the study of gene expression

The ability to sequence entire genomes has stimulated research directed not only at producing DNA sequence, but also at defining the function of genes on a genome-wide level. Given that genes with related functions are likely to be regulated together, techniques that evaluate global gene expression provide a mechanism for the initial identification and clustering of novel gene sequences with related functions.

The techniques for the evaluation of gene expression have progressed from methods developed for the analysis of single, specific genes (e.g., Northern, slot, and dot blotting; semi-quantitative and quantitative-PCR; and nuclease protection assays) to techniques focused on identifying all genes that differ in expression between or among experimental samples (e.g., subtractive hybridization, sequencing of expressed sequence tags, serial analysis of gene expression, hybridization to microarrays and differential display).
2.12.1 Subtractive Hybridization

Subtractive hybridization methods were first described in the early 1980s for the purposes of creating cDNA libraries and generating probes of differentially expressed genes (Sargent and Dawid, 1983). Differentially expressed genes were isolated by hybridizing cDNA representing one sample (tester) to an excess of mRNA representing a second sample (driver). Transcripts expressed in both the tester and driver would form a mRNA/cDNA hybrid molecule, whereas a cDNA sequence unique to the tester would remain single-stranded. Single- and double-stranded molecules were separated using hydroxylapatite chromatography, and the eluted single-stranded cDNA represented genes expressed only in the tester. Differentially expressed cDNA could then be cloned or used directly as probes for screening libraries. Two significant limitations of the original protocol were the requirement of large quantities of mRNA and a bias against the identification of rare transcripts (Blumberg & Belmonte, 1999).

Modifications to the original protocol include the production of cDNA with biotin or oligo(dT)30-latex (Hara et al., 1991) to improve the separation of single- and double-stranded molecules. Amplification of selected cDNA by PCR has also been incorporated to decrease the initial amount of mRNA required and to improve the cloning efficiency of selected transcripts (Hara et al., 1991). Suppression subtractive hybridization (SSH) described by Rebrikov et al., (2004) is a method designed to selectively amplify differentially expressed transcripts while suppressing the amplification of abundant transcripts, thus eliminating the need to separate single- and double-stranded molecules. In addition, SSH normalizes target transcripts to approximately equal abundance.
Subtractive hybridization is the first technique to be widely used for the purpose of identifying differentially expressed genes on a global scale. Advantages of the technique include the ability to isolate genes with no prior knowledge of their sequence or identity and the use of common molecular biology techniques that do not require specialized equipment or analyses. Several limitations of the original protocols, such as requirements of large quantities of RNA and bias toward abundant genes, have been overcome by incorporation of PCR into the SSH technique (Pääkkönen & Tjäderhane, 2009). However, SSH remains applicable only to pair-wise treatment comparisons and must be replicated with the tester and driver reversed to identify gene expression changes in both directions. Additionally, subtractive hybridization does not provide a quantitative measure of expression differences and is most efficient at identifying genes that are completely absent, rather than expressed less abundantly, in the driver sample (Rebrikov et al., 2004; Moody, 2001).

2.12.2 Expressed Sequence Tag Sequencing

Technological advances facilitating high-throughput sequencing led to the concept of expressed sequence tag (EST) libraries in the early 1990s (Adams et al., 1991). The EST sequences are generated by randomly picking clones from a cDNA library and performing a single sequencing reaction to produce 300 to 500 bp of sequence per clone. Differences in gene expression may be identified by counting the number of times a particular sequence appears in EST libraries of genes from different sources. However, EST sequences are often generated from cDNA libraries that have been normalized to equalize the abundance of clones representing different transcripts (Hackett et al., 2005). The EST sequenced from normalized cDNA libraries may be compared to
identify transcripts that are expressed in one library and absent in another, but accurate quantitative data describing the relative abundance of genes within a library can only be obtained from EST generated from non-normalized cDNA libraries. Despite this restriction, the power of discovering differences in gene expression by comparing the abundance of sequences in EST databases increases with growth of the databases.

2.12.3 Serial Analysis of Gene Expression

Serial analysis of gene expression (SAGE) is a technique that is based on the concept that short sequence tags are sufficient to identify a gene transcript, provided the tags represent a known location within the gene (Hackett et al., 2006). A SAGE tag is generated from an mRNA transcript by extracting a short sequence from a defined location. Typically, a SAGE tag includes nine bases of sequence downstream from the last endonuclease recognition site of a transcript. Multiple SAGE tags are ligated together in a cloning vector so that a typical sequencing reaction of 300 to 500 bp generates the sequences of 20 to 30 SAGE tags. Because each unique SAGE tag theoretically represents a unique mRNA transcript, an overview of all genes expressed in the original RNA sample is generated by SAGE tag sequences. Differences in gene expression among experimental samples can then be identified by comparing the relative abundance of specific SAGE tags in different libraries. Genes or EST sequences represented by the SAGE tags are identified by searching sequence databases to identify gene or EST sequences that contain the SAGE tag in the appropriate location. Important advantage of SAGE over subtractive hybridization and differential display is that SAGE data are quantitative and cumulative.
One limitation of SAGE is identification of the genes represented by SAGE sequence tags. In addition, certain transcripts may fail to be represented by SAGE tags depending on the specific enzymes used to generate the SAGE library and other transcripts may be represented by multiple SAGE tags because of single nucleotide polymorphisms or alternative splicing of transcripts (Hackett et al., 2006).

2.12.4 Microarray Hybridization

The evaluation of gene expression using microarray technology that was originally described by Stears et al. (2003) an effective approach used to measure the expression levels of a large number of genes simultaneously. The essence of microarray technology is the parallel hybridization of a mixture of labelled nucleic acids called target, with thousands of individual nucleic acid species called probes, which can be identified by their spatial position in a single experiment. The location of a specific probe on the array is termed spot or feature. Whereas the probes are immobilized on a solid support, the targets are applied as a solution onto the array for hybridization after fluorescent labelling (Ehrenreich, 2006).

There are three general types of microarrays include oligonucleotide chips made by the synthesis of short oligos directly on a glass wafer, oligonucleotide arrays made by spotting pre-synthesized oligos onto glass slides or nylon membranes, and cDNA arrays, made by spotting PCR amplified inserts of cDNA library clones onto glass slides or nylon membranes (Stears et al., 2003).
Regardless of the experimental design, microarray analysis is based on the premise that the intensity of the resulting hybridization signal for any sequence on the microarray is proportional to the amount of mRNA corresponding to that sequence in the original mRNA sample. Relative expression of each sequence represented on the microarray is evaluated by comparing hybridization intensity signals generated by different experimental samples (Ehrenreich, 2006).

2.12.5 Differential Display

Techniques known in general as RNA fingerprinting include differential display (Liang & Pardee, 1992) and RNA fingerprinting by arbitrary primed PCR (RAP-PCR); Welsh et al., 1992). Both methods were introduced in 1992 and are based on PCR amplification of random subsets of genes from two or more RNA samples. Since the introduction of these techniques, differential display has been more widely applied than RAP-PCR. The first step of either differential display or RAP-PCR is to reverse-transcribe random subsets of mRNA to cDNA. In differential display, this is done using an anchored primer, which is typically a polyT oligonucleotide with one or two additional bases. These primers anneal to the polyA tail of mRNA, anchoring the resulting cDNA to the 3’ end of the RNA transcript. In contrast, RAP-PCR uses arbitrary primers in reverse transcription. These primers are typically 10 bp in length and may anneal to complementary sequence and prime reverse transcription from any point along an RNA transcript (Liang & Pardee, 1992; McClelland & Welsh, 1994).

Following reverse transcription of first-strand cDNA, segments of the cDNA transcripts are amplified using multiple PCR primer pairs. For both differential display and RAP-PCR, the
forward PCR primer is an arbitrary primer, typically 10 bp in length, that couples with the anchored (differential display) or arbitrary (RAP-PCR) primer used to produce the cDNA. It has been estimated that PCR products from 240 unique primer pair combinations would represent all mRNA that were present in the original RNA sample (Liang and Pardee, 1992). The PCR products may be labelled by incorporation of a radio-labelled nucleotide such as $^{33}$P-dCTP or by a fluorescently labelled primer (Ito et al., 1994).

Visualization of PCR products is achieved by electrophoresis on polyacrylamide gels followed by appropriate imaging. Images are then evaluated by comparing the relative intensities of bands produced from different experimental samples. Bands that are present in one sample and absent in another or bands that are present at different relative intensities across experimental treatments represent potentially differentially expressed mRNA transcripts. Typically, bands are evaluated only if they are amplified consistently in duplicate PCR reactions from the same experimental sample (Moody, 2001).

The final phase of differential display is to identify the sequence of the transcript represented by the differentially displayed PCR product and to confirm that the transcript is truly differentially expressed. These steps are accomplished by physically locating and excising the acrylamide gel containing the PCR product of interest, based on alignment of the imaged PCR products with the dried acrylamide gel. The PCR products are purified from the gel and reamplified by PCR. Various strategies and techniques have been used to confirm differential expression, including the use of reamplified PCR products as a probe for Northern hybridizations, spotting multiple
reamplified PCR products on membranes for reverse dot- or slot-blot analysis, and cloning and sequencing reamplified PCR products so that gene specific primers may be designed for use in semiquantitative PCR of cDNA. Regardless of the strategy followed to confirm differential expression, sequences of differentially expressed gene fragments will ultimately be required to begin to understand the function of a gene (Moody, 2001; Liang & Pardee, 1992).

Two important advantages of RNA fingerprinting methods relative to subtractive hybridization are the abilities to compare multiple experimental samples simultaneously and to identify genes that are either up or down-regulated in one sample relative to another. However, RNA fingerprinting shares with SSH the limitation that it is not a quantitative method. Additionally, RNA fingerprinting experiments are often plagued with a high rate of false positives, or gene fragments that seem to be differentially expressed as an artefact of PCR. Finally, investigation of all potentially differentially expressed genes requires high-throughput PCR and extensive investments of time and labour to follow up and confirm differential expression on an individual gene basis (Moody, 2001). Among the available differential display methods, one increasingly popular tool is complementary DNA-Amplified fragment length polymorphism (cDNA-AFLP, Bachem et al., 1996).

2.12.5.1 The cDNA-AFLP transcript profiling

The AFLP-based transcript profiling protocol (AFLP-TP) is an improved version of the cDNA-AFLP method described by Bachem et al., (1996). AFLP-TP is a fragment-based, genome-wide expression technique allowing the analysis of “all” genes involved in a particular biological process
or expressed under certain conditions. In the approach, unique transcript tags derived from the 3’ end region of expressed genes are PCR amplified and displayed on acrylamide gels. Selective amplification of subsets of transcript tags allows to fractionate the initial pool of tags and to detect low abundant messengers. Because it is PCR based and because of the fractionation, the sensitivity is significantly higher compared to other techniques including microarrays. Furthermore, as it is a fragment-based technique, it allows to discriminate between homologous sequences provided that distinct restriction fragments are obtained from the homologues (Vuylsteke et al., 2006).

The first step in the AFLP-TP approach is the conversion of mRNA into ds cDNA using a biotinylated oligo-dT primer which is bound to streptavidin coated wells (Figure 5). The cDNAs are digested with two restriction enzymes in a two-step reaction. After digestion with the first enzyme, the 5’ most regions are released in the reaction mixture and washed away. Digestion with the second enzyme releases the restriction fragments or transcript tags. The enzymes have been chosen such that the majority of the cDNAs is cut. The combination used in the protocol (BstYI and MseI) cuts around 80% of all messengers (based on in silico analysis using available full-length cDNA sequences). Around 10% of the fragments is however too small or too large to be displayed on acrylamide gels. Almost 70% of the fragments are BstYI/MseI fragments, while the other 10% are MseI/BstYI fragments. To screen all of them, the two series of analyses have to be done, with as first cutting enzyme BstYI and MseI, respectively (Vuylsteke et al., 2006).
BstYI recognizes the sequence PuGATCPy and after digestion and adapter ligation, a pre-amplification is done using a BstYI-primer with a T or C as 3’ nucleotide. In this way, the total pool of fragments is subdivided in two. In a subsequent amplification, selective primers are used. Increasing the number of selective nucleotides reduces the number of fragments amplified but results in amplification of fragments that are derived from low abundant messengers. Usually, primers with 1 or 2 selective nucleotides are used. In several plant species, a total of 3 selective nucleotides (e.g. BstYI+1 and Msel+2 primers) results in profiles that are not too dense, while the sensitivity is already high enough to detect low abundant messengers (Vuylsteke et al., 2006).

Figure 5: Schematic description of cDNA-AFLP
Abbreviations; S=susceptible, R=resistant, PAGE=polyacrylamide gel electrophoresis.
After generating profiles, differentially expressed genes can readily be detected and isolated from the gels for further characterization. Depending on the density of the gel pattern, good quality sequence can be obtained for 50% to 80% of the isolated fragments by direct sequencing (Vuylsteke et al., 2006).

The differences in the intensity of the bands that can be observed provide a good measure of the relative differences in the levels of gene expression. Further characterization of interesting transcripts often requires the identification of the corresponding full-length cDNA (Vuylsteke et al., 2006).
CHAPTER THREE
MATERIAL AND METHODS

3.1 Plant material and growth conditions

Two sorghum genotypes were used in this study; AF28 (sorghum midge resistant) and seredo (sorghum midge susceptible). The resistance and susceptibility of sorghum midge in these genotypes had been confirmed by Sharma et al. (1999), Harris (1976) and Rossetto et al. (1975). The experimental line AF28 is a landrace of guinea subspecies while Seredo is an improved variety of caudatum subspecies particularly belonging to the Dobb subspecies which is common in Eastern Africa. The seeds of AF28 and Seredo were kindly donated by International Crops Research Institute of the Semi-Arid Tropics (ICRISAT) and Kenya Seed Company, respectively.

The seeds for each genotype were surface sterilized for 30 minutes in 33% sodium hypochlorite followed by a brief rinse with 70% ethanol before planting in sterile soil. The plants were grown in controlled environment of a growth chamber under 12-hour photoperiod with fluorescent light (750 μmol E m⁻² s⁻¹) at 32 °C (day)/ 22 °C (night) and constant 50% humidity.

3.2 Experimental design

Rearing of the sorghum midge was done in Johnson grass at the Ecology Department, Swedish University of Agricultural Sciences, Uppsala, Sweden. The larvae of the sorghum midge was separated from their stock in Johnson grass and starved for 1 hour in empty Petri dishes prior to
the start of the time course experiments. One sorghum midge larvae was placed in the first floret of the flowering sorghum plant and allowed to feed before harvesting the florets at different time intervals of 3, 6, and 12 hours (Figure 6).

Figure 6: The experimental design
Sorghum florets were infested with sorghum midge larvae. The florets were excised and samples of RNA were isolated from the inoculated and control sorghums at the indicated time points.
Key: O = not infested and X = infested.
The harvested florets were immediately frozen in liquid nitrogen, and stored at -80 °C until use. Uninfected florets were used as control and were pooled and harvested at the same time intervals as the midge larvae infested florets.

3.3 Preparation of poly (A)⁺ RNA and cDNA synthesis

Total RNA was isolated from frozen floret tissue using the AURUM Total RNA Mini Kit (Bio-Rad, Hercules, U.S.A.), following the manufacturer’s described protocol with slight modifications. Briefly, the harvested floret tissues were ground into fine powder under liquid nitrogen. The tissue (60 mg) was transferred to 2 ml capped tubes and 700 µl of lysis solution was added and the solution disrupted vigorously by a rotor for 60 seconds, followed by centrifugation at 14000 revolutions per minute (rpm) for 3 minutes. The supernatant was transferred to new 2 ml tubes and 700 µl of ethanol (70%) was added to the supernatant followed by a 30 seconds homogenization of the resulting solution with a rotor. The lysate was transferred to RNA binding columns which had been inserted into 2 ml capless tubes, and thereafter centrifuged at 14000 rpm for 60 seconds. The RNA binding columns were removed from the wash tubes and the filtrate in the wash tubes was discarded, before replacing the columns into the same tubes. The same procedure was repeated for the remainder of the homogenised lysate. The bound RNA was first washed with 700 µl low stringency (4 volumes of 100% ethanol and 1 volume of 5X low stringency solution provided) solution by centrifuging the columns at 14000 rpm for 30 seconds and the supernatant was discarded.
To remove any traces of DNA, 80 µl of DNase I was added to the membrane stack at the bottom of the columns. For digestion to occur, the columns were incubated at 25 °C for 15 minutes. When digestion was complete, the columns were centrifuged for 30 seconds at 14000 rpm. The bound RNA was then washed with 700 µl high stringency wash solution and the supernatant was discarded after centrifuging the columns for 1 minute. To elude the RNA, the columns was transferred to a 1.5 ml capped microcentrifuge tubes and 80 µl of pre-warmed elusion solution (70 °C) was pipetted onto the membrane stack at the bottom of the RNA binding columns, and 1 minute was allowed for the elution solution to saturate the membranes before centrifuging the columns at 14000 rpm for 2 minutes to elude the RNA. The isolated RNA was thereafter quantified by absorbance at A_{260} using ND-1000 Nanodrop Spectrophotometer (Nanodrop Technologies, Wilmington, USA).

Poly(A)^+ RNA was prepared using the mRNA Capture Kit (Roche Applied Science, Mannheim, Germany) according to the manufacturer’s instructions. The first strand of the cDNA was synthesized from 5 µg total RNA. The reverse transcription was started by adding 50 µl of first-strand cDNA synthesis mixture (31.2 µl water, 10 µl 5X first-strand buffer (Roche), 5 µl 0.1 M dithiothreitol (DDT), 2.5 µl 10-mM dNTPs and 1.2 µl 200 U/µl SuperScript III (Invitrogen, San Diego, U.S.A.). The reaction mixture was incubated for 1 hour at 50 °C.

The second strand of the cDNA was synthesized by adding 91 µl water, 16 µl 10x *E. coli* ligase buffer, 6 µl 0.1 M DDT, 3 µl 10 mM dNTPs, 1.5 µl *E. coli* ligase, 1.5 µl (5 U), *E. coli* DNA
polymerase I, and 0.5 μl RNase H (Fermentas) and the mixture was incubated for 1 hour at 12 °C and subsequently for 1 hour for 22 °C in a thermocycler. After incubation, the reaction mixture was discarded leaving the cDNA attached to the tube wall. The cDNA was washed three times with 200 μl wash buffer (Roche) by incubating the sample for 2 minutes at room temperature and liquid was subsequently removed by gently pipetting the liquid out.

3.4 AFLP procedures

The subsequent AFLP analysis was performed as described by Vuylsteke et al. (2006) using BstYI and MseI (New England Biolabs, Beverly, USA) restriction enzymes but with slight modifications. The first cDNA digestion was started by adding a mixture containing 38.8 μl water, 10 μl 5X restriction ligation (RL) buffer [the RL buffer was made by mixing 50 mM Tris-HAc (pH 7.5), 50 mM MgAc, 250 mM KAc, 25 mM DTT] and 1.2 μl BstYI (10 U/μl) restriction enzyme. After incubation for 3 hours at 60 °C, the tubes were washed three times with 100 μl wash buffer (Roche). Digestion with the second enzyme to release the transcript tags was initiated by adding a mixture of 38.8 μl water, 10 μl RL buffer, and 1.2 μl MseI (10 U/μl) restriction enzyme (New England Biolabs) followed by incubation for 3 hours at 60 °C.

After digestion, the restriction fragments were ligated with the following adapters: BstYI-F: 5′-CTCGTAGACTGCGTAGT-3′; BstYI-R: 5′-ATCACTACGCAGTCTAC-3′; MseI-F: 5′-GACGATGAGTCCTGAG-3′; and MseI-R: 5′-TACTCAGGACTCAT-3′ (All adapters were
non-phosphorylated oligonucleotides). The ligation mixture [1 μl (5 pmol) BstYI adapter, 1 μl (5 pmol) MseI adapter, 0.1 μl 100 mM ATP (Pharmacia Biotech, Piscataway, U.S.A), 2 μl 5 x RL-buffer, 1 μl T4 DNA ligase (5 U/μl; Fermentas, Burlington, Canada), 0.5 μL (10 U/μl) BstYI, 0.5 μl 10 U/μl) MseI and 3.9 μl of water] was added to the restriction fragments and the reaction incubated for 3 hours at 37 °C. The ligated sample was diluted 2-fold with 1 mM Tris (pH 8.0) and 5 μl was used as a template in the pre-amplification procedures. Pre-amplification was performed by a MseI primer (MseI: 5’- GATGAGTCCTGAGTAA - 3’) without a selective nucleotide combined with a BstYI primer (BstYI: 5’- GACTGCGTAGTGATC(T/C)NN - 3’) containing either T or C at the 3’ end. The pre-amplification reaction was carried out using 20 cycles (94 °C for 30 seconds; 56 °C for 1 minute; 72 °C for 1 minute).

The amplified reaction was diluted 300-fold and 5 μl was used for final selective amplification using a touchdown amplification program described by Vos et al. (1995). The BstYI+2 and MseI+2 primers with one selective nucleotide respectively were used for the cDNA-AFLP analysis and all 12 primers combinations were performed (Table 1). Selective 33P labeled amplification products were separated on 4.5% polyacrylamide gels using the Sequigen system (Bio-Rad) run at constant power (100 Watt) until 4,300 V was reached. Gels were dried onto 3mm Whatman paper for 1.5 hour before exposing to Amersham Hyperfilm film (GE Healthcare, Buckinghamshire, England) for 10 days.
3.5 Isolation, quantification and sequencing of transcript derived fragments

Gel profiles were quantified using Quantity One gel image analysis software (version 4.4.1, Bio-Rad) resulting in measurement of band intensities per lane for each time interval. Lane-based background subtractions were carried out to remove background intensity from lanes. The bands were normalized to compensate for the differences in any loading effects in the different lanes of the gel. The intensity of each band was quantified using volume units (intensity $\times$ mm$^2$).

<table>
<thead>
<tr>
<th>Forward primer ($Bs$TI+2)* 5'-3'</th>
<th>Reverse primer ($Mse$I+2) 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GACTGCGTAGTGATCTTTA</td>
<td>GATGAGTCCTGAGTAAG</td>
</tr>
<tr>
<td>GACTGCGTAGTGATCTTTC</td>
<td>GATGAGTCCTGAGTAAG</td>
</tr>
<tr>
<td>GACTGCGTAGTGATCTTTG</td>
<td>GACTGCGTAGTGATCTTTA</td>
</tr>
</tbody>
</table>

*All $Bs$YI+2 were labeled with $^{33}$P prior to selective amplification. The $Bs$TI+2 and $Mse$I+2 primers were combined so as to give 12 primer combinations without any repetition.

Bandsof interest were marked and excised from the dried gel. The excised piece of gel was incubated in 100 µl of water and boiled for 15 minutes. The solution was then cooled on ice and 10 µl 3 M NaAc, pH 5.2, 2µl 20 mg/ml glycogen and 200µl 96% ethanol was added and incubated overnight at -20°C. After precipitation and drying, the cDNA was dissolved in 10 µl of water. 1 µl of eluted DNA was re-amplified using the $Bs$YI (T)-0, $Bs$YI(C)-0 and $Mse$I-0
primers. For PCR, the same reaction conditions as in the pre-amplification were used. After re-amplification, the PCR products were checked in 1% agarose gel and ligated into a pJET1.2 cloning vector (Fermentas) and transformed to DH5α cells. Sequencing of the cloned fragments was done using the pJET1.2 forward/reverse sequencing primers (Fermentas).

3.6 Reverse transcription PCR

Total RNA samples were DNase I-treated prior to reverse transcription PCR (RT-PCR). First strand cDNA was synthesized in a 20 μl reaction mixture with a RevertAid First Strand cDNA Synthesis Kit (Fermentas) according to manufacturer’s protocol, using 5 μg total RNA isolated as previously stated. A 20 μl reaction mixture containing 5 μg total RNA and 15-20 pmol specific primers, 4 μl 5X reaction buffers, 1 μl RiboLock RNase inhibitor (20 U/μl), 2 μl dNTP mix (10mM), RiverAid H Minus M-MuLV Reverse Transcriptase (200U/μl) and diethylprocarbonate (DEPC) water was gently mixed, and the mixture incubated for 60 minutes at 42°C. The reaction was terminated by heating the mixture at 70°C for 5 minutes. The specific primers for the control gene and the transcripts were designed using Primer3 (v. 0.4.0) software (Steve & Skalesky, 2000). Thermo-cycling parameters were as follows: after an initial denaturation at 94°C for 2 minutes, samples were subjected to a cycling regime of 30 cycles at 94°C for 30 s, 57°C for 30 s, and 72°C for 1 minute. At the end of the final cycle, an additional extension step was carried out for a further 5 minutes at 72°C.
Quantitative real-time-PCR (qRT-PCR) analyses were performed using the same pooled samples used in the cDNA-AFLP analysis. To remove any traces of genomic DNA, the total RNA was treated with DNase I (Invitrogen) according to the manufacturer's instructions. DNA-free total RNA (5 μg) was converted into cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas) as previously described. The cDNA was then quantified by PCR. Primer sequences for the selected genes were designed using Primer3 (v0.4.0, http://frodo.wi.mit.edu/, Table 2) based on cDNA sequence of interest. Sorghum α-tubulin gene (forward primer: 5'-GACTTGGAGCCACTGTGAT-3' and reverse primer: 5'-GCAGCATCTTCTTGTGTGA - 3') were used as the reference gene control. All primers were synthesized by Invitrogen.

Quantitative RT-PCR analysis was done in optical 96-well plates using a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). SYBR Green was used to monitor dsDNA synthesis. Each reaction contained 10 μl 2X Maxima SYBR green qPCR Master mix (Fermentas), 5 ng cDNA, and 0.2 μM of each gene-specific primer (forward and reverse) in a final volume of 20 μl. All qRT-PCR reactions were performed in triplicate. To standardize the data, the amount of target gene was normalized over the abundance of the constitutive sorghum α-tubulin. The following PCR program was used for all PCR reactions: 95°C for 10 minutes; 40 cycles of 95°C for 30 seconds and 60°C for 45 seconds.
To detect primer dimerisation or other artifacts of amplification, a melting curve analysis was performed immediately after completion of the real-time PCR by incubating the reaction at 95°C for 15 seconds, annealing at 55°C for 20 seconds and then slowly increasing the temperature to 95°C at a 2% ramp rate, with continuous measurement of fluorescence. Dissociation curves were generated for each reaction to ensure specific amplification. The threshold cycle (CT) values generated from the ABI PRISM 7500 Software Tool were employed to quantify relative gene expression using comparative $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen, 2001). Relative expression levels for each of the selected genes in non-infested floret tissues were also computed using the comparative CT method (Livak & Schmittgen, 2001).

### Table 2: Primers used for both RT-PCR and qRT-PCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize protease inhibitor</td>
<td>TGGAGGAAGCAAGAGAAGGTG</td>
<td>CACAAAATCTGCAGTCACAA</td>
</tr>
<tr>
<td>Cinnamoyl-CoA</td>
<td>GTTGTGTGTCTCCTCGTCCTGTCT</td>
<td>GGTGCTGCAGTCCTCCTCAC</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>GCCAAGAAACCATCCTCAA</td>
<td>ACGACGTGTCGACGAAGA</td>
</tr>
<tr>
<td>α-Tubulin</td>
<td>GACTTGGAGCCACTGTGAT</td>
<td>GCAGCATCTCTCCTGTGTGA</td>
</tr>
</tbody>
</table>
CHAPTER FOUR

RESULTS

4.1 cDNA-AFLP analysis of differentially expressed genes

The cDNA-AFLP templates prepared from florets infested with sorghum midge larvae using 12 different primer combinations of BstYI and MseI primers having 1–2 selective nucleotides at their 3’-ends, resulted in the detection of approximately 6,720 cDNA fragments with an average of 40 discrete and clearly visible bands amplified with a given cDNA-AFLP primer combination (Figure 7). Amplification products size ranged from 40 to 500 bp. Overall, cDNAs isolated from the AF28 and Seredo sorghum varieties displayed almost identical patterns on the polyacrylamide gel with a given primer combination. However, a comparison of cDNA-AFLP patterns revealed that: of all the bands, 118 were novel, 17 up-regulated and 23 were down-regulated genes. Ninety-three of the 158 differentially expressed fragments were successfully cloned into pJET1.2 cloning vector. Attempts to isolate and clone the remaining fragments were unsuccessful because of their small quantities.

4.2 Characterization of Sorghum midge responsive-genes in sorghum

To identify and characterize the putative genes involved in resistance against sorghum midge, a NCBI’s BlastX and BlastN search was performed in the database for fragments which had more than 100bp and these results are summarized in Table 3 below. The sequences were compared to
those in the NCBI and Phytozomev4.0 databases and broadly characterized according to their homology and the putative function with known nucleotide and protein sequence.

Figure 7: A typical cDNA-AFLP gel
Expression of sorghum and sorghum midge transcripts displayed by cDNA-AFLP. Key: a- novel transcript in AF28 and b- common transcript between AF28 and Serodo after of selective amplification. Abbreviations: L=ladder, Ac=AF28 control, Ai=AF28 infested, Sc=Seredo control and Si=Seredo infested at 0, 3, 6 and 12 hours time points.
Characterization of Sorghum midge responsive-genes in sorghum

BlastX and BlastN search was performed in the database for fragments which had more than 00bp and these results are shown in Table 3 below. The sequences were compared to those in the NCBI and Phytozomev4.0 databases and broadly characterized according to their homology and the putative function with known nucleotide and protein sequence. A total of 53 genes regulated in response to sorghum midge feeding were listed and categorized into nine groups encoding proteins functioning in defense, signal transduction, cell wall fortification, oxidative stress/stress, photosynthesis, development, cell maintenance, abiotic stress, and unknown function.

Table 3: Summary of differentially expressed fragments generated by cDNA-AFLP between AF28 and Seredo upon sorghum midge infestation and homology of differentially expressed fragments with nucleotide or protein sequences in the NCBI database

<table>
<thead>
<tr>
<th>DNA-AFLP fragment ID</th>
<th>Size (bp)</th>
<th>GenBank Homolog</th>
<th>Species</th>
<th>Blast E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abiotic stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0G12</td>
<td>134</td>
<td>HEAT shock protein</td>
<td>O. sativa</td>
<td>1e-72</td>
</tr>
<tr>
<td>0G22</td>
<td>189</td>
<td>Aquaporin</td>
<td>S. officinarum</td>
<td>1e-54</td>
</tr>
<tr>
<td>1G22</td>
<td>175</td>
<td>Glycosyltransferase (pglcat7 gene)</td>
<td>S. bicolor</td>
<td>2e-70</td>
</tr>
<tr>
<td>Cell maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5G22</td>
<td>378</td>
<td>50S ribosomal protein L33</td>
<td>Z. mays</td>
<td>1e-25</td>
</tr>
<tr>
<td>8G11</td>
<td>549</td>
<td>Vesicle transport v-SNARE 13</td>
<td>Z. mays</td>
<td>4e-82</td>
</tr>
<tr>
<td>DNA-AFLP Segment ID&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Size (bp)</td>
<td>GenBank Homolog&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Species</td>
<td>Blast E-value</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>------------------------------</td>
<td>---------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Cell wall fortification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10G11</td>
<td>339</td>
<td>S-adenosyl-L-methionine:phosphoethanolamine N-methyltransferase</td>
<td><em>Z. mays</em></td>
<td>4e-38</td>
</tr>
<tr>
<td>13G23</td>
<td>127</td>
<td>Caffeic acid 3-O-methyltransferase</td>
<td><em>S. bicolor</em></td>
<td>1e-57</td>
</tr>
<tr>
<td>13G43</td>
<td>297</td>
<td>Membrane protein</td>
<td><em>Z. mays</em></td>
<td>1e-127</td>
</tr>
<tr>
<td>12G11</td>
<td>502</td>
<td>d-TDP-glucose dehydratase</td>
<td><em>P. australis</em></td>
<td>7e-89</td>
</tr>
<tr>
<td><strong>Defense related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10G11</td>
<td>214</td>
<td>Transcriptionally-controlled tumour protein</td>
<td><em>O.sativa</em></td>
<td>1e-140</td>
</tr>
<tr>
<td>11G11</td>
<td>239</td>
<td>EDS1-like protein</td>
<td><em>Z. mays</em></td>
<td>2e-160</td>
</tr>
<tr>
<td>16G21</td>
<td>175</td>
<td>Maize protease inhibitor</td>
<td><em>Z. mays</em></td>
<td>2e-07</td>
</tr>
<tr>
<td>16G13</td>
<td>194</td>
<td>Lipase class 3</td>
<td><em>O.sativa</em></td>
<td>1e-09</td>
</tr>
<tr>
<td>10G13</td>
<td>322</td>
<td>Subtilisin/chymotrypsin inhibitor CI-1B</td>
<td><em>Z. mays</em></td>
<td>2e-70</td>
</tr>
<tr>
<td>15G12</td>
<td>102</td>
<td>Polyphenol oxidase</td>
<td><em>T. aestivum</em></td>
<td>1e-09</td>
</tr>
<tr>
<td>13G7</td>
<td>207</td>
<td>Thaumatin-like protein</td>
<td><em>Z mays</em></td>
<td>1e-14</td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11G13</td>
<td>375</td>
<td>Homocysteine S-methyltransferase 4</td>
<td><em>Z. mays</em></td>
<td>2e-17</td>
</tr>
<tr>
<td>19G1</td>
<td>122</td>
<td>Scarecrow-like 1</td>
<td><em>Z. mays</em></td>
<td>9e-11</td>
</tr>
<tr>
<td>20G12</td>
<td>255</td>
<td>Arginyl-tRNA synthetase</td>
<td><em>Z. mays</em></td>
<td>2e-33</td>
</tr>
<tr>
<td>13G21</td>
<td>135</td>
<td>Histone H3</td>
<td><em>Z. mays</em></td>
<td>1e-69</td>
</tr>
<tr>
<td>18G12</td>
<td>172</td>
<td>Auxin-induced protein</td>
<td><em>S. officinarum</em></td>
<td>4e-14</td>
</tr>
</tbody>
</table>
Table 3 continued

<table>
<thead>
<tr>
<th>cDNA-AFLP fragment ID</th>
<th>Size (bp)</th>
<th>GenBank Homolog</th>
<th>Species</th>
<th>Blast E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1G241</td>
<td>250</td>
<td>Phosphoenolpyruvate carboxylase</td>
<td>Z. mays</td>
<td>2e-20</td>
</tr>
<tr>
<td>Oxidative burst/stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2G12</td>
<td>145</td>
<td>Peroxidase 2</td>
<td>V. vinifera</td>
<td>1e-16</td>
</tr>
<tr>
<td>U1G6</td>
<td>184</td>
<td>Quinone oxidoreductase</td>
<td>Z. mays</td>
<td>4e-53</td>
</tr>
<tr>
<td>Photosynthesis-related</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N5G21</td>
<td>199</td>
<td>NADP-isocitrate dehydrogenase</td>
<td>C. limon</td>
<td>1e-19</td>
</tr>
<tr>
<td>N5G13</td>
<td>307</td>
<td>Putative SecA protein</td>
<td>O. sativa</td>
<td>6e-42</td>
</tr>
<tr>
<td>N2G22</td>
<td>269</td>
<td>Chlorophyllide a oxygenase</td>
<td>Z. mays</td>
<td>2e-45</td>
</tr>
<tr>
<td>U2G7</td>
<td>163</td>
<td>Cyclophilin (peptidyl-prolyl cis-trans isomerase)</td>
<td>S. bicolor</td>
<td>2e-82</td>
</tr>
<tr>
<td>N1G243</td>
<td>602</td>
<td>Phosphoenolpyruvate carboxylase 2</td>
<td>H. vulgare</td>
<td>4e-39</td>
</tr>
<tr>
<td>Signal transduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1G11</td>
<td>223</td>
<td>ADP-ribosylation factor (GTPase-activating protein 3)</td>
<td>Z. mays</td>
<td>3e-48</td>
</tr>
<tr>
<td>N5G32</td>
<td>159</td>
<td>R2R3 Myb transcription factor</td>
<td>Z. mays</td>
<td>2e-19</td>
</tr>
<tr>
<td>N8G32</td>
<td>238</td>
<td>Leucoanthocyanidin dioxygenase</td>
<td>Z. mays</td>
<td>3e-106</td>
</tr>
<tr>
<td>N6G11</td>
<td>314</td>
<td>Wall-associated protein kinase</td>
<td>O. sativa</td>
<td>7e-123</td>
</tr>
<tr>
<td>DNA-AFLP fragment ID</td>
<td>Size (bp)</td>
<td>GenBank Homolog</td>
<td>Species</td>
<td>Blast E-value</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------</td>
<td>---------------</td>
</tr>
<tr>
<td>10G5</td>
<td>120</td>
<td>GTP-binding protein sarl</td>
<td><em>O. sativa</em></td>
<td>5e-85</td>
</tr>
<tr>
<td>12G4</td>
<td>223</td>
<td>ADP-ribosylation factor GTPase-activating protein 3</td>
<td><em>Z. mays</em></td>
<td>8e-54</td>
</tr>
<tr>
<td>7G22</td>
<td>350</td>
<td>Cinnamoyl-CoA reductase 2</td>
<td><em>Z. mays</em></td>
<td>5e-55</td>
</tr>
<tr>
<td>1G12</td>
<td>125</td>
<td>Ras-GTAase activating protein binding protein2</td>
<td><em>Z. mays</em></td>
<td>3e-33</td>
</tr>
<tr>
<td>4G3</td>
<td>308</td>
<td>Ankyrin-like protein</td>
<td><em>O. sativa</em></td>
<td>2e-65</td>
</tr>
<tr>
<td>3G2</td>
<td>249</td>
<td>Flavonol glucosyltransferase</td>
<td><em>S. bicolor</em></td>
<td>1e-10</td>
</tr>
<tr>
<td>1G32</td>
<td>222</td>
<td>Cyanogenic beta-glucosidase dhurrinase 2 genes</td>
<td><em>S. bicolor</em></td>
<td>8e-32</td>
</tr>
<tr>
<td><strong>Signal transduction continued</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1G21</td>
<td>108</td>
<td>Unknown</td>
<td><em>E. coli</em></td>
<td>2e-21</td>
</tr>
<tr>
<td>10G13</td>
<td>120</td>
<td>Unknown</td>
<td><em>H. sapiens</em></td>
<td>1e-18</td>
</tr>
<tr>
<td>11G1</td>
<td>161</td>
<td>Unknown</td>
<td><em>P. fluorescens</em></td>
<td>5e-13</td>
</tr>
<tr>
<td>1G4</td>
<td>107</td>
<td>Putative hypothetical protein</td>
<td><em>Z. mays</em></td>
<td>0.016</td>
</tr>
<tr>
<td>12G31</td>
<td>123</td>
<td>Unknown</td>
<td><em>C. roseus</em></td>
<td>0.33</td>
</tr>
<tr>
<td>1G42</td>
<td>282</td>
<td>Putative protein</td>
<td><em>Z. mays</em></td>
<td>3e-39</td>
</tr>
<tr>
<td>3G31</td>
<td>170</td>
<td>Unknown</td>
<td><em>Penaeus monodon</em></td>
<td>4e-08</td>
</tr>
<tr>
<td>1G32</td>
<td>144</td>
<td>Hypothetical protein</td>
<td><em>Z. mays</em></td>
<td>7e-17</td>
</tr>
<tr>
<td>2G33</td>
<td>116</td>
<td>Hypothetical protein</td>
<td><em>T. brucei</em></td>
<td>0.088</td>
</tr>
</tbody>
</table>
### Table 3 continued

<table>
<thead>
<tr>
<th>DNA-AFLP fragment ID(^a)</th>
<th>Size (bp)</th>
<th>GenBank Homolog(^b)</th>
<th>Species</th>
<th>Blast E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1G33</td>
<td>285</td>
<td>Unknown clone</td>
<td><em>H. sapiens</em></td>
<td>3e-07</td>
</tr>
<tr>
<td>D2G32</td>
<td>129</td>
<td>Unknown/ cDNA-AFLP sex and growth traits</td>
<td><em>P. monodon</em></td>
<td>5e-06</td>
</tr>
<tr>
<td>810G6</td>
<td>122</td>
<td>Hypothetical protein</td>
<td><em>Z. mays</em></td>
<td>9e-15</td>
</tr>
<tr>
<td>11G11</td>
<td>127</td>
<td>Hypothetical protein</td>
<td><em>O. sativa</em></td>
<td>5e-12</td>
</tr>
</tbody>
</table>

The first letter of the fragment name corresponds to the extracted cDNA-AFLP fragment type, e.g. D1G33, U1G11 and D1G33 represent novel, up-regulated and down-regulated respectively. BlastX was used to determine homologous genes and putative functions of genes. BlastN was used in case of failure to return any hits by BlastX.

Of all the 53 induced characterized genes, 6% are due to abiotic stress, 4% are responsible for cell maintenance, 8% are involved in cell wall fortification, 13% are defense-related, 11% related to development, oxidization and reduction related genes accounted for 4%, 9% are involved in photosynthesis, 21% are responsible for signal transduction and 24% are of unknown function (Figure 8).
Figure 8: Functional categories of the sorghum genes responsive to sorghum midge floret feeding. In the pie chart, values of percentage indicate the proportion of a number of genes in each category to a total number of genes (53 genes). The functional categories included novel, up-regulated and down-regulated genes.

4.3 Validation of expression patterns using RT-PCR and qRT-PCR analyses

The RT-PCR analyses was used to verify the transcripts which were suspected to be responsible for sorghum midge defense in the sorghum. Figure 9 shows RT-PCR validation of maize protease inhibitor, the subtilisin/chymotrypsin inhibitor CI-1B and cinnamoyl-CoA reductase 2 (Table 3).
A. Maize protease inhibitor

B. Chymotrypsin inhibitor CI-1B

C. Cinnamoyl-CoA reductase 2

Figure 9: RT-PCR analysis of three selected differentially expressed cDNA-AFLP transcripts at three different time points after infesting with sorghum midge. (Abbreviation: L=ladder)

According to the cDNA-AFLP data, induction of the sorghum midge defense genes in the sorghum midge resistance genotype AF28 occurred. The same pattern of gene expression was obtained by RT-PCR: the 100b expected bands were amplified only in AF28 and were absent in the sorghum midge susceptible genotype Seredo. This was in agreement with the same pattern as found by cDNA-AFLP. The results therefore suggest that these genes are only expressed in the
sorghum upon sorghum midge attack. Overall, the RT-PCR data confirmed the robustness of the DNA-AFLP method and enabled further characterization of gene expression.

The qRT-PCR analyses carried out for 3 cDNA-AFLP transcript fragments (maize protease inhibitor, subtilisin/chymotrypsin inhibitor CI-1B, and cinnamoyl-CoA reductase 2) were selected based on their interesting cDNA-AFLP expression patterns in the time-course of the DNA-AFLP experiment and homology to genes known to play a role in plant insect herbivore defense. Expression profiles of the 3 cDNA-AFLP transcript fragments in sorghum florets after 6, 12, and 16 hour post-infestation with sorghum midge are shown in Figure 9. For each cDNA-AFLP transcript fragment, the same expression pattern was found with qRT-PCR analyses as observed in the cDNA-AFLP experiments demonstrating the reliability of the cDNA-AFLP data. The results from the controls at all time points support the conclusion that these genes changed their expression levels in response to the pest infestation.
Figure 10: Relative expression of the three target gene transcripts; a. maize protease inhibitor, b. chymotrypsin inhibitor CI-1B and c. cinnamoyl-CoA reductase 2, using qRT-PCR in sorghum midge infested sorghum florets. Expression was normalized according to α-tubulin.
CHAPTER FIVE
DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

1 Discussion

Understanding the complex transcriptional changes occurring after insect attack in plants is becoming increasingly important for the global study of plant–insect interactions and thus for an efficient management of insect pests. While the signal transduction pathways that are activated by mechanical wounding or insect feeding have been studied quite intensively during the past years (Maffei et al., 2007), far less is known about the changes in gene expression in plants as a response to herbivore feeding insects.

This study demonstrated that cDNA-AFLP is a powerful technique to study genes involved in the sorghum-sorghum midge system. Because AFLP primers are universal, the cDNA-AFLP technique is applicable to any organisms or any host-pathogen systems for comprehensive transcript profiling. This technique is robust, very sensitive due to PCR amplification steps and highly reproducible. The technique is simple and affordable.

The band patterns were obtained with cDNA-AFLP analysis and all the cDNAs obtained by using BstYI-MseI enzyme combination resulted in a large volume of the transcripts. Of the 6,720 bands detected from 12 primer combinations, 118 were novel, 17 were up-regulated and 23 were own regulated genes. Homology analysis of the cDNA-AFLP fragments sequenced after
Sorghum midge larvae feeding showed that 53 genes could be characterized into known or tentative function categories. In all the cases hypothesized to be vital for enabling the plant to overcome the stresses inflicted by sorghum midge feeding, RT-PCR and qRT-PCR analyses confirmed the result of the cDNA-AFLP experiment consistently.

Seven genes involved in direct defense responses were differentially expressed upon midge infestation. Members encoding for defense proteins such as pathogenesis-related, protease inhibitor genes are known to be involved in defense molecule biosynthetic pathways. Genes coding for enhanced defense susceptibility (EDS1)-like protein and thaumatin-like protein were induced. EDS1-like protein has previously been identified in plants under bacterial, fungal and insect stresses (Pegadaraju et al., 2007; Park et al., 2006; Hu et al., 2005). Thaumatin, an ethylene related protein has only been recorded to be expressed in plants attacked by insects (Zhu-Salzman et al., 2004).

A gene encoding for maize protease inhibitor (subtilisin/chymotrypsin inhibitor) was induced from the earlier stage of sorghum midge feeding (3 hours) and maintained at high levels until 12 hours post-infestation. Proteinous inhibitors from plants can inhibit insect digestive proteases and therefore suppress insect growth, development, and survival (Murdock & Shade, 2002).
Protease inhibitors are the first identified wound-inducible proteins and are important defensive in plant against chewing insect herbivores. They are antidigestive and antinutritive proteins which inhibit elastases in larval midgut (Van Loon et al., 2006) and hence decrease insect herbivore’s performance while in the host plant. The induced polyphenol oxidase encoding gene, known a wound response protein, catalyses biosynthesis of active quinones which are toxic to insect parasitic pathogens and herbivores due to their ability to produce indigestible amino acids and proteins (Haruta et al., 2001).

However, several genes previously induced in Poaceae upon insect infestation were not identified in this study. Such genes include chitinases and glucosidases (Zhang et al., 2004; Zhu-Salzman et al., 2004). The probable reasons behind this are that the primer combinations used in this study did not amplify these genes and that the cDNA in some bands were not successfully extracted.

Plant defense against invading pathogens and pests involves a crosstalk of signaling pathways which are highly controlled. As a whole, 11 genes involved in signal transduction were expressed in response to sorghum midge feeding (Table 3). The number in this category makes up for the second largest category, after that of unknown function, reflecting the significance of signaling pathways in plant defense. Among these genes, cinnamoyl-CoA reductase 2, a key enzyme in lignin biosynthesis and Ras-GTPase activating protein binding protein 2 are effectors of small GTPase Rac. The GTPase Rac is known to play a crucial role in controlling mitogen-
Activated protein kinases and transduces diverse signals in plants (Yang & Fu, 2007; Kawasaki et al., 2006). Cinnamoyl-CoA reductase is a key enzyme in lignin biosynthesis (Peters & Constabel, 2002). Lignin, a major component of secondary cell walls, is a heterogeneous multidimensional phenolic polymer resulting from the oxidative polymerization of monolignols. Deposition of lignin during defense responses is considered to function as a physical barrier against tissues attack (Dixon et al., 2002). A gene encoding for ankyrin- induced protein was up-regulated. Ankyrin regulates the salicylic acid defense reactions, including systemic acquired resistance (Durrant & Dong, 2004).

The central role of octadecanoid pathway in plants responding to chewing insects was demonstrated by Rayan (2000). This lipid-based signaling cascade involving jasmonate production via octadecanoid pathway and leading to direct and indirect defenses against chewing insect herbivores is activated (Schilmiller & Howe, 2005). Of all the nine genes involved in signal transduction, none of those involved in octadecanoid pathway were expressed.

Cell wall reinforcement is another major defense strategy employed by plants (Collins et al., 2003). In transcript profiles obtained, 4 genes which are related to cell wall fortification were included. They are; S-adenosyl-L-methionine, phosphoethanolamine N-methyltransferase, caffeic acid 3-O-methyltransferase and d-TDP-glucose dehydratase, which have all been recorded as insect feeding cell wall inducible genes (Park et al., 2006; Cruz-Ramirez et al., 2004). The four genes play an important role in the biosynthesis of lipids which are an integral
Component of the membrane phospholipid phosphatidylcholine biosynthesis. Caffeic acid o-ethyltransferase participates in lignification of cell walls (Do et al., 2007). This supports the proposal that cell wall fortification played a fundamental task in resistant genotype AF28 against sorghum midge feeding.

Sorghum midge induced a dhurrinase transcripts in 3 hours after infestation with sorghum midge vae. A putative flavonol glucosyltransferase and cyanogenic beta-glucosidase dhurrinase 2 enes are involved in the synthesis of phenylpropanoid/ secondary compounds, which are ecursors for dhurrin synthesis (Sue et al., 2006). The three gene pathway for synthesis of efensive compound dhurrin has been transferred from sorghum into Arabidopsis, resulting in ect resistant plants (Tattersall et al., 2001). Sorghum is known to synthesize numerous unique anogenic glycoside, flavonoids and other compounds as part of its broad response to insect tack (Lo et al., 1999; Hipskind et al., 1990; Nicholson et al., 1987), reflective of its gnificance genetic diversity (Menz et al., 2004). This demonstrates the potential of sorghum as resource for gene discovery and supplementation for the current understanding of plant efense.

Plants respond to many forms of biotic stress by generating reactive oxygen species (ROS) that articipate in defensive signaling and participate in hypersensitive response. ROS are reactive molecules that accumulate to toxic levels with water deficit and other stresses. Enzymes that etabolite ROS are therefore important in preventing the damage that excess ROS could cause.
Several transcripts for proteins that constitute intracellular ROS were induced by sorghum midge feeding. Genes encoding peroxidase 2 and quinone oxidoreductase were induced (Table 3). Both peroxidases and reductases play a role in controlling ROS leading to oxidative signal transduction (Apel & Hirt, 2004). Peroxidases exist as isoenzymes with diverse expression profiles participating in various physiological processes, such as lignification, suberization, auxin catabolism, wound healing and defense mechanisms against pathogen and insect attack (Cosio & Dunand, 2009). Quinone oxidoreductase, which was up-regulated, is known for hunting for toxic free radicals of semiquinones using divergent reduction in Arabidopsis (Mano et al., 2002).

Sorghum midge infestation induced a water responsive protein, aquaporin, and temperature inducible protein, HEAT shock protein (Table 3). Aquaporins are proteins embedded in the cell membrane that regulate the flow of water. They facilitate osmosis by forming water-specific pores as an alternative to water diffusion through the lipid bilayer, thus increasing water permeability. It has been shown that suppression of aquaporin reduces draught tolerance (Bartels & Sunkar, 2005). However, its induction can be attributed to the osmotic changes in the affected tissue, caused by the sorghum midge larvae (Thompson & Goggin, 2006). Heat shock proteins are molecular chaperones for protein molecules and play an important role in protein-protein interactions such as folding, assisting in the establishment of proper protein shape and prevention of unwanted protein aggregation. These proteins are induced by abiotic stresses like drought, cold and high-salinity. It has previously been shown to be induced by mechanical, insect or pathogen attack (Hamilton III & Coleman, 2001).
The results also support substantial effect on housekeeping genes and intermediary metabolism probably in order to redistribute resources from growth into defense. The housekeeping genes were categorised as cell maintenance involved genes, photosynthesis-related genes and genes involved in development. Resource distribution is fundamental for plant survival during insect attack (Bezemer et al., 2003) and identification of a number of cDNA-AFLP fragments that change during sorghum midge larvae herbivory open up possibilities for testing their relative role in defense and resistance.

Among numerous cell maintenance genes, only two genes, were induced on sorghum by sorghum midge feeding; 50S ribosomal protein L33 and vesicle transport v-SNARE 13 (Table 3). The 50S ribosomal protein L33 is a ribosomal protein which is involved in protein synthesis. Of interesting significance in this category is the vesicle transport v-SNARE 13 which is responsible for targeting vesicles involved in the secretory pathway. The hypothesis connecting transport v-SNARE 13 to insect herbivory defense is that defense related secondary metabolites may be packaged in vesicles and their secretion and targeting may be aided by v-SNARE 13 proteins.

The relationship between photosynthesis and plant defense was established by Creelman & Mullet, (1997), where an inverse correlation between photosynthesis and defense-related gene regulation was observed in plants subjected to fungal infection, elicitor treatment or insect herbivory. This response seemingly allows energy reallocation to defense responses, with inhibition of less important functions, upon attack by insects or pathogens. Among the genes
induced are Phosphoenolpyruvate carboxylase and Chlorophyllide a oxygenase (Table 3). Chlorophyllide a oxygenase catalyses a two-step oxygenation reaction involved in the synthesis of chlorophyll b and therefore this gene was presumed to have been suppressed.

Six genes encoding for proteins responsible for plant development were induced by sorghum midge larvae feeding (Table 3). Among them, the Scarecrow-like1 is a member of the GRAS gene family which encode transcriptional regulators that have diverse functions in plant growth and development such as gibberellin signal transduction, root radial patterning, axillary meristem formation, phytochrome A signal transduction, and gametogenesis (Ricci et al., 2008). Auxin induced protein was down regulated upon sorghum midge feeding on sorghum florets. The plant hormones such as auxin and gibberellin have been widely involved in plant development. When plants are negatively stressed, they develop a strong response with the corresponding activation of a plethora of defense-related genes to thwart the effects of the stressor. The expression of the development hormones in stressed plants have been shown to have antagonistic relationships with defense related proteins such as abscisic acid and ethylene (Mayda et al., 2000).

The genes with unknown function ranked the highest in all the 9 categories (Table 3). A total of 13 cDNAs failed to hit any matched sequence from GenBank databases by both BlastN and BlastX or matched to sequences whose functions have not been characterized yet. Some of them showed up-or down regulation by sorghum midge feeding. This implies that these genes could be involved in regulation of sorghum defense responses against sorghum midge.
This is the first report of expression of a large number of gene fragments related to the defense response in sorghum upon sorghum midge infestation. A significant outcome of this study is the identification of novel transcriptionally-derived fragments that will provide new insights and better understanding of various components involved in sorghum midge defense for improvement of sorghum cultivation in the near future.

5.2 Conclusions

i. The cDNA-AFLP technique was successfully used to determine gene expression patterns in an interaction between sorghum and the sorghum midge larvae. Genes involved in compatible interactions between sorghum and the sorghum midge pest were identified and their expression patterns were determined. The genes and their putative functions provide insight in understanding the sorghum-sorghum midge interaction, and provide candidate genes for future function analysis.

ii. Quantification of the level of gene expression by qRT-PCR analysis provided the necessary molecular tools for a better understanding of sorghum defense reactions against sorghum midge, and therefore with the knowledge to design improved strategies for breeding or genetically engineering sorghum resistant apple cultivars in the future.

iii. An integration of the genomic information obtained from this study with functional proteomics and metabolomics will enable researchers understand the mechanism of sorghum midge resistance. Such information can be used by breeders for selection of candidate genes, and their transfer to agronomically important sorghum varieties.
1.3 Recommendations

i. The results of this study are a valuable resource to help further the understanding of the mechanisms of sorghum midge defense in sorghum, and of how these defense mechanisms are regulated at a transcriptional level.

ii. Further studies on the over expression and RNA silencing of the genes hypothesized to be key in sorghum midge defense to determine the functions of transcripts and characterization of the transcripts identified in this study. In addition, identifying the functions of the numerous ‘unknown’ fragments will provide further exciting investigations.

iii. Other transcript profiling techniques such as microarrays should be used to profile genes expressed in AF28 sorghum genotype to validate the results obtained in this study. This will also solve the problem of cDNA-AFLP fragments which were not recovered.

iv. Further studies need to be carried out to construct transcript expression maps of other sorghum genotypes known to be resistant to sorghum midge attack using more marker types locate these markers on chromosomes and convert AFLP markers to sequence characterized amplified region markers for genome mapping and marker-assistant selection.

v. Other work needs to be carried out on the signalling pathways that activate sorghum defense responses. This would provide a basis for understanding the defense mechanisms
of *Poaceae* in florets and growing embryos. This is because much has been done on signal transduction on various plant parts with an exception of flowers and their developing embryos. As sorghum midge feeds on developing embryos, future research should therefore focus on unveiling the signal transduction pathways on these organs, following perception of the pest by the host and the complex ways in which they interact.
REFERENCES


Phytoalexin Deficient4 without its signaling partner Enhanced Disease Susceptibility 1. 
Plant Journal 52:332-341.


APPENDICES

Appendix 1: Sequences of differentially expressed fragments generated by cDNA-AFLP between AF28 and Seredo upon sorghum midge infestation

cDNA-AFLP fragment ID | Sequence
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D3G12</td>
<td>ATGACTGCGTAGTGATCTTCAGTACACCTGTAAGCTCAGTTTCGACGATTATTT ATTGAATTTTCTTTACTCAGGACTCATCATGCTCTTTGCTGATGCTGAGAACGAGTCTAC GCATTTTACTCAGGACTCATC</td>
</tr>
<tr>
<td>N3G22</td>
<td>GATGAGTAGCTGTTGAGTCAAATCCAAAGAAAAATCTGAATCAAATTTGACTTTGAGTCAATG GATCAGTGACATGACTGAGATTTATTTTCTTTACTCAGGACTCATCATACTGCGTAGTGATCTTGTATCTACAATGCAGAACTGT</td>
</tr>
<tr>
<td>N12G13</td>
<td>GACTGCGTAGTGATCTTCCATTATTTCATACAAATGAGTGCAAGATCAGTTTCGATGATC AACTACGAGATCATCAGGACTGAGATTTCTTTACTCAGGACTCATCATACTGCGTAGTGATCTTGTATCTACAATGCAGAACTGT</td>
</tr>
<tr>
<td>N5G22</td>
<td>TTCATGTCTACACTCCCTCTCCTACATTTGAGGCTTTTGTCACTTCATTTACCTCAAGGACTCATCAGTTCGACGTTTCGACGACTGACATTTTTTACAGGACTCATCATACTGCGTAGTGATCTTGTATCTACAATGCAGAACTGT</td>
</tr>
<tr>
<td>N8G11</td>
<td>GAGATCCAATCCGGAGCTCGAGTAGTGTTGAGCTCAGTTTCAGCAAGTGCAAGATCAGTTCGACGTTTCGACGACTGACATTTTTTACAGGACTCATCATACTGCGTAGTGATCTTGTATCTACAATGCAGAACTGT</td>
</tr>
</tbody>
</table>
GGATTTTACCTGAAAGCAGCGTCTCCGCCTCCGTGATGCTACATCACCACACTGA
ATGTCTGCAAGCCTACAGGATGAAAGCCGATCCGTGACCAGATTACCATGTGACA
CTCGGGGCCCCACTGTGCCAAGGAAATCATTGAGACTGGAACAGCCTGCTGAGAGC
TCGTCAACGCTGAAACCCGACATGAGTTATGCTCCGGGGCTGAGGACACCCCTAC
CTGACCATTGAAAGGGCATTGCTGCTGGGATGCAAGAACCAGTCTGTTAGCACAAGATGTG
ACCATACCGGAGTGGGAGTCACTACACTACCTACGACATGAGTTACATCTGAGTAA

\text{N12G11} \quad \text{GATGAGTCCTGAGTAAAGCAGCCCTACAGGATGAAAGCCGATCCGTGACCAGATTACCATGTGACA}
CTCGGGGCCCCACTGTGCCAAGGAAATCATTGAGACTGGAACAGCCTGCTGAGAGC
TCGTCAACGCTGAAACCCGACATGAGTTATGCTCCGGGGCTGAGGACACCCCTAC
CTGACCATTGAAAGGGCATTGCTGCTGGGATGCAAGAACCAGTCTGTTAGCACAAGATGTG
ACCATACCGGAGTGGGAGTCACTACACTACCTACGACATGAGTTACATCTGAGTAA

\text{N5G12} \quad \text{TGCCCTACCACCCAGGGCGACCACCGAGACCCAGGCGACAGTAGAGAGCGT}
CCGCACCGCCCAATGACTACACTTCTGGGAGGCCACACCGGACGACCATCTACGTC

\text{N8G32} \quad \text{GATGAGTCCTGAGTAAATCTCTGGGCCACAATTTGTTCAAAGCGACCAGCAATCAAAT}
AAACGTCCCTACCAGCGGCCCAATTTGCGCCTGAGAACTTAAAGGAGATACGTGAC
CTTACAACTTTTTAATATTTTTTACGAGGACATGACCTGATGACACCTAGACCTTCCA
CGGACTCAATAACTCTCCCTTCTTCTAGAGAAATTTGTATGAGATTACGTC

\text{N6G11} \quad \text{GATGAGTCCTGAGTAAATCTCTGGGCCACAATTTGTTCAAAGCGACCAGCAATCAAAT}
AAACGTCCCTACCAGCGGCCCAATTTGCGCCTGAGAACTTAAAGGAGATACGTGAC
CTTACAACTTTTTAATATTTTTTACGAGGACATGACCTGATGACACCTAGACCTTCCA
CGGACTCAATAACTCTCCCTTCTTCTAGAGAAATTTGTATGAGATTACGTC

\text{N10G5} \quad \text{TGATACACAAGATGATGAGTCCTGAGTAAAGAGCTCGCTCTCTTCTATAATCAAG}
GTGCTCCTCTCTCCAGATCTCTGGCTGTAATACTACGAAAATAATGGGATCAGATCAA
CTACCTAC

\text{N3G2} \quad \text{TGGGGCGAGCACTCAGAGATAGAAGATTGGGAGGTAAAGGATAGAGCGGCTAAGAGC}
ACTCAAAGAAGGGCGGCTCACATCTCATGATGCTTTGTTGGAATTCTTGAGAAGTGAAAGG
TATCGAATGCCTACATACTTCCCTACTAAGGGAGGCTCTCATTTAGGACAGC
ACGTTCATAGCTCAAAATACAGAGCCTCAAAAGACATTAGGCAACTCTGTGATAG
CATAAGTCAACCCACATCTGCTTACATTG

\text{N13G2} \quad \text{CTGGTGAGAACAATCTCGAGATGGTCTAGTGCGCTACCCAGGGCTCCTCGGCACTCGCTA}
CTCGATCGTGAAGATTGGGTGCAAGCAGCGACCTGTAAGAGCGGTCAAGGAGG
AAAGAGTTACACCGAGCTGCAAAAGGCGGCTGAAATCTGATTAAAGGCAAAATTTGGA
GCGAAATAAAATAGATCAAAGAGTTGTTACGCAGCTTATAACACTGGTGATGTTTA
N12G4  GATGAGTCTCTGAGTAAAGCAAAAGTATATATTCGAAACCGAAATAAAAAACCAACTT
CAAAAAAGACTTCCCCGATACGACTCCGATACCATGACTTAGGAAAGTACAG
ATTGCCCATATAAAAGTTTCTTTCTTTCCCTCTACACATTTACCTGAGTTTT
CAGATGCCTTTCCGAACCTCTAGACAGATCACTACGCAGTGATGAGTCCTGAGTAA

N7G22  CTGCGTAGTGATCTTTGTGCCTGCGTAGTGCTTCTACGAGAC
AGTGAAGAGCTCCAGGAGAAAGGCACTGCTGGGCGCTCCCCCATTGGGAAAGAGCAC
AGAACCAACAATCCTACAACTACATATATATTTTATTTTATATTATATATGAGCATTTTACGAGACTTAC
ATCATAGCTGGCTGTAGCTTCTTGACTGGCTGATGCTTCTTGACTTGAGTACTTCTTA
CTCAGGACTCATC

N1G12  AGAGCCAAATCTTGTATTGGGCAACGTCAGCAGAAATATGGAAGTCTGACT
TGAAAGTATATTCGTACAAAGTTTCTTTCTTGCTGCTGTTCTGACGGGGATCATC
TCAAGAGCCAGAGGACTGCTCTTCTCAGGACTCATC

U4G3   ATGACACCTCCTGAAGAGCATTTGACCTACCCCTCGAGTGGAGCTCTAGTTGCCTCTGCCTGCACAT
CTCCGAAGCCACTTAAACGCAGCCCTGACCCAGAAGCCAGATGACTCCAGACT
GGTACAGGCATATGTTTCTCACTAAACTAGGAGCAAGATCACAAGGCACATC
GGTTAAGAGCTGATCATGACTGACTACCTCCTCTGTCCCTGGGCTGAGGACTGACTTCA
AGAATGCTGCCCTACCAATGCTATACATCGCGAGCCTGACTATGGCAGCATGATG
GGGCCAAAGCCACGCCCAGTCATTCTAGATG

N1G21  ATCGACCAAGCCTTCGCTGATAGTCAGTGGGATGATGTTTTTCTGCAGGAGGACTACACG
ACTTATGCTTAAAATTTCATCGTGACCGACTTTTCTCGATATGCTGTAATTTCTCT

N11G13  GAAGATGACTGCTTACTGAGTATTTGACTGATTAGTATACTGCTGATAC
CGTGGTATACCCGAAACCGCCAAATCTCCTTTGAGTACATGGCACCCATGCTCCCG
CTTACACCTCATTCTTGCTAGAGGACTCAGACTGCTTCAGAGACTCATC

N8G4   GATGAGTCTCTGAGTAAAGTTTACTTACTACAAAGACTTTGCTGCGGCGGACTATAGTTAT
GATACACACATCTTAAGTTACTGCAAAACCAGCAATAAGATACACTACGAGTC

N12G31  CTGCGTAGTGATCTTTCTTTCTTTACTCAGGACTCATCATTAGCAGAAGATGATTACT
CAGGACTCATCATAACCGTAGTATGCTGAGTCTAGTTAGAAAATCATC
GGATTACTCAG

U3G31  GACTGCGTACTCTGTCAGAAGATGACTTCAGTATTAGCAGAAGATGATTACT
CAGGACTCATCATAACCGTAGTATGCTGAGTCTAGTTAGAAAATCATC
GGATTACTCAG
Appendix 2: qRT-PCR quantification of transcripts for the selected cDNA-AFLP fragments and α-tubulin in infested and non-infested floret tissues of AF28 sorghum at various time-points after infestation with sorghum midge larvae

1. Maize protease inhibitor (MPI)

<table>
<thead>
<tr>
<th>Time(h)</th>
<th>α-tubulin</th>
<th>CT1</th>
<th>Genes</th>
<th>CT2</th>
<th>△△CT</th>
<th>2^{△△CT}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>α-tubulin-C-0</td>
<td>18.5484</td>
<td>MPI-C-0</td>
<td>31.286</td>
<td>0.000</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>α-tubulin-3</td>
<td>18.654</td>
<td>Infested-3</td>
<td>31.041</td>
<td>-0.351</td>
<td>1.275</td>
</tr>
<tr>
<td>6</td>
<td>α-tubulin-6</td>
<td>22.067</td>
<td>Infested-6</td>
<td>34.506</td>
<td>-0.298</td>
<td>1.229</td>
</tr>
<tr>
<td>12</td>
<td>α-tubulin-12</td>
<td>17.423</td>
<td>Infested-12</td>
<td>30.073</td>
<td>-0.087</td>
<td>1.062</td>
</tr>
</tbody>
</table>

2. Chymotrypsin inhibitor CI-1B (Chy)

<table>
<thead>
<tr>
<th>Time(h)</th>
<th>α-tubulin</th>
<th>CT1</th>
<th>Genes</th>
<th>CT2</th>
<th>△△CT</th>
<th>2^{△△CT}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>α-tubulin-C-0</td>
<td>19.857</td>
<td>Chy-C-0</td>
<td>33.763</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>α-tubulin-3</td>
<td>19.260</td>
<td>Chy - Infested-3</td>
<td>32.872</td>
<td>-0.294</td>
<td>1.226</td>
</tr>
<tr>
<td>6</td>
<td>α-tubulin-6</td>
<td>19.286</td>
<td>Chy - infested-6</td>
<td>32.730</td>
<td>-0.463</td>
<td>1.378</td>
</tr>
<tr>
<td>12</td>
<td>α-tubulin-12</td>
<td>19.899</td>
<td>Chy - Infested-12</td>
<td>33.537</td>
<td>-0.268</td>
<td>1.204</td>
</tr>
</tbody>
</table>
### 3. Cinnamoyl-CoA reductase 2 (CCoA)

<table>
<thead>
<tr>
<th>Time(h)</th>
<th>α-tubulin</th>
<th>CT1</th>
<th>Genes</th>
<th>CT2</th>
<th>ΔΔCT</th>
<th>$2^{\Delta\Delta CT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>α-tubulin-C-0</td>
<td>21.990</td>
<td>CCoA-C-0</td>
<td>30.382</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>α-tubulin-3</td>
<td>21.127</td>
<td>CCoA - Infested-3</td>
<td>29.228</td>
<td>-0.2906</td>
<td>1.222</td>
</tr>
<tr>
<td>6</td>
<td>α-tubulin-6</td>
<td>22.082</td>
<td>CCoA - Infested-6</td>
<td>30.032</td>
<td>-0.441</td>
<td>1.3578</td>
</tr>
<tr>
<td>12</td>
<td>α-tubulin-12</td>
<td>21.198</td>
<td>CCoA - Infested-12</td>
<td>29.409</td>
<td>-0.180</td>
<td>1.1328</td>
</tr>
</tbody>
</table>

**Formula:** Fold Change = $2^{-\Delta\Delta CT}$ ($\Delta CT$ gene=gene CT infested - gene CT control; $\Delta CT$ α-tubulin=CT α-tubulin infested - CT 18S control; $\Delta\Delta CT=\Delta CT$ gene - $\Delta CT$ α-tubulin) CT1 and CT2 are the mean values of threshold cycle for α-tubulin and each gene in three biological replicates. Figures represent relative expression changes for each gene in control and infested floret tissues across time points.