CLONING AND CHARACTERIZATION OF MICRORNAS FROM SORGHUM

By

LI ZHANG

Master of Science in Microbiology

Sichuan University

Chengdu, Sichuan, China

2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2010

CLONING AND CHARACTERIZATION OF

MICRORNAS FROM SORGHUM

Thesis Approved:

Ramanjulu Sunkar

Thesis Adviser

Ramamurthy Mahalingam

Yinghua Huang

Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

My deepest gratitude goes first and foremost to Dr. Ramanjulu Sunkar, my advisor, for his constant encouragement and guidance. Without his consistent and illuminating instruction, the research work could not be completed successfully and this thesis could not have reached its present form.

Secondly, I should give my hearty thanks to Dr. Ramamurthy Mahalingam and Dr. Yinghua Huang for their patient instructions in various courses and their precious suggestions for my research.

I also owe my sincere gratitude to the labmates-Drs. Guru Jagadeeswaran, Yongfang Lee, Guosheng Li and Jessica Matts, who gave me their help and time in listening to me and helping me work out my problems. I also want to thank the core facility of BMB department for their help in my research work

Lastly, my thanks would go to my beloved family for their loving considerations and great confidence in me all through these years.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

I. microRNAs: Inroduction

Highly coordinated multiple gene regulatory mechanisms involving transcriptional, post-transcriptional and post-translational regulations in a spatio- and temporal-specific manner determine the optimal plant growth and development, plant progression into different phases as well as other physiological processes including stress responses (Mallory and Vaucheret, 2006; Jones-Rhoades et al., 2006; Sunkar, 2010). Of the different gene regulatory mechanisms, transcriptional regulation, which is dependent on the action of specific transcription factors that bind to specific *cis*-elements in the promoter region, is relatively the best understood phenomena. Although post-transcriptional gene regulation was thought to be one of the critical mechanisms of gene regulation but the components that mediate this process was relatively unknown and only recently small RNAs, which act as guide molecules in this process has been uncovered in plants and animals.

In plants, endogenous small non-coding RNAs vary from 21 to 30 nucleotides in size, which can be divided into two major classes such as microRNAs (miRNAs) and small interfering RNAs (siRNAs). MicroRNAs (miRNAs) are a class of evolutionarily conserved small RNAs and regulates the target gene expression at post-transcriptional level by guiding cleavage and/or attenuating the translation of target mRNA (Carrington and Ambros, 2003; Bartel, 2004; Molnár et al., 2007). In rice, 24-nt long miRNAs (lmiRNAs) have been recently discovered (Wu et al., 2010). Interestingly, their biogenesis dependent on dcl3a but not on dcl1 and are capable of directing DNA methylation at the target locus, implying that lmiRNAs play a role in transcriptional gene silencing (Wu et al., 2010).

Future production of renewable transportation fuels will require a consistent supply of biomass produced specifically for biofuel production. Sweet sorghum can be grown throughout temperate climate zones of the United States, including Oklahoma. It is poised to become one of the major sources of biofuel plants in the United States, because of its high biomass yield with low inputs and drought tolerance coupled with its genetic diversity. Sorghum (*Sorghum bicolor* L. Moench), an African grass related to sugarcane and maize, is grown for food, feed, fiber and fuel (Paterson et al., 2009). As a C4 plant species, sorghum has high photosynthetic efficiency for converting solar energy to biomass; high water use efficiency for growing in high temperature and drought prone areas; and can be grown on poor and marginal lands (Sasaki and Antonio, 2009; Wang et al., 2009). Because, sorghum can thrive in hot, semidry places, it feeds more than 500 million people in 98 countries especially in arid and semi-arid regions (Pennisi, 2009, Xin et al., 2009). Sorghum is not used widely as food grain in the US, but recently attracted much attention as a versatile feedstock for large-scale bioenergy production given its sugar from stem juice, cellulose/hemicellulose from stalks, and starch from grain. Realizing the importance of the sorghum, recently its genome has been sequenced (Paterson et al., 2009). Nevertheless, we know little about the basic biology of sorghum. Better knowledge of gene-regulatory processes controlling biomass accumulation, plant architecture, cell wall composition, nutrient uptake and assimilation and stress responses could assist in designing rational strategies for improving biomass production and other traits important for biofuel production and processing.

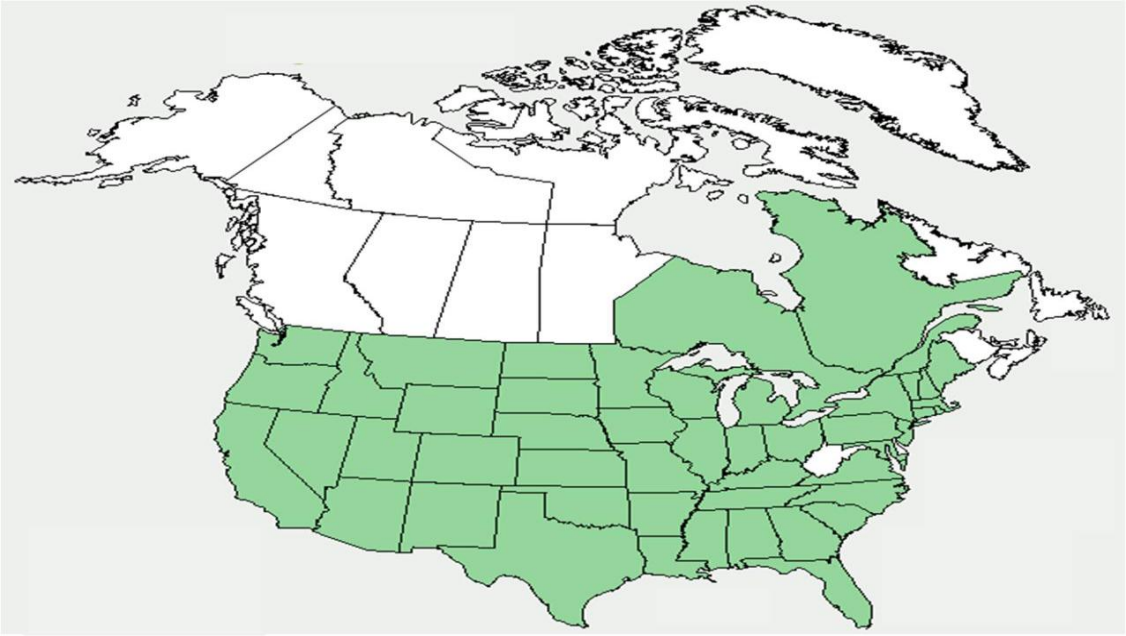


Figure 1. Distribution of Sorghum bicolor (L.) Moench in the US and Canada. The presence and absence are representative by green and white, respectively. (Modified from: http://plants.usda.gov/java/profile?symbol=SOBI2).

II. Project objectives

Our current knowledge about the regulatory roles of miRNAs and their targets point to the fundamental functions in various aspects of plant development, including auxin signaling, meristem boundary formation and organ separation, leaf development and polarity, seedling development, embryo development, phyllotaxy, lateral root formation, transition from juvenile-to-adult vegetative phase and from vegetative-to-flowering phase, floral organ identity, petal number and reproduction (Mallory and Vaucheret, 2006; Jones-Rhoades et al., 2006). In addition to their roles in development, miRNAs play important roles in adaptation to biotic and abiotic stresses including phosphate, sulfate and copper-deprived conditions (Sunkar et al., 2007; 2010). Furthermore, recent studies in Arabidopsis have shown that overexpression of miR156 causes a moderate delay in flowering, initiate leaves faster and also cause a severe decrease of apical dominance (Schwab et al., 2005). Combination of these traits leads to a ten-fold increase in total

leaf number in transgenic plants compared to wild-type plants. Similarly, better understanding of miRNA-guided gene regulations can lead to improving abiotic stress tolerance in plants (Sunkar et al., 2006). The transgenic Arabidopsis plants overexpressing a miR398-resistant Cu-Zn superoxide dismutase (*CSD2*) exhibited improved tolerance to diverse stress conditions (Sunkar et al., 2006). Thus, finding miRNAs in biofuel plants have implications both for improving biomass accumulation and stress tolerance of the plant. Currently, with the development of powerful tools for genetic manipulation and the completion of sorghum genome sequencing, genomics-based approaches hold great promise for molecular breeding of sorghum with novel or improved quality traits. Thus far, the miRNA component of sorghum is unknown. Here, I focused on identifying miRNAs from sorghum by sequencing a small RNA library. Additionally, the genes that the miRNAs are targeting in Sorghum were predicted and a few such target genes were validated.

CHAPTER II

REVIEW OF LITERATURE

I. Endogenous small RNAs, biogenesis and function in plants

High throughput sequencing of small RNA libraries has revealed an unexpected diversity and greater abundance of endogenous siRNAs in plants (Lu et al. 2005; Sunkar et al. 2005b; Rajagopalan et al. 2006; Johnson et al. 2007). Like miRNAs, endogenous siRNAs are 21-24-nt small RNAs thus are structurally related to miRNAs but differs in their biogenesis. Endogeneous siRNAs are derived from the processing of typically long dsRNAs whereas miRNAs are processed from a single stranded RNA that can adopt hairpin-like structure (Plasterk 2002; Waterhouse et al. 2001; Doench et al. 2003; Tang et al. 2003). The synthesis of dsRNA is largely dependent on specific RNA-dependent RNA polymerase (RDR) activity. The dsRNA is processed by the DCL family of enzymes (DCL2, DCL3 and DCL4) to produce the predominant 21 and 24-nt siRNAs. These endogenous siRNAs can fall into different sub-classes such as transacting siRNAs (ta-siRNAs), natural antisense siRNAs (nat-siRNAs), heterochromatic siRNAs and long siRNAs (lsiRNAs) based on their biogenesis and function (Vaucheret, 2006). Transacting siRNAs are encoded by TAS loci but their biogenesis is dependent on miRNA-pathway. TasiRNAs can regulate the target gene expression similar to that of miRNAs, i.e., degradation of target mRNA (Allen et al., 2005). Natural cis-antisense transcripts-associated siRNAs (nat-siRNAs) are derived from the expression of convergent gene pairs (Borsani et al., 2005; Katiyar-Agarwal et al., 2006). Heterochromatic siRNAs, are specifically derived from the transcription of repetitive sequences (centromeric repeat sequences, retro-elements, transposons etc.,) and are involved in DNA and histone methylation in ARGONAUTE4 (AGO4) dependent pathway (Chen et al., 2004; Xie et al., 2004; Zilberman et al., 2004; Kasschau et al., 2007). Long-

siRNAs (l-siRNAs, 30 – 40 nt long), which are longer than conventional small RNAs and are generated from convergent pairs of genes, by a mechanism that is similar to the biogenesis of nat-siRNAs and were shown to be generated in response to bacterial infection (Katiyar-Agarwal, 2007).

II. MicroRNAs

The characteristics of miRNA (lin-4) was described in 1993 but were termed as 'short-temporal RNA (stRNA)' in C. elegeans. The lin-4 controls the timing of larval development by regulating the target gene lin-14 in C. elegans (Lee et al., 1993). It took almost additional seven years to discover a second miRNA gene, let-7, encoding another small non-coding regulatory RNA in the C. elegans, which is also involved in heterochronic pathway (Reinhart et al., 2000). Most importantly, it was found that the let-7 sequences are conserved in several higher eukaryotes including human (Reinhart et al., 2000). Subsequently, these were named as "microRNAs" and many miRNAs were discovered in Drosophila, C. elegans and mammalian cells by sequencing small RNA libraries (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Identification of conserved miRNAs from several animals has dramatically changed the perception that gene-regulation guided by such small RNAs is of fundamental importance in diverse biological processes. In 2002, first plant miRNAs were reported from Arabidopsis (Reinhart et al., 2002).

III. MicroRNA biogenesis in plants

In plants, the genes encoding miRNAs often exists as independent transcriptional units although rarely exists as clustered units (Jones-Rhoades et al., 2006). The miRNA genes are transcribed by the RNA polymerase II, resulting in synthesis of long primary miRNA transcripts (pri-miRNAs), which are polyadenylated at their 3'ends and capped at their 5'ends, like RNA polymerase II products. Pri-miRNA transcripts possess self-complementary nucleotides and thus can adapt a

hairpin-like structure. The enzyme, Dicer-like 1 (DCL1) acts on such hairpin-like structures and excises the miRNA and miRNA* duplex possessing 5'phosphate and 3'hydroxyl group with 2-nt overhangs. Recent studies indicated that several other proteins such as HYL1 (hyponastic leaves 1 is a dsRNA binding protein), SE (serrate, a C2H2 zinc-finger protein), DDL1 (Dawdle), assists the DCL-1 enzyme in liberating the miRNA and miRNA duplex from the hairpin-like structure. HYL1 interacts strongly with DCL1 in yeast two-hybrid assays.SERRATE (SE, a C2H2 Zinc finger protein) appears to be another important component of miRNA biogenesis in plants. The in vivo function of SE is unknown, but it interacts with HYL1 in the yeast two-hybrid system. Dawdle (DDL) is a nuclear-localized FHA domain-containing protein and is required for the accumulation of miRNAs in Arabidopsis. It's affinity for RNA, its potential association with DCL1, and the reduction in pri-miRNA levels in ddl loss-of-function mutants suggest that DDL could recruit DCL1 to its substrates (Yu et al., 2008). CBP20 and CBP80, the 2 mRNA cap-binding proteins (CBPs) have been shown to play important roles in assisting the DCL-1 in processing the miRNA:miRNA* duplex from the hairpin-like structure. Then, the last nucleotide of miRNA duplex are methylated by the action of HEN1, a nuclear-localized methyltransferase (Yu et al., 2005). HEN1 methylation prevents plant miRNAs from the 3'-end uridylation (addition of the oligo U to the 3' end of the miRNA strands) which interferes with the miRNA ability to enter the RISC complex and the subsequent degradation (Li et al., 2005). The methylated miRNA:miRNA* duplex is then exported to the cytoplasm by HASTY5 (a plant ortholog of exportin 5) (Park et al., 2005). In summary, DCL1, CBPs, HYL1, DDL-1 and SE form a complex for miRNA biogenesis. The miRNA/miRNA* duplex is stabilized by HEN1 and exported by HASTY5 (Figure. 2) (Liu et al., 2005; Yang et al., 2006; Dong et al., 2008; Kim et al., 2008; Laubinger et al., 2008; Ramachandran and Chen, 2008). In the cytosol, miRNA (miRNA: guide strand, miRNA*: passenger strand) is loaded into RISC containing AGO (Argonaute). Rules that govern strand incorporation into RISC are based on low pairing energy at the 5' end of the incorporated (guide) strand, compared to the discarded (passenger or miRNA* strand) strand

(Khvorova et al., 2003; Schwarz et al., 2003). In plants, miRNAs show perfect complementarity to their target mRNA sequences and guides the degradation of the target mRNA or blocks the translation (Mallory and Vaucheret, 2006; Zhu, 2008; Brodersen and Voinnet, 2009).

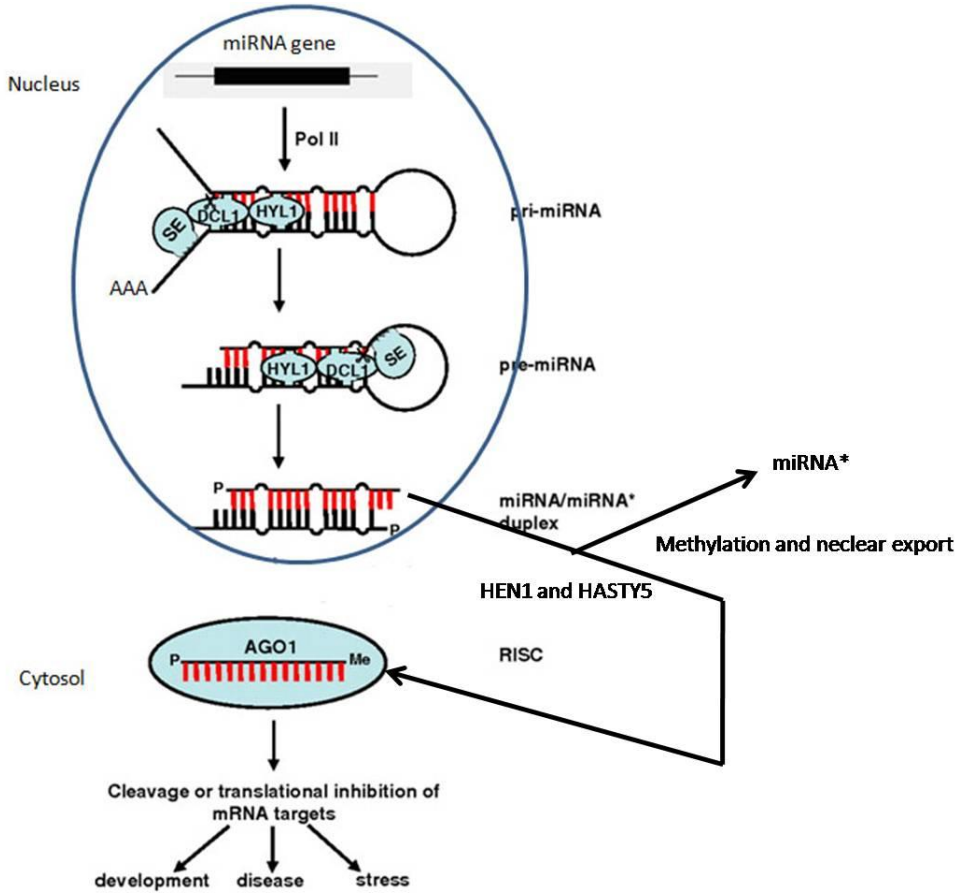


Figure 2. miRNA biogenesis and function in plants (Modified from: Zhu, J.K. (2008). Reconstituting plant miRNA biogenesis. Proc Natl Acad Sci U S A 105, 9851-9852.)

To date, small RNAs have been identified in rice, *Arabidopsis*, *Populus*, *Physcometrella*, *Medicago truncatula*, grapes, switchgrass, tomato and several other plant species (Arazi et al., 2005; Sunkar et al., 2005; Lu et al., 2006; Talmor-Neiman et al., 2006a; Axtell et al., 2007;Fahlgren et al., 2007; Lu et al., 2008; Sunkar & Jagadeeswaran, 2008; Zhu et al., 2008; Jagadeeswaran et al., 2009; Matts et al., 2010). These studies indicated that lineage-specific and species-specific miRNAs are expressed in plants, besides the well-conserved 23 families of

miRNAs. The conserved miRNAs have conserved roles while species-specific miRNAs are thought to play important roles in species-specific processes.

CHAPTER III

METHODOLOGY

I. Collection of different plant tissue and RNA isolations

Different tissue samples ((3-week-old seedlings, middle leaves from the 6-week-old plants, middle leaves from the adult plants, flag leaves, stems, roots, emerging inflorescence and mature inflorescence in which seed setting has initiated) were collected and immediately frozen in liquid nitrogen and stored at −80°C until use. For RNA isolations, the tissue was ground to a fine powder in presence of liquid nitrogen in a clean autoclaved mortar and pestle. To the 100-120 mg tissue powder, 1 ml Trizol (Invitrogen) was used. Chloroform (200µl chloroform per 1ml Trizol) was added and inverted the tube for several times to mix thoroughly and left the tube at room temperature for 2-3 min. Then, the tubes were centriguged at 13,000rpm for 15min and transferred the aqueous phase to a new tube. Equal volume of isopropanol was added to the aqueous phase, mixed thoroughly and incubated on ice for 30min. The tubes were then centrifuged at 13,000rpm for 15min again and the supernatant was discarded. The pellet was washed using 80% ethanol by spinning at 6,000rpm for 5min. The supernatant was discarded and the pellet was air dried briefly (5 min) and completely dissolved in DEPC treated water.

II. Small RNA library construction

Total RNA was isolated from the three-week-old Sorghum seedlings using Trizol reagent. Small RNAs of the desired size range (18-24 nt) were gel-isolated (denaturing 15% polyacrylamide gel) from total RNA. Small RNAs were dephosphorylated and then ligated to a 3’ RNA oligonucleotide adapter. The ligation product was recovered from the gel and re-phosphorylated at the 5’end of small RNAs and receovered after ethanol precipitation. Next, the 5' RNA adapter

was ligated and the ligation product was excised and eluted from the gel. Reverse transcription reaction was performed using the RT primer (AAGGATGCGGTTAAA), subsequently PCR was performed using the forward (TACTAATACGACTCACTAAA) and reverse (AAGGATGCGGTTAAA) primers. A small aliquot (1-2 ul) of the final PCR product electrophoresed using 3% low-melting agarose gel along with a 25bp DNA ladder. The final PCR product was isolated from the gel, purified and sequenced at the Illumina Inc. The schematic presentation of the process is showed in Figure 3.

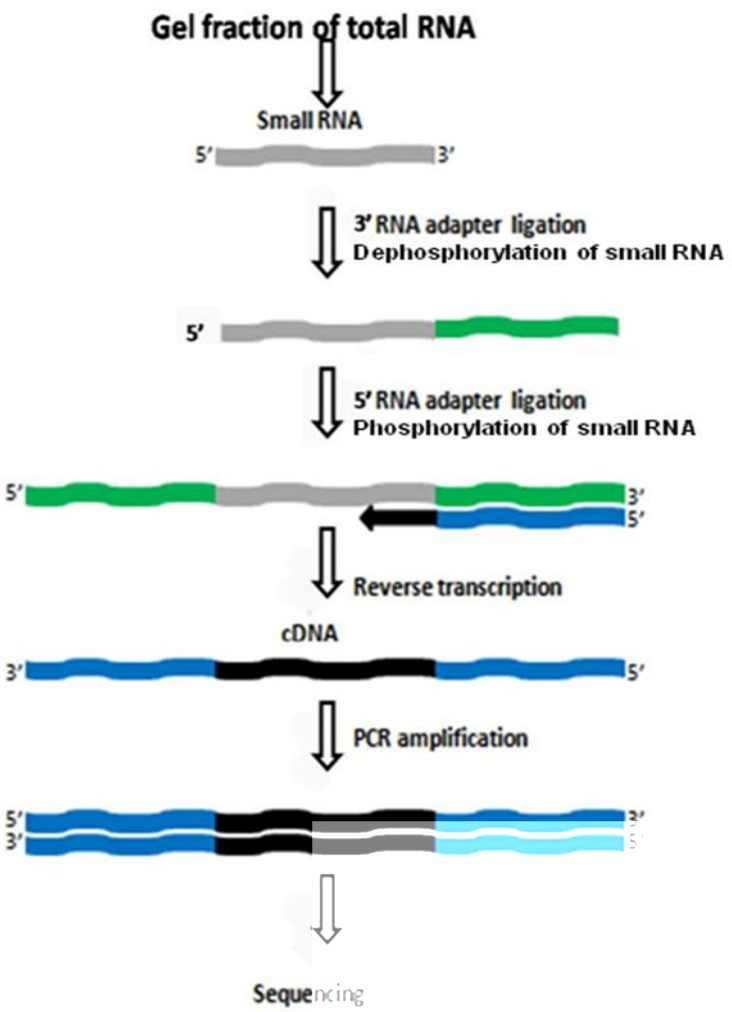


Figure 3. Schematic presentation of construction of a small RNA library (Modified from: Meyers et al., 2006).

III. Sequence analysis of small RNAs

All reads without perfect matches to the most proximal 11 nt of the 5'adaptor sequence were first removed. The adaptor sequences in the remaining reads were trimmed and the small RNAs in between the adaptors were extracted. The redundant sequences were eliminated and the count of unique small RNAs was established. The unique small RNAs were aligned to the Repbase (version 14.01, obtained from http://www.girinst.org), the TIGR Sorghum Repeats DB (http://plantrepeats.plantbiology.msu.edu/downloads.html), known noncoding RNAs (rRNAs, tRNAs, snRNAs, snoRNAs, etc., obtained from http://www.sanger.ac.uk/Software/Rfam/ftp.shtml) and the mRNA of Sorghum bicolor (http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, filtered models) with the NCBI BLASTN. Small RNAs that were mapped to these non-coding RNAs were removed form the data set. Then the small RNAs were mapped to the reported miRNAs in the miRBase (release 13, obtained from http://microrna.sanger.ac.uk/sequences/ftp.shtml). Small RNAs that were mapped to known miRNAs of Sorghum bicolor or other plant species resulted in identification of conserved miRNA homologs in Sorghum. The remaining unique small RNAs were aligned to the genome sequence of Sorghum bicolor (downloaded from http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, masked assembly) with BLASTN. Unique small RNAs with more than 10 genomic hits were removed from further analysis, assuming that these small RNAs might have been derived from repeat-rich loci. The flanking regions of the remaining genome-matched sequences were cut out, and the fold-back structures were predicted using the RNAfold program (Hofacker, 2003). The small RNAs for which a fold-back structure could be predicted were considered as potential new miRNAs. Then, the miRNA* sequence could be predicted based on the criterion that there were 2-nt overhangs at the 3' end and the existence of such sequences in the small RNA populations were searched.

IV. Small RNA Blot Analysis

Low-molecular weight RNA (20µg) from 3-week-old seedlings, middle leaves from 6-week-old plants, middle leaves from adult plants, flag leaves, stems, roots, young inflorescence and mature inflorescence was resolved on a 15% polyacrylamide gel containing 7M urea in TBE buffer (45 mM Tris-borate, pH 8.0 and 1.0 mM EDTA), along with labeled 21-24 nt RNA markers. The small RNAs were size fractionated electrophorectically. Stained the gel in 0.5 × TBE buffer containing ethidium bromide for 5min. RNA was then transferred to Hybond-N+ (Amaresham) membranes using a wet-blot transfer unit (Hoefer). Following the transfer, the membrane was UV cross-linked (Stratalinker) and baked for 1 h at 80°C. Radiolabelled probes were made by end-labeling DNA oligonucleotides complementary to miRNA sequences with γ-32P-ATP by using T4 polynucleotide kinase (NEB). Blots were pre-hybridized for at least 1 h and hybridized overnight using PerfectHYB Plus buffer (Sigma) at 38°C. Blots were washed three times at 50 °C with washing buffer (2xSSC, 0.1% SDS) and autoradiographed using a phosphorimager. The expressed miRNAs gave signals at sizes around ~21 nt. Membranes were stripped and re-probed with a labeled U6 (small nuclear RNA), which served as a loading control.

V. Bioinformatic prediction of miRNA targets

To predict potential targets for Sorghum miRNAs, the annotated Sorghum coding sequences were extracted and used for searching complementary sequences to the miRNAs (http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, filtered models). In predicting targets we allowed a maximum of 3.5 mismatches between the miRNA and its target mRNA (Sunkar et al., 2005; 2008).

VI. Target gene validation by mapping cleavage site on the target mRNA

A modified RNA ligase-mediated rapid amplification of cDNA ends (5'RACE) was used to verify whether or not the predicted miRNA target is subjected to cleavage in vivo (Llave et al., 2002). Messenger RNAs that are not targeted by the miRNAs will be intact with polyA tail at 3'end and cap structure at the 5'end. However, messenger RNAs, which are subjected to the miRNA-guided cleavage are sliced between 10^{th} and 11^{th} of the complementary region. Such cleaved fragments possessing 5'phosphate can be ligated with an RNA adaptor using T4 RNA ligase. The ligated mRNA is reverse transcribed by SuperScript^{TM} II RT and oligo dT primer to synthesize the 1^{st} strand cDNA which has the known priming sites at the 5' and 3' ends. The cDNA was subjected to an amplification procedure with the GeneRacer 5' primer and gene-specific primers followed by a nested PCR with GeneRacer 5' nested primer gene specific nested primers The amplified products were gel purified, cloned and sequenced.

CHAPTER IV

FINDINGS

Currently two approaches (cloning and computational) have been widely used to identify miRNAs in different plant species (Jones-Rhoades and Bartel et al., 2004; Zhang et al., 2006; Sunkar and Jagadeeswaran, 2008). Conserved miRNAs can be identified bioinformatically, but the knowledge of genome sequence of the plant species is a pre-requisite for such a purpose. However, experimental approach is straight forward and has the potential to identify species-specific, novel and atypical miRNAs, in addition to the conserved miRNA homologs.

I. Sequence analysis of the small RNA library

To identify miRNAs expressed in sorghum, a small RNA library was constructed by using RNA isolated from the 3-week-old seedlings. The library was sequenced using Sequencing-By-Synthesis technology (Illumina). A total of 619,010 raw sequences ranging in size between 18 nt and 26 nt were obtained after removing the sequences that does not possesses the recognizable adapter sequence (Table 1). The highest abundance of sequences is around the lengths 21, 22 and 24 nt, which is consistent with the size of the Dicer products (Figure 4) (Lu et al., 2005; Fahlgren et al., 2007; Jagadeeswaran et al., 2009). In plants, small RNA populations have been shown to have two peaks, one at 21 nt and the other at 24 nt. The 24-nt peak is much larger than the others (Figure 4), which is also consistent with several reports from plants (Lu et al., 2005; Fahlgren et al., 2007; Jagadeeswaran et al., 2009). From the total reads, a dataset of unique reads with their read counts was established after removing the redundant sequences. This set of sequences was used to search for the breakdown products from rRNA, tRNA, small nuclear RNA and small nucleolar RNA. Reads with perfect matches with the above non-coding RNAs were excluded from the further analysis. Similarly, the reads with perfect hits to the messenger RNAs in sense

orientation were also eliminated from the analysis assuming that these are degradation products derived from mRNAs. The filtered reads were used to identify conserved miRNA homologs by mapping to the miRBase. The remaining set of sequences with perfect matches to the Sorghum genome was used to identify novel miRNAs in Sorghum.

Reads	Number of unique reads	Number of total reads
Genome matching reads	359936	492803
messenger RNA	58397	109087
miRBase	260	25733
Rfam	26456	71099
Repeats	24574	55954
Mitochondrion/chloroplast	10638	18342
Total	464763	619010

Table 1. Summary of sequence analysis of small RNA library

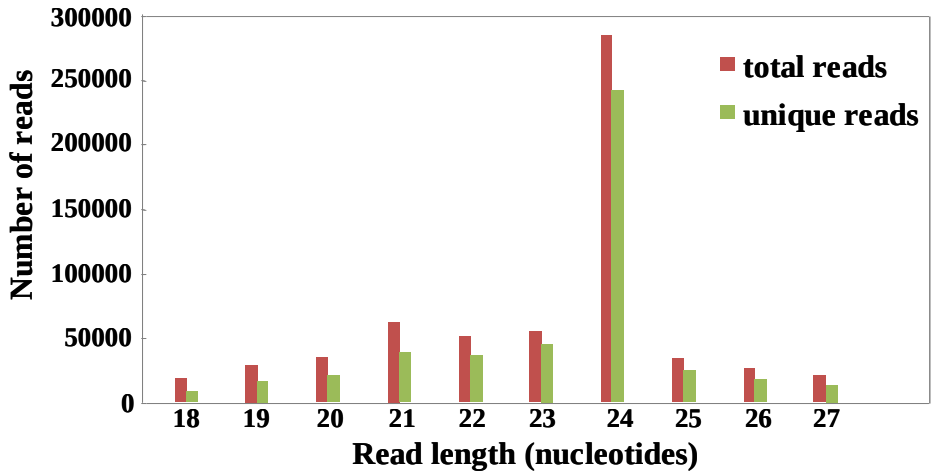


Figure 4. Abundance of 18-27 nucleotide small RNAs in the small RNA library.

II. Identification of conserved miRNAs

Twenty three miRNA (miR156/157, miR159, miR160, miR162, miR164, miR165/166. miR167, miR168, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396,

miR397, miR398, miR399, miR408 and miR827) families are highly conserved between dicotyledonous and monocotyledonous plants, whereas miR403 is conserved in dicotyledonous plants only and miR396d/e, miR437 and miR444 are conserved in monocotyledonous plants only (Sunkar and Jagadeeswaran, 2008; Jones-Rhoades et al., 2006). On the other hand, a few miRNA homologs such as miR158 and miR391 are conserved only among the members of Brassicaceae and miR2118, miR2119 and miR2199 are conserved only among the members of leguminaceae, which can be designated as lineage-specific miRNAs. Besides these, studies in Arabidopsis, rice, Medicago truncatula and other plant species indicated the existence of several species-specific miRNAs (Arazi et al., 2005; Lu et al., 2006; Axtell et al., 2007;Fahlgren et al., 2007; Lu et al., 2008; Jagadeeswaran et al., 2009).

After extrcating the small RNAs from the library, the conserved miRNAs in Sorghum were identified by homology searches against the miRBase. This analysis revealed the identification of 113 conserved miRNA homologs belonging to 28 distinct miRNA families in Sorghum (Table 2). With the exception of miR162 and miR399 families, the remaining miRNA families that are broadly conserved between the dicots and monocots (miR156/157, miR159, miR160, miR164, miR165/166. miR167, miR168, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396, miR397, miR398, miR408 and miR827) were identified in Sorghum (Table 2). miR399 is known to be specifically induced under phosphate-limiting conditions (Fujii et al., 2005; Chiou et al., 2006). miR162 is highly conserved miRNA and targets DCL-1, the component that processes miRNAs from their precurosors. Bioinformatic analysis suggested the conservation of miR162 in sorghum but the absence of miR162 sequence in the library implies it is not expressed or expressed only at extremely low levels in seedlings. Consistentwith our observation, miR162 was expressed at extremely low levels in rice seedlings (Sunkar et al., 2008).

miRNA families such as miR396d/e, miR437 and miR444, which are known to be expressed among the monocots were also recovered in the library (Table 2). Few other homologs of known but less conserved miRNA families such as miR529, miR530, miR894, miR896, miR1318, miR1436 and miR2910 were also identified in sorghum (Table 2). The fold-back structures for all these Sorghum miRNA precursors could be predicted using mfold program (Zuker, 2003) (Figure 5).

The detection of a few miRNA homologs in Sorghum is worth pointing here in view of their distribution in different plant species. For instance, miR529 in rice and maize, miR530 in rice and Populus, miR894 and miR896 in Physcometrella patens, miR2118 in several legumes is known (miRBase). Similarly, miR1318, miR1436 and miR2920 from rice have been reported (miRBase). These results suggest that besides the well-conserved miRNAs such as miR396d/e, miR437 and miR444 among the monocots, miR529, miR1318 and miR1436 also conserved at least in some of the monocots, but not in all monocots. Similarly though absent in Arabidopsis, miR530 has been detected among several dicots and monocots. Thus far, miR894 and miR896 have been reported only from Physcometrella patens, a moss, whose homologs were not found in Arabidopsis, rice, Populus, Medicago truncatula and maize for which the genomes were sequenced and high-throughput sequencing of small RNA populations have been performed suggesting that these two miRNAs are ancient but were lost in many plant species although appeared to be retained in Sorghum.

The abundance of different miRNA families can be inferred from the frequency of their appearance in the library. The read abundance for each of these miRNA families is highly varied, ranging from single read to 2339 reads. miR166 family is the most abundantly expressed, followed by miR167, miR159, miR169 and miR444 in Sorghum seedlings. By contrast, the least abundant expression was found for miR393 and miR1436, both of which are represented by single read only. Similarly miR390 was represented by 2 reads only in the library. Interestingly,

tas3-siRNAs frequency is much greater compared to the miR390 in Sorghum seedlings. TAS3-siRNA generation is dependent on miR390, because miR390 directed cleavage on TAS3-precursors sets the stage for converting it into dsRNA and subsequenctly processing of TAS3 siRNAs in a phased manner (Allen et al., 2005; Jagadeeswaran et al., 2009).

Most conserved miRNA families are represented by multiple loci (similar in sequence or vary in sequence by one-to-two nucleotides) in plants and each of theses loci appears to differ in their expression, which could confer a tissue- or cell-specific expression of different members within a miRNA family. Therefore, it is important to assess which locus is highly expressed. The expression from each of these loci can be assessed from the frequency of their appearance in the library provided at least these members vary in one nucleotide.

It is also interesting to note that a greater disparity exists among different members of the same miRNA families, i.e., few variants/loci are most abundantly expressed than the others. For instance, miR166a-3 variant is the most abundantly expressed (2339) where as miR166j is the least expressed (36) in seedlings; similarly, miR159a is represented by 368 reads and miR159d is represented by 2 reads; miR167 b by 349 reads and miR167m by 14 reads; miR172a is most abundant (318 reads) whereas miR172e is the least abundant (133 reads), miR169c is the most abundant (298 reads) whereas miR169j,l,m.o.q is least abundant (single read each) (Table 2). Eight of the miR168 members have shown almost similar level of expression as represented by their frequency (Table 2). Interestingly, miR168 is represented by 8 loci in Sorghum whereas miR168 in Arabidopsis, rice and others have fewer loci (one or two), suggesting that miR168 has undergone not only additional duplications but also diverged in its sequence.

On the basis of slight variation in nucleotide sequence, miR169 is the largest miRNA family and is represented by 14 members in Sorghum. This is followed by miR156 (10 members), miR160 (9 members), each of the miR168, miR171 and miR167 families are represented by 8 members,

whereas several miRNA families such as miR319, miR390, miR393, miR398, miR437, miR529 are represented by single locus in sorghum.

Some of the highly conserved miRNAs are induced when specific nutrients are deprived suggesting that miRNAs play an important role in nutrient homeostasis (Sunkar et al., 2007; 2010). The induction of miR395 under sulfate deficiency, miR399 under phosphate deficiency, miR397, miR398 and miR408 under copper deficiency is known in Arabidopsis, Brassica sps and M. truncatula (Yamasaki et al., 2007; Abdel-Ghany & Pilon, 2008; Buhtz et al., 2008). Our library, which was generated from 3-week-old seedlings grown on optimal nutrient levels has miR395 reads although represented by a low frequency (Table 2). The recovery of miR395 reads from sorghum seedlings suggests a low basal expression of miR395 under normal conditions. On the other hand we did not recover any reads belonging to miR399 family, which is induced under phosphate deficiency.

miRNA id#	miRNA sequence	Frequency	Normalized frequency (TPM)
miR156a-c	UGACAGAAGAGAGUGAGCAC	104	168
miR156e	UGACAGAAGAGAGCGAGCAC	14	23
miR156f	UGACAGACGAGAGUGAGCAC	1	2
miR156g	UGACAGAAGAGAUUGAGCAC	1	2
miR156h	UGACAGAAGAGAAUGAGCAC	2	3
miR156i	UGACAGAAGAGAGUGCGCAC	1	2
miR156j	UGACAGAAUAGAGUGAGCAC	2	3
miR156k	UGACAAAAGAGAGUGAGCAC	3	5
miR156l	UGACGGAAGAGAGUGAGCAC	1	2
miR156m	UGACAUAAGAGAGUGAGCAC	1	2
miR159	UUUGGAUUGAAGGGAGCUCUU	430	695
miR159	UUUGGAUUGAAGGGAGCUCUA	428	691
miR159	UUUGGAUUGAAGGGAGCUCUG	368	594
miR159b	CUUGGAUUGAAGGGAGCUCCU	41	66
miR159c	UUUGGAUUGAAGGGCGCUCUA	5	8
miR159d	UUUGGAUUGAAGGGGGCUCUGA	2	3

miR159f	UUUGGGUUGAAGGGAGCUCUGAA	1	2
miR159e	UUUGGAUUGAAUGGAGCUCUU	10	16
miR160a-e	UGCCUGGCUCCCUGUAUGCCA	76	123
miR160f	UGCCGGGCUCCCUGUAUGCCA	1	2
miR160g	UGCCUGACUCCCUGUAUGCCA	2	3
miR160h	UGCCUGGCUCCCUAUAUGCCA	1	2
miR160i	UGCCUGGCUCCCUGAAUGCCA	9	15
miR160j	UGCCUGGCUCCCUGAAUGCCU	9	15
miR160k	UGCCUGGCUCCCUGUAAGCCA	1	2
miR160l	UGCCUGGCUCCCUGUAGGCCA	1	2
miR160m	UGCCUGGCUCCCUGUAUGCCU	7	11
miR164	UGGAGAAGCAGGGCACGUGCA	7	11
miR164b	UGGAGAAGCAGGGCACGUGCU	7	11
miR164c	UGGAGAAGCAGGACACGUGAG	8	13
miR164d	UGGAGAAGUAGGGCACGUGCA	3	5
miR164e	UGGAGAAGUAGGGCACGUGCU	3	5
miR165	UCGGACCAGGCUUCAUCCCCC	71	115
miR166a-d	UCGGACCAGGCUUCAUUCCC	2300	3715
miR166e	UCGGACCAGGCUUCAAUCCCU	228	368
miR166f	UCGGACCAGGCUUCAUUCCUC	2331	3765
miR166f-2	UCGGACCAGGCUUCAUUCCUCA	2331	3765
miR166g	UCGGACCAGGCUUCAAUCCCU	228	368
miR166h	UCGGACCAGGCUUCAUGCCCC	92	149
miR166i	UCGGACCAGGCUUCAUGCCUC	87	141
miR166j	UCUGACCAGGCUUCAUUCCCC	36	58
miR166k	UCUGACCAGGCUUCAUUCCUC	38	61
166a-2	UCGGACCAGGCUUCAUUCCCCC	2311	3733
166a-3	UCGGACCAGGCUUCAUUCCCCU	2339	3778
miR167a,b	UGAAGCUGCCAGCAUGAUCUA	349	564
miR167c-g	UGAAGCUGCCAGCAUGAUCUG	356	575
miR167h	UGAAACUGCCAGCAUGAUCUAU	4	6
miR167i	UGAAACUGCCAGCAUGAUCUGA	4	6
miR167j	UGAAGCUGCCAGCAUUAUCUA	9	15
miR167k	UGAAGCUGCCAGCCUGAUCUGA	26	42
miR167l	UGAAGCUGCCCGCAUGAUCUGA	28	45
miR167m	UGAAGCUGCCCGCCUGAUCUGA	14	23
miR168	UCGCUUGGUGCAGGUCGGGAA	20	32
miR168	UCGCUUGGUGCAGAUCGGGAC	50	81
miR168b	UCGCUUGGUGCAGCUCGGGAC	52	84

miR168c	UCGCUUGGUGCAGAUCUGGAC	33	53
miR168d	UCGCUUGGUGCAGAUUGGGAC	38	61
miR168e	UCGCUUGGUGCAGCUCGGGCC	49	79
miR168f	UCGCUUGGUGCAGGUCGGGAC	21	34
miR168g	UCGCUUGGUGCAUAUCGGGAC	30	48
miR168h	UCGCUUGGUGCCGCUCGGGAC	32	52
miR169a	CAGCCAAGGAUGACUUGCCGA	49	79
miR169b	CAGCCAAGGAUGACUUGCCGG	49	79
miR169c,d	UAGCCAAGGAUGACUUGCCUA	149	241
miR169e-h	UAGCCAAGGAUGACUUGCCUG	148	239
miR169i	UAGCCAAGAAUGACUUGCCUA	45	73
miR169j	CAGCCAAGAAUGACUUGCCGU	1	2
miR169k	CAGCCAAGGAUUACUUGCCGG	2	3
miR169l	CAGCCAAGUAUGACUUGCCGG	1	2
miR169m	CAGCCAAUGAUGACUUGCCGA	1	2
miR169n	UAGCCAAGAAUGGCUUGCCUAUC	36	58
miR169o	UAGCCAAGCAUGACUUGCCUA	1	2
miR169p	UAGCCAAGGAUAACUUGCCUC	2	3
miR169q	UAGCCAAGGAUGACUUCCCUA	1	2
miR169R	UAGCCAAGGAUGACUUGCCUA	50	81
miR170	UGAUUGAGCCGUGUCAAUAUC	3	5
miR171a,b,d	UGAUUGAGCCGUGCCAAUAUC	126	204
miR171e,f	GUGAGCCGAACCAAUAUCACU	11	18
miR171g	UGAUUGAGCCGGGCCAAUAUC	3	5
miR171h	UGAUUGAGCCGUGCCAAUAUU	124	200
miR171i	UGAUUGAGCCGUGCCACUAUC	14	23
miR171j	UGAUUGAGCCGUGCCCAUAUC	10	16
miR171k	UUGAUUGAGCCGUGCCAAUAUC	71	115
miR172	AGAAUCUUGAUGAUGCUGCAG	146	236
miR172a-d	AGAAUCUUGAUGAUGCUGCA	159	257
miR172e	UGAAUCUUGAUGAUGCUGCAC	133	215
miR319	UUGGACUGAAGGGUGCUCCC	84	136
miR319a-2	UUGGACUGAAGGGUGCUCCC	84	136
miR390	AAGCUCAGGAGGGAUAGCGCC	2	3
miR393	UCCAAAGGGAUCGCAUUGAUC	1	2
miR394a,b	UUGGCAUUCUGUCCACCUCC	21	34
miR394c	UUGGCAUUCUGUCCUCCUCC	1	2
miR395a-e	GUGAAGUGUUUGGGGGAACUC	15	24
miR395f	AUGAAGUGUUUGGGGGAACUC	15	24

miR395g,h	UGAAGCGUUUGGGGGAACUC	1	2
miR395i	UUGAAGUGUUUGGGGGAACUC	15	24
miR396	UCCACAGGCUUUCUUGAACUG	254	410
miR396	UCCACAGGCUUUCUUGAACGG	228	368
miR396a,b	UUCCACAGCUUUCUUGAACUG	66	107
miR396c	UUCCACAGCUUUCUUGAACUU	65	105
miR396d,e	UUCCACAGCUUUCUUUAACUG	5	8
miR397a	UUGAGUGCAGCGUUGAUGAGC	14	23
miR397b	UUGAGUGCAGCGUUGAUGAGU	14	23
miR398	UGUGUUCUCAGGUCGCCCCCG	2	3
miR408a	CUGCACUGCCUCUUCCCUGGC	31	50
miR408b	CUGCACUGCCUCUUCUCUGGC	1	2
miR437	AAAGUUAGAGAAGUUUGACUU	2	3
miR444a	UGCAGUUGCUGCCUCAAGCUU	38	61
miR444b	UGUUGUCUCAAGCUUGUUGCC	37	60
miR444c_1	UGCAGUUGUUGUCUCAAGCUU	250	404
miR444c_2	UGCAGUUGUUGUCUCACGCUU	10	16
miR444c_3	UGCAGUUGUUGUCUCCAGCUU	14	23
miR444c_4	UGCAGUUGUUGUCUCGAGCUU	4	6
miR444c_5	UGCAGUUGUUGUCUCUAGCUU	2	3
miR444c_6	UGCAGUUGUUGUCUUAAGCUU	3	5
miR444c_7	UGCAGUUGUUGUUUCAAGCUU	4	6
miR444d_1	UUUGCUGCCUCAAGCUUCCUGC	1	2
miR444d_2	UGCAGUUGUUGCCUCAAGCUU	57	92
miR444d_3	UUGUGGCUUUCUUGCAAGUUG	1	2
miR528	CCUGUGCCUGCCUCUUCCAUU	5	8
miR529a	GCUGUACCCUCUCUCUUCUUC	4	6
miR530	UGCAUUUGCACCUGCACCUCC	5	8
miR827a	UUAGAUGACCAUCAGCAAACA	41	66
miR827b	UUAGAUGACCAUCAGUAAACA	2	3
miR827c	UUAGAUGACUAUCAGCAAACA	1	2
miR827d	UUAGAUGAUCAUCAGCAAACA	1	2
miR827e	UUAGAUGACCAUCAACAAACA	1	2
miR827f	UUAGAUAACCAUCAGCAAACA	1	2
miR894	CGUUUCACGUCGGGUUCACCA	555	897
miR896	GCGGAUUUGGCCGAGUGGUUAAGG	9	15
miR1126	UCCACUAUGGACUACAUACGGAG	1	2
miR1318	CUCAGGAGAGAUGACACCGA	6	10
miR1436	UAUUAUGGGACGGAGGGAGUAGU	1	2

miR2118	UUCCUGAUGCCUCCCAUGCCUA	7	11
miR2118	UUCCUGAUGCCUCUCAUUCCUA	1	2
miR2910	UAGUUGGUGGAGCGAUUUGUC	473	764
tasiRNA3a	UUCUUGACCUUGCAAGACCUU	6	10
tasiRNA3b	UUCUUGACCUUGUAAGACCCA	45	73

Table 2. Conserved miRNA homologs and their frequency in the small RNA library.

miR156

```
u        -     -         a      -c      acg
 gacagaa gaga gugagcac cggcg  gacga   gcaua
 ||||||| |||| |||||||| |||||  |||||   |||| a
 cugucuu cucu cacucgug gccgc  cugcu   uguau
a        u     u         c      uc      gua
```

miR159

```
a   a       a   u     u  ag    cuuuu  -u   -    u c   g    u     guuc    uau         ucauguaugugguauguacucuagaggg
 gcg agcuccu uca uccaa ga  ggcc     ca  ggg uggu c gcu cucg ucaug    ccac   ccuaucuca                            c
 ||| ||||||| ||| ||||| ||  ||||     ||  ||| |||| | ||| |||| |||||    ||||   |||||||||                            c
 cgu ucgaggg agu agguu cu  ccgg     gu  ccc gcca g cga gagc aguac    ggug   ggauagagu                            c
a   c       a   u     u  cg    ---uc  uc   u    c u   g    c     guuu    uuc         uucugacugcuggguguacuuagagaagag
```

miR160

```
u  c        c           a    -     cc      caua
 gc uggcucc uguaugccac cau guag  caacc    u
 || ||||||| |||||||||| ||| ||||  |||||
 cg accgggg acgugcggug gug cguc  guugg    g
a  a        a           g    u     -c      aauc
```

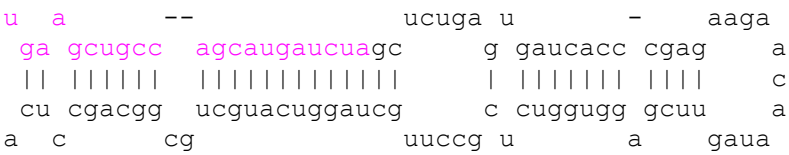
miR164

```
u        ca          cauu         augccgccaagcucgauccuccucugagcuu
 ggagaag  gggcacgug     accaucca                                  g
 |||||||  |||||||||     ||||||||
 ccucuuc  cccguguac     ugguaggu                                  c
a        ua          uucu         aggucgguaccgaccgcucgacuaccucgau
```

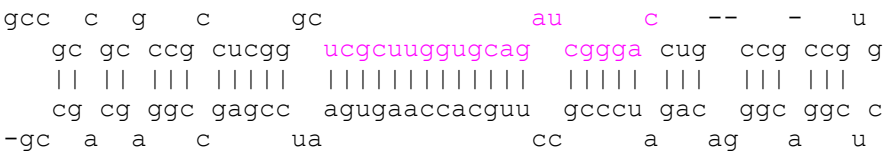
miR166

```
gu       uu          a         cuugu           -   uuug   a
  ggaaug  gucugguuc aggucucg      gauuuaagga uga    ugc u
  ||||||  ||||||||| ||||||||      |||||||||| |||    |||
  ccuuac  cggaccagg ucuagagu      cuaaguuccu auu    aug g
cc       uu          c         ---au           u   uuua   c
```

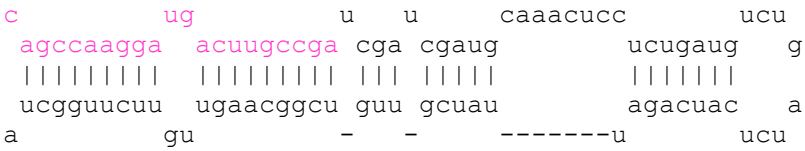
miR167



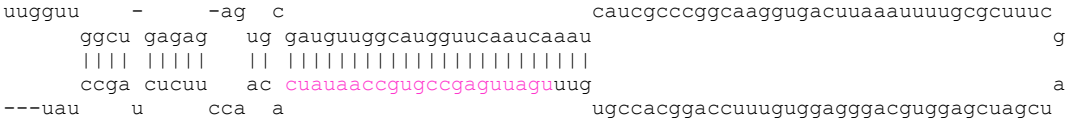
miR168



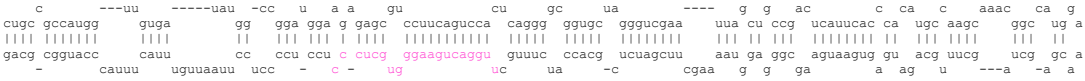
miR169



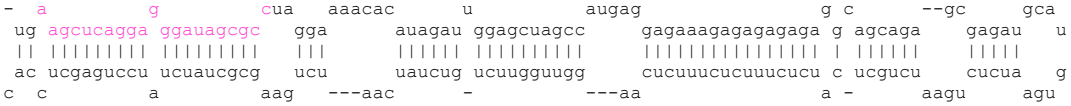
miR171



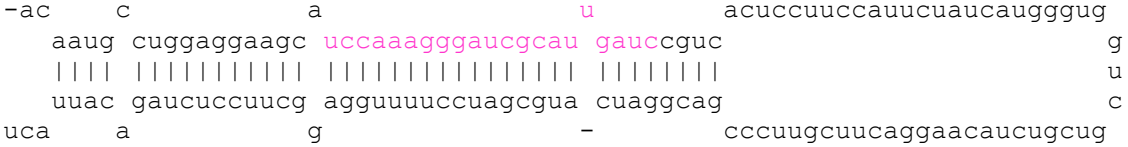
miR319



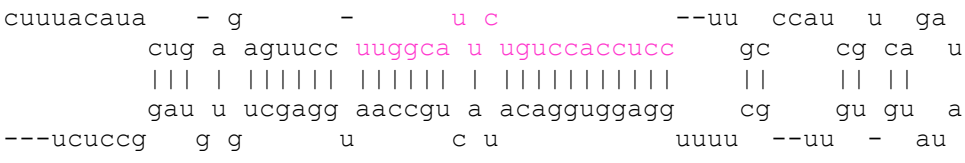
miR390



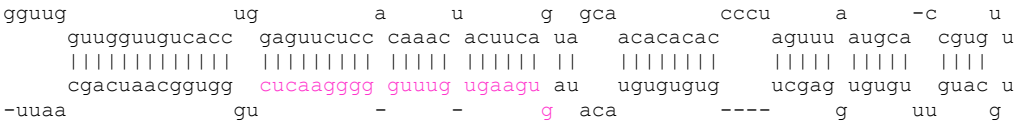
miR393



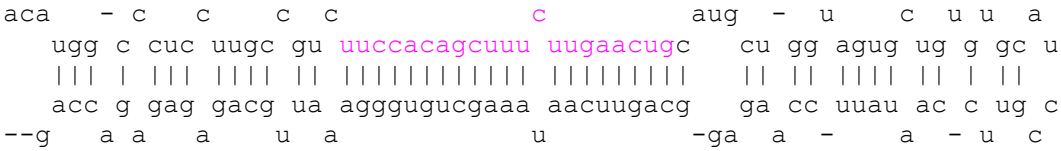
miR394



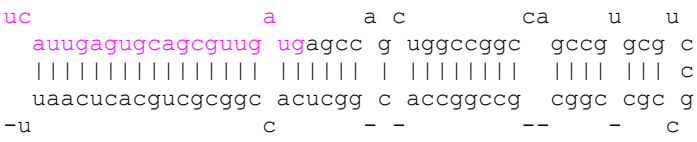
miR395



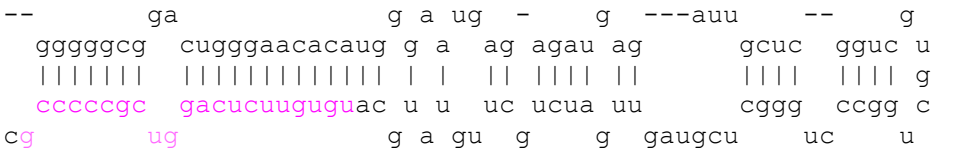
miR396



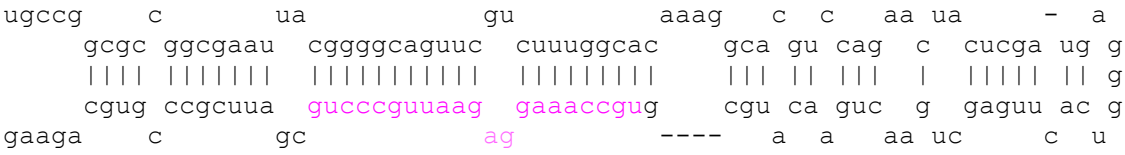
miR397



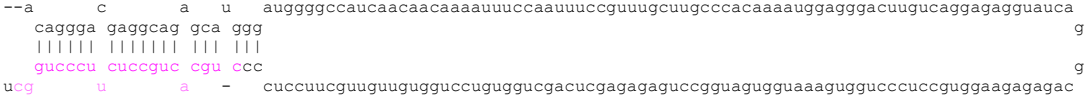
miR398



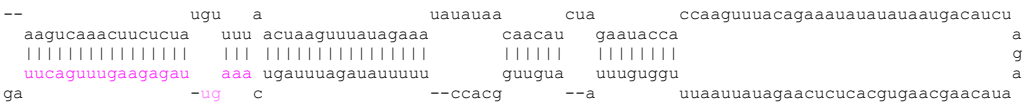
miR399



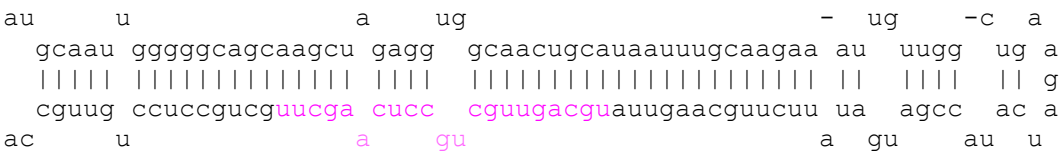
miR408



miR437



miR444



miR529

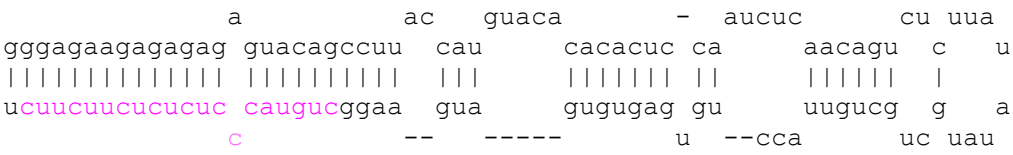


Figure 5. Predicted hairpin-like structures for representative conserved miRNA families in Sorghum. The mature miRNA sequence is highlighted in pink letters.

III. Expression analysis of conserved miRNAs

MicroRNAs described in this study are cloned from Sorghum seedlings. Previous reports from rice, *Arabidopsis*, *Medicargo truncatula* and switchgrass indicated that some miRNAs expressed only in certain cell types or tissues, or certain developmental stages (Chen, 2004; Sunkar & Zhu, 2004; Combier et al., 2006; Boualem et al., 2008; Jagadeeswaran et al., 2009; Matts et al., 2010). It is essential to determine their spatio- and temporal-expression in order to discern their functions. The expression of 15 conserved miRNAs was analyzed in eight different tissues (3-week-old seedlings, middle leaves from 6-week-old plants, middle leaves from adult plants, flag leaves, stems, roots, emerging inflorescence and mature inflorescence in which seed setting has initiated) of sorghum (Figure 6).

We used 20µg of LMW RNA to determine the expression analysis of conserved miRNAs. Of the different tissues, inflorescence had abundant expression of several miRNAs. For instance,

miR172, miR156, miR319, miR159, miR529, miR164, miR160, miR166, miR167 and miR169 showed the highest expression levels in inflorescence relative to other tissues (Figure 6). It is also interesting that the expression levels highly varied between mature and young inflorescence tissues, i.e., particularly, miR156, miR159, miR160, miR164, miR319 and miR172 expressed at very low levels in young inflorescence, whereas miR396 expressed at high levels in mature inflorescence. Also miR156 and miR159 expression levels greatly varied between the leaves from young plants or adult plants, i.e., lower expression levels in leaves from young plants whereas high expression levels in leaves from the adult plant (Figure 6). miR529 had almost similar level of abundance in both young and mature inflorescence tissues (Figure 6).

Both, miR168 and miR167 showed almost uniform expression in all tissues examined with the exception that mature inflorescence had relatively high-level expression (Figure 6). In sorghum seedlings, miR156 and miR169 are abundantly expressed and miR167, miR390 and miR444 are moderately expressed, whereas the expression of miR393, miR396 and miR319 is almost below the detection limit. With the exception of miR169 and miR396 (miR169 and miR396 had low expression levels in stem and root, respectively), the remaining 13 miRNAs did not differ in their abundance between roots and stems.

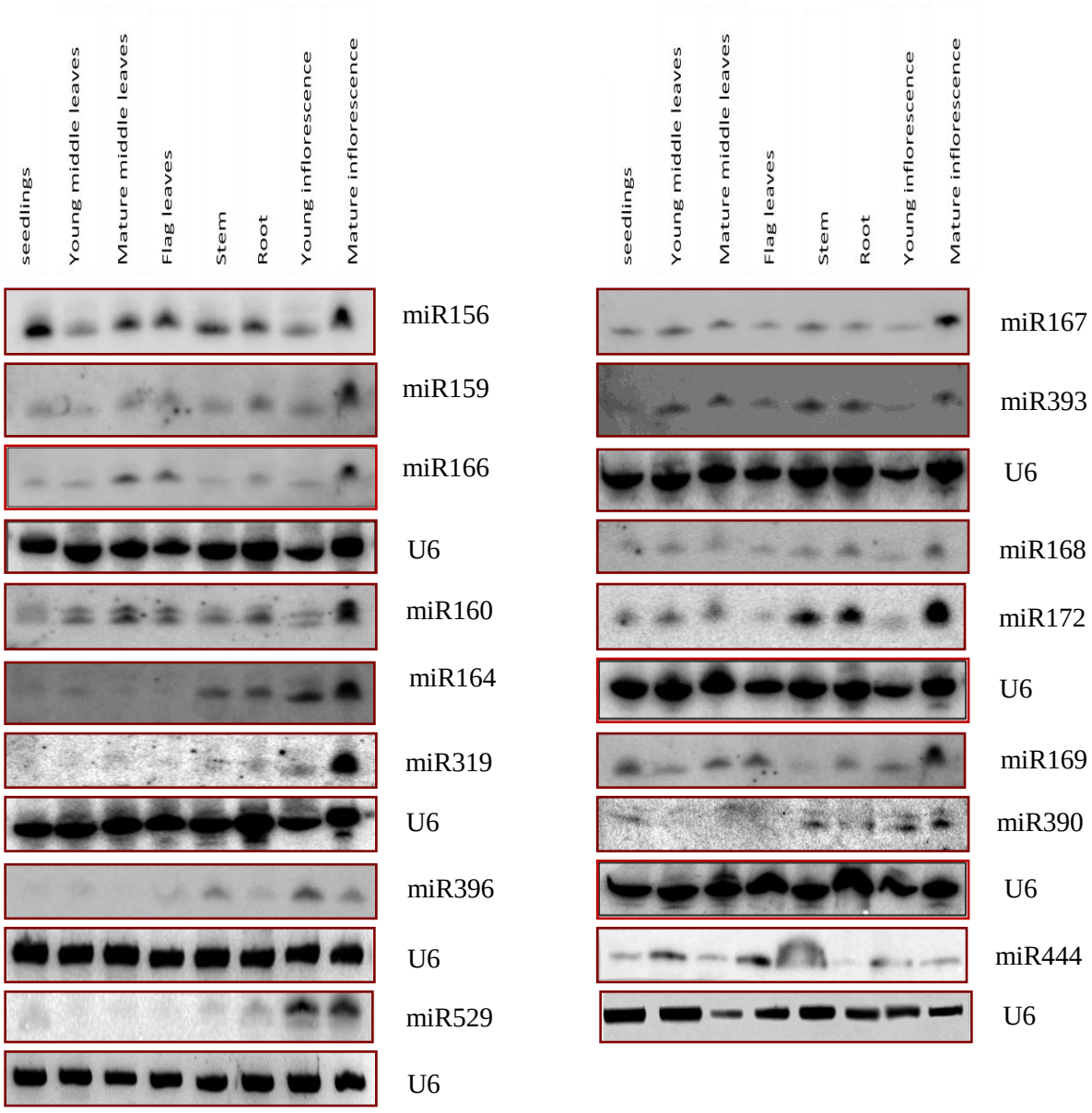


Figure 6. Expression patterns of conserved miRNAs in different tissues. Blots were stripped and re-probed with ^{32}P-end-labelled oligonucleotides complementary to the U6 probe, which served as a loading control.

IV. miRNAs induced under nutrient deprivation

In *Arabidopsis*, miR395, miR399 and miR398 are induced in response to sulfate-, phosphate- and copper-deprived conditions, respectively (Jones-Rhaodes et al., 2004; Fujii et al., 2005; Abdel-

Ghany and Pilon, 2008). Additionally, the expression of miR397 and miR408 are also elevated in response to copper deficiency in Brassica and Arabidopsis (Yamasaki et al., 2007; Abdel-Ghany and Pilon, 2008; Buhtz et al., 2008). Recently similar findings were also reported for *M. truncatula* (Jagadeeswaran et al., 2009). By contrast, miR395 and miR399 are constitutively expressed in plants grown on optimal levels of nutrients but not induced in response to sulfate- and phosphate-deprivation in switchgrass, a plant species adapted to marginal soils with poor nutrient availability (Matts et al., 2010). To analyze the response of miR395, miR397, miR398, miR399 and miR408 in sorghum, 3-week-old seedlings were transferred exposed to sulfate, phosphate or copper-deficit conditions. The results indicated that all these miRNAs are found to be induced both in shoot and root tissues in response to the deficiency of sulfate, phosphate and copper levels in the hydroponic culture (Figure 7).

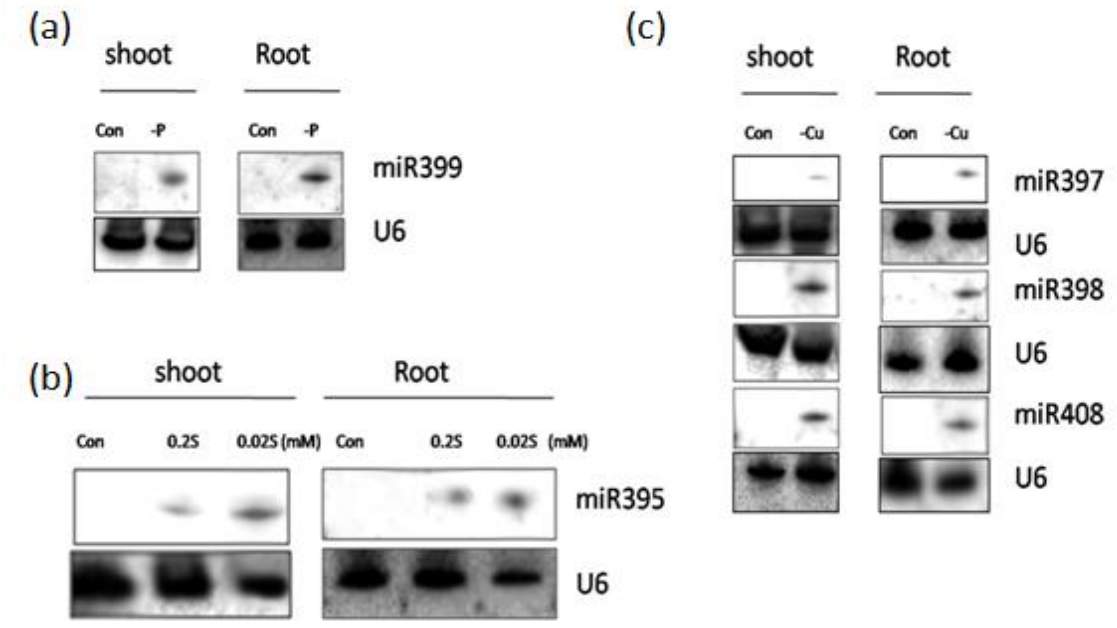


Figure 7. miRNAs induced in response to deprived nutrient levels in the growth medium. (a-c) Small RNA blots of 20µg low molecular weight RNA isolated from sorghum grown continuously in the same growth medium (control) or transferred to medium without (a) phosphate, (b) sulfate or (c) copper. Blots were rehybridized with the U6, which served as a loading control.

V. Prediction of miRNA targets and their validation in Sorghum

A total of 100 genes were predicted as targets for some of the conserved miRNAs (Table 4). We were unable to predict the targets for some of the conserved miRNAs, because the sorghum genome annotation is still incomplete. The predicted targets are predominantly transcription factors: miR156 is predicted to target 9 Squamosa promoter binding transcription factors; miR159 is predicted to target 4 MYB domain containg transcription factors; miR160 is predicted to target 5 auxin response factors; miR164 is predicted to target 6 No Apaical Mersitem containing proteins; miR166 is predicted to target 4 Homeobox domain containing transcription factors; miR167 is predicted to target 3 auxin response factors; miR168 is predicted to target 4 piwi-domain containing Argonaute proteins; miR169 is predicted to target one nuclear transcription factor Y subunit; miR171 is predicted to target 3 SCARECREW (GRAS domain containing) transcription factors; miR172 is predicted to target 2 Apetala 2 related transcription factors; miR396 is predicted to target 6 growth regulating transcription factors and miR444 is predicted to target 5 MADS box transcription factors (Table 3). In addition to these, genes encoding F-box containing proteins (miR394), a sulfate transporter and an ATP sulfurylase (miR395), eleven laccases (miR397), Cu/Zn superoxide dismutases (miR398), two inorganic phosphate transporters (miR399), three plastocyanin-domain containing proteins (miR408), were also among the predicted targets (Table 3). These target genes are likely to be involved in wide variety of physiological processes whose roles remain unknown. To confirm that the predicted targets are true targets in Sorghum, a few predicted target genes were verified using 5'-RACE assay. ATP sulfurylase targeted by miR395, laccase (multicopper oxidase) targeted by miR397, Cu/Zn superoxide dismutase targeted by miR398 and phosphate transporter targeted by miR399 were validated by mapping the respective miRNA-directed cleavages *in vivo* (Figure 8).

miRNA families	Predicted target accession ID	Target gene family
miR156	jgi|Sorbi1|5232347|estExt_fgenesh1_pg.C_chr_33889	SBP domain containing protein
	jgi|Sorbi1|5271494|Sb06g024630	SBP domain containing protein
	jgi|Sorbi1|4977676|estExt_Genewise1.C_chr_28260	SBP domain containing protein
	jgi|Sorbi1|4201067|gw1.2.18533.1	SBP domain containing protein
miR159	jgi|Sorbi1|4982901|estExt_Genewise1.C_chr_39472	MYB family transcription factor
miR160	jgi|Sorbi1|5291326|Sb10g027790	Auxin response factor
	jgi|Sorbi1|5277928|Sb01g019130	Auxin response factor
miR164	jgi|Sorbi1|5121666|estExt_Genewise1Plus.C_chr_81551	NAC domain transcription factor
miR166	jgi|Sorbi1|5230460|estExt_fgenesh1_pg.C_chr_30266	HD-Zip protein
	jgi|Sorbi1|5289000|Sb08g021350	HD-Zip protein
	jgi|Sorbi1|5102053|estExt_Genewise1Plus.C_chr_112335	HD-Zip protein
	jgi|Sorbi1|5098418|estExt_Genewise1Plus.C_chr_14113	HD-Zip protein
	jgi|Sorbi1|5097817|estExt_Genewise1Plus.C_chr_12763	HD-Zip protein
miR167	jgi|Sorbi1|5061512|e_gw1.6.16287.1	Auxin response factor
miR168	jgi|Sorbi1|4995896|estExt_Genewise1.C_chr_90067	argonaute protein
miR169	jgi|Sorbi1|5101414|estExt_Genewise1Plus.C_chr_111272	nuclear transcription factor Y subunit
miR171	jgi|Sorbi1|5235905|estExt_fgenesh1_pg.C_chr_61794	hypothetical protein
	jgi|Sorbi1|5233945|estExt_fgenesh1_pg.C_chr_	hypothetical protein
miR172	jgi|Sorbi1|4987016|estExt_Genewise1.C_chr_48442 jgi|Sorbi1|5267190|Sb06g030070	hypothetical protein AH2 domain transcription factor
miR319	jgi|Sorbi1|5228446|estExt_fgenesh1_pg.C_chr_20329	TCP family transcription factor
	jgi|Sorbi1|5257392|Sb01g006020	TCP family transcription facto
miR390	jgi|Sorbi1|127730|fgenesh1_pm.C_chr_4001033	Leucine rich repeat protein
miR393	jgi|Sorbi1|5082622|e_gw1.9.1294.1	F-box family protein
miR394	jgi|Sorbi1|4983693|estExt_Genewise1.C_chr_310757	F-box domain containing protein
	jgi|Sorbi1|5234325|estExt_fgenesh1_pg.C_chr_43425	F-box domain containing protein
miR395	jgi|Sorbi1|147166|estExt_fgenesh1_kg.C_chr_10143	sulfurylase(sulfate adenylyltransferase)
miR397	jgi|Sorbi1|123904|fgenesh1_pm.C_chr_1000958	Multicopper oxidases
	jgi|Sorbi1|146423|estExt_fgenesh1_pm.C_chr_90369	Laccase
	jgi|Sorbi1|5109791|estExt_Genewise1Plus.C_chr_39807	Laccase
miR398	jgi|Sorbi1|131357|fgenesh1_kg.C_chr_1000400	Cu2+/Zn2+ superoxide dismutase SOD1
miR399	jgi|Sorbi1|142168|estExt_fgenesh1_pm.C_chr_11153	Phosphate transporter
miR408b	jgi|Sorbi1|5257619|Sb01g010520	blue protein precursor
miR444	jgi|Sorbi1|147544|estExt_fgenesh1_kg.C_chr_10531	MADS box transcription factor
	jgi|Sorbi1|144632|estExt_fgenesh1_pm.C_chr_40774	MADS box transcription factor

Table 3. Predicted targets for the conserved miRNAs in Sorghum

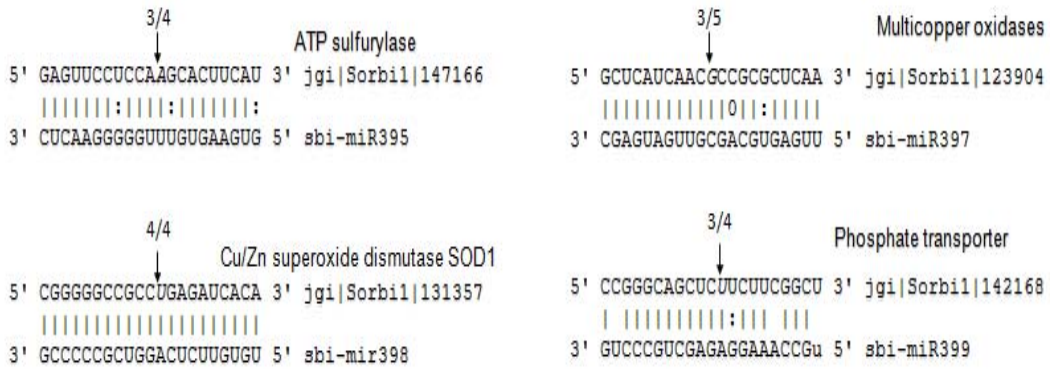


Figure 8. Validation of miRNA (miR395, miR397, miR398 and miR399) target (ATP sulfurylase, multicopper oxidase, Cu/Zn superoxide dismutase and phosphate transporter) genes in sorghum. The perfect matches, mismatches and G-U wobbles are represented by straight lines, colons and circles, separately.

VI. Identification, characterization and expression analysis of novel miRNAs

Identifying novel miRNAs which are not conserved or conserved only in closely related species is a challenge given the fact the plant small RNA populations include not only miRNAs but also endogenous siRNAs, which out numbers the miRNA population both in number and diversity. Because of this and the fact that some of the siRNAs in the miRBase have been misannotated as miRNAs, plant small RNA community has established a set of criteria for correct annotation of miRNAs particularly 'non-conserved miRNAs' in plants, which demands the sequencing of a miRNA* read in the library (Meyers et al., 2008). Our sequence analysis revealed 14 candidates as potential novel miRNAs based on the predicted fold-back structures (Figure 9) but only 9 small RNAs are supported by the miRNA* and the remaining five did not have their corresponding star-sequence in the library (Table 4). Limited sequencing depth (about 600,000 reads in this study relative to several millions of reads in other studies) may also be one of the causes for not sequencing a miRNA* read for some of the novel low-abundant miRNAs in sorghum. For five cases (s91586, s449185, s197538, s50685 and s418541) which appeared like novel miRNAs but without their corresponding miRNA* sequences in the library, we conducted

miRNA*expression using small RNA blot analysis. Since the abundance of miRNA* species usually at low levels compared to the miRNA abundance, we used relatively higher amounts of RNA (50µg of LMW RNA) for detecting miRNA*, whereas 20µg was used for detecting the expression of novel miRNAs. By using small RNA blot analysis we were able to detect the miRNA* expression for 4 novel (s91586, s449185, s197538 and s50685) miRNAs. Conservation of novel miRNAs in closely related species could be another supporting evidence for annotation of novel small RNAs as 'novel miRNAs' (Jagadeeswaran et al., 2009). In order to examine the conservation of these novel miRNAs in closely related monocots, blast searches were performed against the EST database. Surprisingly, 7 of these novel miRNA sequences are conserved in sugarcane, maize, and wheat, suggesting that some of these could be designated as monocot-specific or lineage-specific miRNAs (Appendix 1). Predicted fold-back structures for the novel miRNA precursors from sugarcane, wheat and maize are shown (Figure 10). Taken together, on the basis of appearance of miRNA* read in the library or detection of miRNA* expression using small RNA blot analysis and/or their conservation in related monocots, 14 small RNAs were annotated as "novel miRNAs" in Sorghum. Seven novel miRNAs are conserved at least in one another monocot plant species, thus designated as lineage-specific miRNAs where as the remaining seven appears to Sorghum-specific miRNAs. The frequency of a few novel miRNAs (s412459, s91586 and s50685) is substantially higher and even greater than the frequency of several conserved miRNAs (Table 2 and Table 4).

We also analyzed the expression of these novel miRNAs in several tissues. Abundant and ubiquitous expression of s412459 could be detected in all tissues analyzed (Figure 10). By contrast most other novel miRNAs showed low and uniform expression in all tissues with the exception that s252721 and s91586 showed high level expression in root tissue, relative to other tissues.

We also predicted 25 genes as potential targets for the 14 novel miRNAs (Table 5). The novel miRNA and target mRNA alignments are shown in Appendix 3. Most of the predicted novel miRNA targets are hypothetical proteins implying that the predicted targets are novel genes and may have specific functions in sorghum. However, some known genes such as E3 ubiquitin protein ligase (s71509), SNARE protein Syntaxin (s76707), SAM decarboxylase (s2122), putative receptor kinases (s418541) and oxidoredoreductase and arabinogalactan protein (s449185) are also predicted targets for the novel miRNAs in Sorghum (Table 5).

miRNA	miRNA sequence	miRNA frequency	NB	miRNA* Sequence	miRNA* frequency	NB	Conservation
s412459	UGGGGAAGCAAUUCGUCGAAC	1472	+	UCGGCGACUUGCUUCACCCAUG	5		surgarcane
s373158	UCACCGGCGCUGCACUCAAUU	15	+	UUGAGUGCAGCGUUGAUGAGC	14		switchgrass
s252721	GAACAGCGGGGAGGUGCUGCC	2	+	UCAGCAUCACCUCCCUGUUGU	3		
s13121	AAAUUGUAAGUCGUUCUGGCU	7	- -	CAGAGCGACUUACAAUUUGGA	1	- -	wheat
s229270	CGUGGCUCUGACCGGUGCUAAAGG	1	- -	UUUAGCACCGGUUCGUGUUACGAA	1	- -	
s438157	UUGAACUAUGGUAAAAAUUUC	1	- -	AACUUUUACCAUAGUUCAAGC	1	- -	
s2122	AAAAGACAAAUCAGCAUGUCA	1	- -	ACAUGCUGAUUUGUCUUUUGU	1	- -	
s71509	ACUCCAACACAUGUGGAUUGAG	7	- -	UCAAUCUACAUGUGUUGGAGUGG	1	- -	Sugarcane wheat
s76707	ACUUCAAUCCAUGUAUGUUGGUGU	1	- -	CUCAACACAUGUGGAUUGAUGUA	1	- -	
s91586	AGCAAUUCGUCGAACAGCUUGU	130	+		0	+	
s449185	UUGUUUGGAUGUUGUCGGAUUCAC	43	+		0	+	Wheat maize
s197538	CAUCGAAUCUUUAGACGUAUGCAU	55	+		0	+	
s50685	ACACAUGUGGAUUGAGGUGAA	154	+		0	+	Wheat sugarcan
s418541	UGUGGAUUGAGGUGAAUCCGA	34	+		0	-	wheat

Table 4. Identified novel miRNAs in Sorghum on the basis of expression/detection of miRNA* or their conservation in related monocots

s412459

```
cuc   ga  u-      c      -       u            -  -     ca----      ---  c
   ccg  gc  ggccug cguggg gaagcaa ucgucgaacagcu gc ugcgc      gagguu   gg a
   ggc  cg  ccggac guaccc cuucguu agcggcuugucga cg acgug      cuccga   cc g
u--   uc  uc      u      a       c            g  c     caccug      guc  g
```

s91586

```
  uuc    c   c   c     ca               u   -   c-  ---  g     c
gc   ugcc ug cgug gggag  auucgucgaacagcu gag gcg  gc   cg ccggg \
cg   acgg ac gcac uccuc  uaagcagcuugucga cuc cgc  cg   gc ggccc a
  ua-    -  c    c     cg               c   a   ac  acu  g     a
```

s449185

```
 a---------                u  c      acc             u    a  ag   a           ag
          agguuuguuuggaugu gu ggauuc   ucaauccacaugu uugg gu  auu ggguggaauuu  \
          uccggaacaaaccuaca ca ccuaag   aguuagguguaca aacc ca  uaa cucaccuugaa  u
 ugaauauaaa                u  a      cgu             u    c  cu   c           cu
```

s418541

```
.-uuu      a           a                   uggacugaagugg
     ggaugu gucggauuuau ucaauccacauauguuggg             \
     ccuaca cagccuaagug aguuaggugu auacaaccu             a
\ ---      a           g                   uaaaucaaguuua
```

s373158

```
-        auc               a      agc        ca    u   uc
 gcaaaggc   auugagugcagcguug ugagcc   uggccggc  gccg gcg  \
 cguuuccg   uaacucacgucgcggc acucgg   accggccg  cggc cgc  c
u        cau               c      c--        --    -   cg
```

s252721

```
      a   u                          aac   cc
aguuuc gca caccucccuguuguucuccggguacac   cuc  c
uuaaag cgu guggaggggcgacaagagguuuaugug   gag  g
      c   c                          ca-   cu
```

s197538

```
       a             c           a   --   uaa--          a gg -------                  g cuu
auuuaaug uucauguaugugc aaagauucgaug gau  gga     aaaaauuuug u  a       gaacuaaacagggccuaag c   g
uaaauuau agguacguaugcag uuucuaagcuac cug  ccu     uuuuuaaaac g  u       cuugauuuguucuggauuu g   u
       g             a           a   cu   uagac          - uu uaaaaac                  - aua
```

s50685

```
  ag          a               uggu         a   aaacuuga
    ucggauucgc ucaauccacaugugu    guggauugg gug        \
    agccuaagug aguuagguguacaca    caccuaacc cac        a
  ag          g               u---         c   cuuaaauu
```

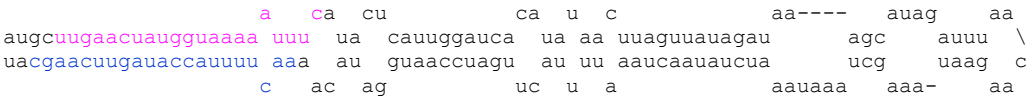
s13121

```
.-u    -    ucccaaca-                                  a   g  ugcu   gc
   gcua guac          uuccaaauuguaagucguucuggcuuuucua gua au    gug  u
   cgau uaug          aagguuuaacauucagcgagaccgagaagau uau ua    cac  a
\ -    a    uuauauaac                                  g   g  uc--   gu
```

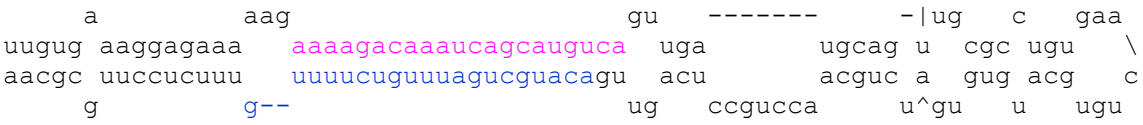
s229270

```
-----    u---        .-a      ucu                   u        acaac  cu
     gguu    gcuucggc   cguggc   gaccggugcuaaaggguc cugcccca     ac  g
     ccag    uggagccg   gcauug   uuggccacgauuuccuag gacggggu     ug  u
uuuua    uuuu        \ -      ugc                   -        gccga  ac
```

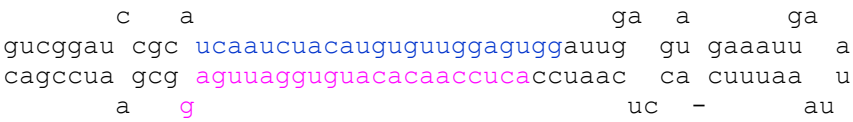
s438157



s2122



s71509



s76707

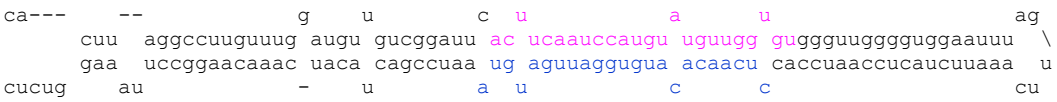


Figure 9. Predicted fold-back structures for the novel miRNA precursor sequences. The sequences colored in red and blue are the mature miRNA sequence and miRNA* sequences, respectively.

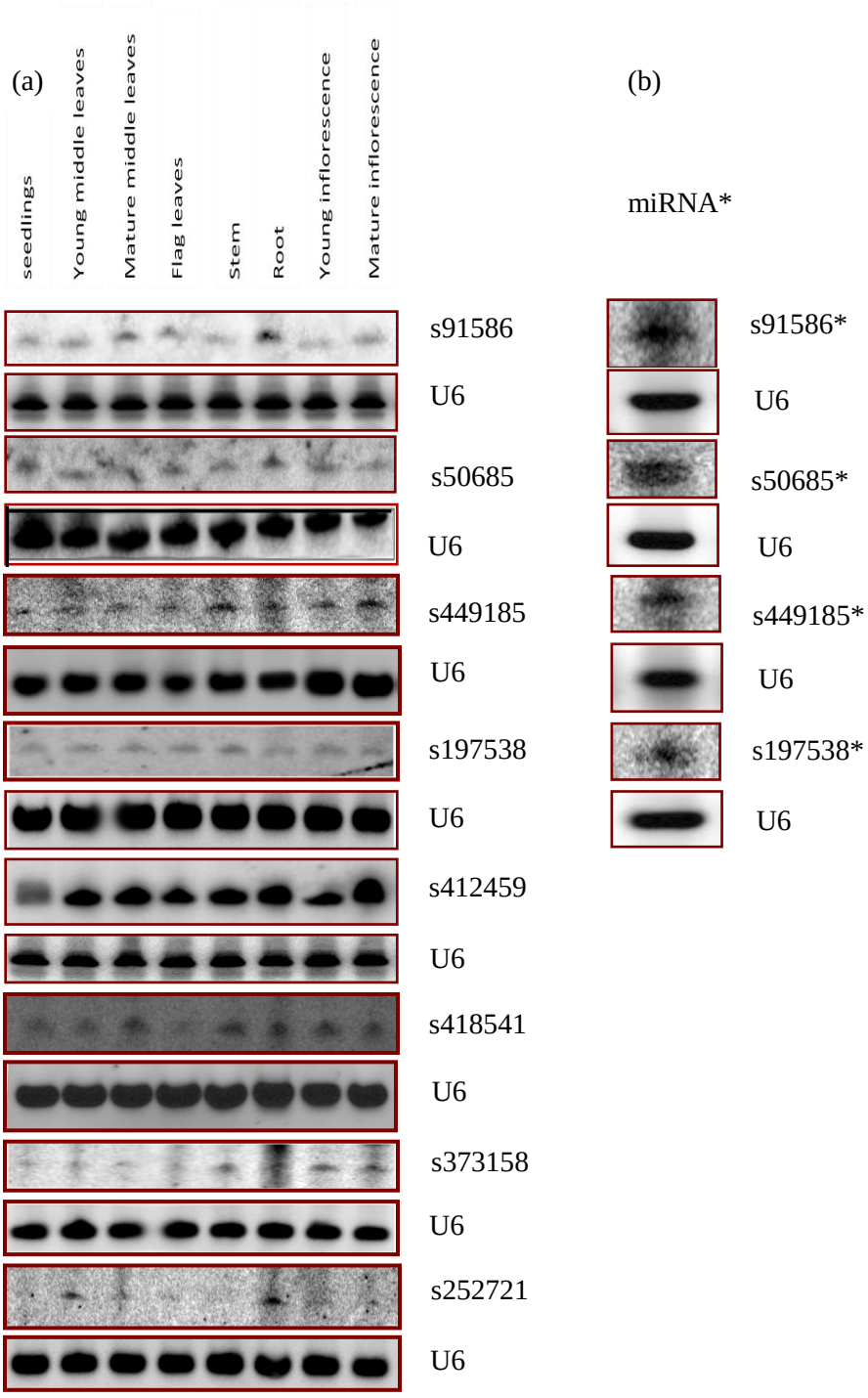


Figure 10. Expression patterns of newly identified miRNAs in Sorghum as determined by small RNA blot analyses. (a) 20µg low molecular weight RNA was used for detection of the miRNA and (b) 50µg low molecular weight RNA was used to detect the expression of miRNA* reads

miRNA	Predicted target gene accession#	Target gene family
s412459	jgi|Sorbi1|5108736|estExt_Genewise1Plus.C_chr_3 8093	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5021956|e_gw1.2.939.1	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|4975996|estExt_Genewise1.C_chr_20831	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5268653|Sb05g002310	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5221943|fgenesh1_pg.C_chr_9002484	Sorghum bicolor hypothetical protein
s91586	jgi|Sorbi1|5108736|estExt_Genewise1Plus.C_chr_38093	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5260497|Sb10g009870	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5268653|Sb05g002310	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5048590|e_gw1.5.179.1	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5234440|estExt_fgenesh1_pg.C_chr_50254	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5268646|Sb05g002250	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5268643|Sb05g002220	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5268631|Sb05g002040	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5215956|fgenesh1_pg.C_chr_7000859	Sorghum bicolor hypothetical protein
s449185	jgi|Sorbi1|5239195|estExt_fgenesh1_pg.C_chr_92061	arabinogalactan protein
	jgi|Sorbi1|5278077|Sb01g022490	oxidoreductase
	jgi|Sorbi1|5277399|Sb01g009180	Sorghum bicolor hypothetical protein
	jgi|Sorbi1|5290360|Sb10g001370	chloroplast zebra-necrosis protein (ZN)
s418541	jgi|Sorbi1|5278077|Sb01g022490	putative receptor protein kinase
s252721	jgi|Sorbi1|124111|fgenesh1_pm.C_chr_1001165	appr-1-p processing enzyme family protein
s197538	jgi|Sorbi1|149258|estExt_fgenesh1_kg.C_chr_70040	Splicing coactivator
s13121	jgi|Sorbi1|5046708|e_gw1.4.48.1	Sorghum bicolor hypothetical protein
s438157	|Sorbi1|147278|estExt_fgenesh1_kg.C_chr_10257	Putative methionine aminopeptidase
s71509	jgi|Sorbi1|4982924|estExt_Genewise1.C_chr_39502	E3 ubiquitin-protein ligase
s76707	jgi|Sorbi1|147110|estExt_fgenesh1_kg.C_chr_10083	SNARE protein (Syntaxin 1)

Table 5. Predicted targets for the novel miRNAs in sorghum.

CHAPTER V

CONCLUSIONS

I. Overview of sequencing results

Unlike in animals, in which miRNAs represent a larger proportion of total small RNA populations, miRNAs only represent a minute fraction (1-2%) of total small RNAs in plants. Unlike, a single peak at 22 nucleotides in the small RNA libraries, which represent miRNAs in animals, usually two peaks (one at 21 nt and the other at 24 nt) were observed in plant small RNA poulations (Figure 4). Of the two peaks, 24 nt size class of small RNAs are the most abundant and very diverse (Figure 4). Because of the established importance of miRNAs in diverse aspects of plant biology, this study focused on miRNAs in sorghum, i.e, identification of conserved and novel miRNAs, their expression patterns and identification of miRNA target genes.

II. Identification and expression analyses of conserved miRNAs

The frequencies of conserved miRNA families varied greatly between 1 and 12093 in the library (Table 2). miR166 family represented as many as 12093 times, of which miR166a alone accounted for 6950. The second most-abundant miRNA family was miR167 represented by 1146 times (Table 2). A total of 28 families represented by either highly conserved (miR156, miR159, miR160, miR164, miR165/166, miR167, miR168, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396, miR397, miR398, miR399, miR408) miRNA families and monocot-specific miRNA families (miR437 and miR444) as well as miRNA families conserved only in some of the monocots (miR529, miR530, miR827, miR894, miR896, miR1318 and miR1436), have been identified in Sorghum. Fifteen of these miRNA families (miR156, mIR159, miR160, miR164, miR167, miR168, miR169, miR172, miR319, miR390,

miR393, miR396, miR444 and miR529) have been analysed for their expression patterns in eight different ttissues (three-week-old seedlings, leaves from young and adult plants, flag leaf, stem, root, young and mature inflorescence) of sorghum plants. Some miRNA families (miR168, miR393, miR444) were expressed at low levels in all tissues examined in this study (Figure 6). On the other hand the expression of some other miRNAs was substantially different in different tissues as the plant develops from juvenile-to-vegetative phase and vegetative-to-reproductive phase. For instance, miR156 showed high expression level in seedling, flag leaves, mature middle leaves and stem than other miRNAs (Figure 6). Ten miRNAs (miR156, miR159, miR160, miR164, miR166, miR167, miR169, miR172, miR319 and miR529) showed high expression levels in mature inflorescence, relative to the other tissues. miR529 is preferentially expressed in inflorescence (young or mature). miR164 and miR319 could not be detected in leaves (middle or flag). Interestingly, the abundance of several miRNAs did not differ between flag leaves and other leaves from the young plant or mature plant implying that the miRNA-guided target gene regulation is not very distinct in flag leaves.

III. Characterization of miRNAs induced under nutrient deprivation

miR395 was induced in shoots and roots, and shoots had a slightly stronger response to low sulfate. Similarly, miR399 was upregulated on phosphate-deficiency in both roots and shoots. miR397, miR398 and miR408 were induced in shoots and roots without copper (Figure 7). As determined by small RNA blot analysis, miR395 could not be detected in seedlings grown on optimal levels of sulfate but the accumulation of miR395 was clearly evident as the sulfate levels dropped in the medium (Figure 7). Similarly, miR399 was also induced under phosphate deficiency (Figure 7). It is also clear that the reported Copper-responsive miRNAs (miR397, miR398 and miR408) in Arabidopsis are also induced under Cu^{2+} deficiency in Sorghum (Figure 7).

IV. Novel miRNAs identification

Recent deep sequencing efforts in Arabidopsis, rice, Medicago truncatula and several other plant species, led to the identification of several novel miRNAs, which are also conserved in closely related spcies (lineage-specific) or not conserved but species-specific miRNAs have been identified (Fhalgren et al., 2007; Rajagopalan et al., 2006; Sunkar et al., 2008; Lu et al., 2008; Jagadeeswaran et al., 2009). Consistent with these reports, deep sequencing of small RNA populations in sorghum also revelaed 14 novel miRNAs; 9 were based on cloning of the miRNA* (s412459, s373158, s252721, s13121, s229270, s438157, s2122, s71509 and s76707); small RNA blot analysis revealed the detection of the miRNA* reads for 4 novel miRNAs (s91586, s449185, s197538 and s50685). Of these seven novel miRNAs are conserved (s412459, s449185, s373158, s50685, s418541, s13121 and s71509) in related monocot species such as sugarcane, wheat, maize and switchgrass. Thus, of the 14 novel miRNAs, 7 (s412459, s449185, s373158, s50685, s418541, s13121 and s71509) are 'conserved miRNAs' and the remaining 7 are sorghum-specific (s252721, s91586, s197538, s229270, s438157, s2122 and s76707) miRNAs.

V. Prediction of miRNA targets and their validation.

A total of 100 genes were predicted as targets for some of the conserved miRNAs (Table 4). The predicted targets include squamosa promoter binding proteins, MYB transcription factors, auxin response factors, NAM transcription factors; HD-ZIP transcription factors, Nuclear transcription factor Y subunit, SCARECREW (GRAS domain containing) transcription factors; Apetala-2 transcription factors; growth regulating transcription factors and MADS box transcription factors. In addition to these, genes encoding Argonaute proteins, F-box containg proteins, sulfate and phosphate transporters and an ATP sulfurylase, laccases, Cu/Zn superoxide dismutases, plantacyanain-like proteins were predicted as targets for the conserved miRNAs. Four of the

predicted target genes (ATP sulfurylase targeted by miR395, laccase (multicopper oxidase) targeted by mir397, Cu/Zn superoxide dismutase targeted by miR398 and phosphate transporter targeted by miR399) were confirmed as result of respective miRNA-directed cleavages (Figure 9). We also predicted 25 genes as potential targets for the 14 novel miRNAs (Table 5). Most of the predicted novel miRNA targets belong to hypothetical proteins implying that the predicted targets are novel genes and may have specific functions in Sorghum. However, some known genes such as E3 ubiquitin protein ligase (s71509), SNARE protein Syntaxin (s76707), SAM decarboxylase (s2122), putative receptor kinases (s418541) and oxidoredoreductase and arabinogalactan protein (s449185) are also predicted targets for the novel miRNAs in Sorghum.These results has laid the foundation for probing post-transcriptional gene regulations controlling growth and development as well as other vital processes including nutrient homeostasis in Sorghum, an important biofuel plant species.

REFERENCES

Abdel-Ghany, S.E., and Pilon, M. (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. J Biol Chem. 283(23):15932-45.

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol 18, 758-762.

Ambros, V. (2004). The functions of animal microRNAs. Nature 431, 350-355.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D.C. (2005). Cloning and characterization of micro-RNAs from moss. Plant J 43, 837-848.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell 15, 2730-2741.

Axtell, M.J., and Bartel, D.P. (2005). Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658-1673.

Axtell, M.J., Snyder, J.A., and Bartel, D.P. (2007). Common functions for diverse small RNAs of land plants. Plant Cell 19, 1750-1769.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M., and Frugier, F. (2008). MicroRNA166 controls root and nodule development in Medicago truncatula. Plant J 54, 876-887.

Brodersen, P., and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10, 141-148.

Buhtz A, Springer F, Chappell L, Baulcombe DC, Kehr J (2008) Identification and characterization of small RNAs from the phloem of Brassica napus. Plant J 53: 739–749

Chekulaeva, M., and Filipowicz, W. (2009). Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Curr Opin Cell Biol 21, 452-460.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303, 2022-2025.

Chiou, T.J., Aung, K., Lin, S.I., Wu, C.C., Chiang, S.F., and Su, C.L. (2006). Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. Plant Cell 18, 412-421.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science 301, 336-338.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M., and Niebel, A. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes Dev 20, 3084-3088.

Ding, Y.F., and Zhu, C. (2009). The role of microRNAs in copper and cadmium homeostasis. Biochem Biophys Res Commun 386, 6-10.

Dong, Z., Han, M.H., and Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A 105, 9970-9975.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2, e219.

Fujii, H., Chiou, T.J., Lin, S.I., Aung, K., and Zhu, J.K. (2005). A miRNA involved in phosphate-starvation response in Arabidopsis. Curr Biol 15, 2038-2043.

Gregory, R.I., and Shiekhattar, R. (2005). MicroRNA biogenesis and cancer. Cancer Res 65, 3509-3512.

Han, M.H., Goud, S., Song, L., and Fedoroff, N. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci U S A 101, 1093-1098.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol 57, 19-53.

Jagadeeswaran, G., Zheng, Y., Li, Y.F., Shukla, L.I., Matts, J., Hoyt, P., Macmil, S.L., Wiley, G.B., Roe, B.A., Zhang, W., and Sunkar, R. (2009). Cloning and characterization of small RNAs from Medicago truncatula reveals four novel legume-specific microRNA families. New Phytol 184, 85-98.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A.,

and Carrington, J.C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5, e57.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209-216.

Kidner, C.A., and Martienssen, R.A. (2005). The developmental role of microRNA in plants. Curr Opin Plant Biol 8, 38-44.

Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J., and Chua, N.H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. Plant Cell Physiol 49, 1634-1644.

Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J., and Chua, N.H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. Plant Cell Physiol 49, 1634-1644.

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A 101, 12753-12758.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., Lin, C., Socci, N.D., Hermida, L., Fulci, V., Chiaretti, S., Foa, R., Schliwka, J., Fuchs, U., Novosel, A., Muller, R.U., Schermer, B., Bissels, U., Inman, J., Phan, Q., Chien, M., Weir, D.B., Choksi, R., De Vita, G., Frezzetti, D., Trompeter, H.I., Hornung, V., Teng, G., Hartmann, G., Palkovits, M., Di Lauro, R., Wernet, P., Macino, G., Rogler, C.E., Nagle, J.W., Ju, J., Papavasiliou, F.N., Benzing, T., Lichter, P., Tam, W., Brownstein, M.J., Bosio, A., Borkhardt, A., Russo, J.J., Sander, C., Zavolan, M., and Tuschl, T. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401-1414.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G., and Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proc Natl Acad Sci U S A 105, 8795-8800.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858-862.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell 115, 787-798.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol 15, 1501-1507.

Li, Y.F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M.J., Zhang, W., and Sunkar, R. (2010). Transcriptome-wide identification of microRNA targets in rice. Plant J. 62, 742-759.

Liu, B., Li, P., Li, X., Liu, C., Cao, S., Chu, C., and Cao, X. (2005). Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. Plant Physiol 139, 296-305.

Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C., and Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. EMBO Rep 7, 1052-1058.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science 294, 853-858.

Lu, C., Jeong, D.H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B.C., and Green, P.J. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci U S A 105, 4951-4956.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006). MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res 16, 1276-1288.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat Genet 38 Suppl, S31-36.

Matts, J., Jagadeeswaran, G., Roe, B.A., and Sunkar, R. (2010) Identification of microRNAs and their targets in switchgrass, a model biofuel plant species. J Plant Physiol. 167, 896-904.

Matzke, M., Matzke, A.J., and Kooter, J.M. (2001). RNA: guiding gene silencing. Science 293, 1080-1083.

Meyers, B.C., Souret, F.F., Lu, C., and Green, P.J. (2006). Sweating the small stuff: microRNA discovery in plants. Curr Opin Biotechnol 17, 139-146.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., and Zhu, J.K. (2008). Criteria for annotation of plant MicroRNAs. Plant Cell 20, 3186-3190.

Millar, A.A., and Waterhouse, P.M. (2005). Plant and animal microRNAs: similarities and differences. Funct Integr Genomics 5, 129-135.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C. and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature, 447, 1126–1129.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell 18, 2929-2945.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453, 803-806.

Ori, N., Cohen, A.R., Etzioni, A., Brand, A., Yanai, O., Shleizer, S., Menda, N., Amsellem, Z., Efroni, I., Pekker, I., Alvarez, J.P., Blum, E., Zamir, D., and Eshed, Y. (2007). Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nat Genet 39, 787-791.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102, 3691-3696.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F., Messing, J., and Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551-556.

Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science 321, 1490-1492.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20, 3407-3425.

Reddy, A.M., Zheng, Y., Jagadeeswaran, G., Macmil, S.L., Graham, W.B., Roe, B.A., Desilva, U., Zhang, W., and Sunkar, R. (2009). Cloning, characterization and expression analysis of porcine microRNAs. BMC Genomics 10, 65.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901-906.

Schwab, R., and Voinnet, O. (2009). miRNA processing turned upside down. EMBO J 28, 3633-3634.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev Cell 8, 517-527.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199-208.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16, 2001-2019.

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12, 301-309.

Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. (2005). Cloning and characterization of microRNAs from rice. Plant Cell 17, 1397-1411.

Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.R., Mueller-Roeber, B., Jarmolowski, A., and Szweykowska-Kulinska, Z. (2009). Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. Nucleic Acids Res 37, 3083-3093.

Talmor-Neiman, M., Stav, R., Klipcan, L., Buxdorf, K., Baulcombe, D.C., and Arazi, T. (2006). Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J 48, 511-521.

Trindade, I., Capitao, C., Dalmay, T., Fevereiro, M.P., and Santos, D.M. miR398 and miR408 are up-regulated in response to water deficit in Medicago truncatula. Planta 231, 705-716.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol 14, 346-351.

Wang, M.L., Zhu, C., Barkley, N.A., Chen, Z., Erpelding, J.E., Murray, S.C., Tuinstra, M.R., Tesso, T., Pederson, G.A., and Yu, J. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor Appl Genet 120, 13-23.

Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C, Qi Y. (2010). DNA methylation mediated by a microRNA pathway. Mol Cell. 38(3):465-75.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. (2005). Expression of Arabidopsis *MIRNA* genes. Plant Physiol 138, 2145-2154.

Yamasaki, H., Abdel-Ghany, S.E., Cohu, C.M., Kobayashi, Y., Shikanai, T., and Pilon, M. (2007). Regulation of copper homeostasis by micro-RNA in Arabidopsis. J Biol Chem 282, 16369-16378.

Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. Plant J 47, 841-850.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science 307, 932-935.

Zhong, R., and Ye, Z.H. (2007). Regulation of HD-ZIP III Genes by MicroRNA 165. Plant Signal Behav 2, 351-353.

Zhu, J.K. (2008). Reconstituting plant miRNA biogenesis. Proc Natl Acad Sci U S A 105, 9851-9852.

Zhu, Q.H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., and Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18, 1456-1465.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406-15.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol 14, 1214-1220.

APPPENDICES

s412459 is conserved in Sugarcane

```
cuc   ga  u-      c      -       u             -  -     ca----      ---  c
   ccg  gc  ggccug cguggg gaagcaa ucgucgaacagcu gc ugcgc      gagguu   gg a
   ggc  cg  ccggac guaccc cuucguu agcggcuugucga cg acgug      cuccga   cc g
u--   uc  uc      u      a       c             g  c     caccug      guc  g
```

s412459 in Sugarcane

```
gcuc-   ga  u-      c      -       u     g        c  c--   agc   uu    c
     ccg  gc  ggccug cguggg gaagcaa ucguc aacggcug ag   gcg   uug  aggc u
     ggc  cg  ccggac guaccc cuucguu agcgg uugucgac uc   cgc   agc  uccg g
guauu   uc  uc      c      g       c     -        u  aca   acu   --    a
```

s412459 in Sorghum (2 nt)

```
agcu      ucc    --                    -  a-     a----    g   ua
    ggccug   uggg  aagcaauucgucgaacagcu gc  gcgcg     gagg ugg  g
    ccggac   accc  uucguuaagcgguuugucga cg  cgcgc     cucc acu  g
----      ugu    au                    g  ca     acuag    g   cc
```

s412459 in Sorghum (1 nt)

```
gcu      ucc    --                   -  a-    gagacg     --    cc
    ggccug   uggg  aagcaauucgucgaacagcu gc  gcgc      guggu  aggc  \
    ccggac   accc  uucguuaagcgguuugucga cg  cgcg      cacca  uccg  u
 uc-      ugu    au                   g  ca    ------     gc    ac
```

s449185 is conserved in Maize and Wheat

```
a---------               u  c      acc            u    a  ag   a          ag
          agguuuguuuggaugu gu ggauuc   ucaauccacaugu uugg gu  auu ggguggaauuu  \
          uccggaacaaaccuaca ca ccuaag   aguuagguguaca aacc ca  uaa cucaccuugaa  u
 ugaauauaaa               u  a      cgu            u    c  cu   c          cu
```

s449185 in maize

```
a---      g  ua     u           c     cua    --   c   -  a       -   aauuu   u
    ggccuu uu  gaugu gucggauucac uuaau   caua  ugu gga ug auugggg uag     agu c
    ccggaa aa  uuaca cagucuaagug aguua   guau  aca ccu ac uaacuuc guc     ucg a
aaaa      g  gc     u           u     a--    ag   a   c  c       u   acu--   a
```

s449185 in wheat (1 nt)

```
                  u  c      auc                  ag   a   ag         ag
aggcuuuguuuggaugu gu ggauuc   ucaauccacauguauugg  ugg uug  guggaauuu  \
uccggaacaaaccuaca ca ucuaag   aguuagguguacauaacu  acc aau  caccuugaa  u
                  u  c      cau                  cg   c   cu         cu
```

s449185 in Sorghum (1 nt)

```
---------                 u        acc                 ag   a       a     ag
         aaggccuuguuuggaugu guuggauuc   ucaauccauauguguugg  ugg uuggggug aauuu  \
         uuccggaacaaaccuaca cagccuaag   aguuagguguauacaauc  acc aaccucac uugaa  u
 cacauuugu                 u         cgu                 cu   c       c     cu
```

s418541 is conserved in Wheat

```
.-uuu      a           a                     uggacugaagugg
     ggaugu gucggauuuau ucaauccacauauguuggg              \
     ccuaca cagccuaagug aguuagguguauacaaccu              a
\ ---      a           g                     uaaaucaaguuua
```

s418541 in wheat (1 nt)

```
-   uc           a  c       a         a      -        a      c   a
 agg  uuguuuggaugu gu ggauucgc ucaauccaca guguugg guggauugg guggaa uua \
 ucc  aauaaaccuaca ua ccuaagug aguuaggugu cacaacc caccuaauc cacuuu aau a
g   ga           a  a       g         a      u        c      a   c
```

s373158 is conserved in Switchgrass

```
-         auc                  a       agc          ca     u    uc
 gcaaaggc    auugagugcagcguug ugagcc    uggccggc   gccg gcg   \
 cguuuccg    uaacucacgucgcggc acucgg    accggccg   cggc cgc   c
u          cau                  c       c--          --     -    cg
```

s373158 in switchgrass (1 nt)

```
         auc                  a         c    -     ag
gcgaaggc    auugagugcagcguug ugagccgc ggc ggcg   \
cguuuccg    uaacucacgucgcggc acucggcg ccg ccgc   g
         caa                  c         -    g     cu
```

s13121 is conserved in Sugarcane

```
    -    ucccaaca-                                  a   g  ugcu   gc
gcua guac         uuccaaauuguaagucguucuggcuuuucua gua au    gug  u
cgau uaug         aagguuuaacauucagcgagaccgagaagau uau ua    cac  a
    a    uuauauaac                                  g   g  uc--   gu
```

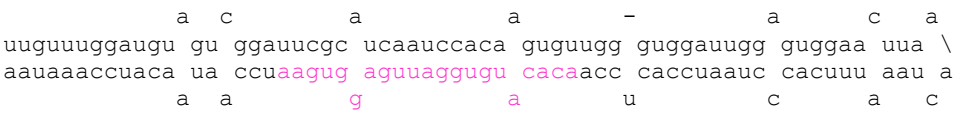
s13121 in Sugarcane

```
ccucc                   u      uu     a        aaaa--       au
     guuccaaauuguaaguugu cuggcu  ucuag uacauagu      guuaugu  c
     caaggguuuaacguucagca gaucga  agauc auguaucg      cgauaca  u
a----                   u      na     c        aagaua       ga
```

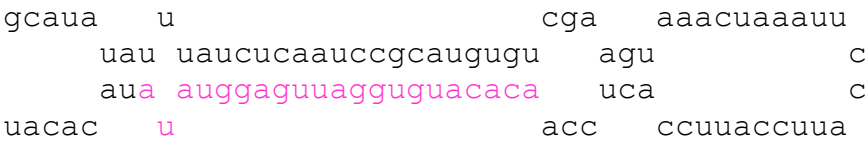
s50685 is conserved in Wheat and Sugarcane

```
ag           a                 uggu          a    aaacuuga
  ucggauucgc ucaauccacaugugu     guggauugg gug         \
  agccuaagug aguuagguguacaca     caccuaacc cac          a
ag           g                 u---          c    cuuaaauu
```

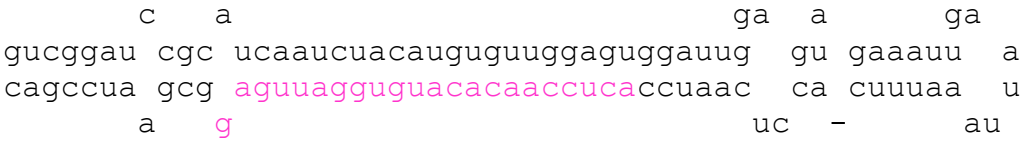
s50685 in wheat



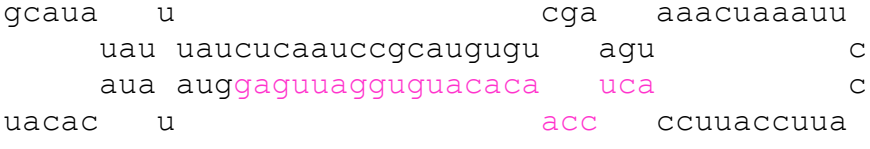
s50685 in sugarcane (2 nt)



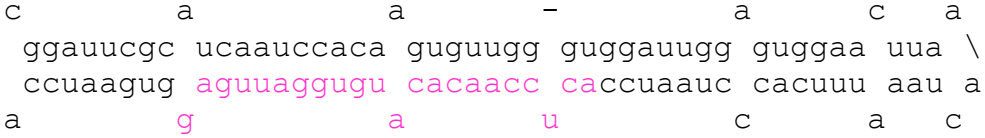
s71509 is conserved in Sugarcane and Wheat



s71509 in sugarcane



s71509 in wheat



Appendix 1. Predicted fold-back structures for the novel miRNA precursors that are conserved in related monocots. The sequences colored in red and blue are the mature miRNA sequence and miRNA* sequences, respectively.

miR156

```
jgi|Sorbi1|5232347   5′ gugcucucucucuucuguca 3′
                        ||||||0||||||||||||
miRNA156             3′ cacgagugagagaagacagu 5′

jgi|Sorbi1|5271494   5′ gugcucucucucuucuguca 3′
                        ||||||0||||||||||||
miRNA156             3′ cacgagugagagaagacagu 5′

jgi|Sorbi1|4977676   5′ gugcucucucucuucuguca 3′
                        ||||||0||||||||||||
miRNA156             3′ cacgagugagagaagacagu 5′

jgi|Sorbi1|4201067   5′ gugcucucucucuucuguca 3′
                        ||||||0||||||||||||
miRNA156             3′ cacgagugagagaagacagu 5′
```

miR159

```
jgi|Sorbi1|4982901   5′ uggagcucccuucacuccaag 3′
                        ::||||||||||||||||||:
miRNA159             3′ gucucgagggaaguuagguuu 5′
```

miR160

```
jgi|Sorbi1|5291326   5′ aggcauacagggagccaggca 3′
                        0||||||||||||||||||||
miRNA160             3′ accguaugucccucgguccgu 5′

jgi|Sorbi1|5277928   5′ aggcauacagggagccaggca 3′
                        0||||||||||||||||||||
miRNA160             3′ accguaugucccucgguccgu 5′
```

miR164

```
jgi|Sorbi1|5121666   5′ agcucgugcccugcuucucca 3′
                        0||0||||||||||||||||
miRNA164             3′ acgugcacgggacgaagaggu 5′
```

miR166

```
jgi|Sorbi1|5230460   5′ cgggaugaagccugguccgg 3′
                        0||:|||||||||||||||:
miRNA166             3′ cccuuacuucggaccaggcu 5′
jgi|Sorbi1|5289000   5′ ugggaugaagccugguccgg 3′
                        0||:|||||||||||||||:
miRNA166             3′ cccuuacuucggaccaggcu 5′

jgi|Sorbi1|5102053   5′ ugggaugaagccugguccgg 3′
                        0||:|||||||||||||||:
miRNA166             3′ cccuuacuucggaccaggcu 5′

jgi|Sorbi1|5098418   5′ ugggaugaagccugguccgg 3′
                        0||:|||||||||||||||:
miRNA166             3′ cccuuacuucggaccaggcu 5′

jgi|Sorbi1|5097817   5′ ugggaugaagccugguccgg 3′
                        0||:|||||||||||||||:
miRNA166             3′ cccuuacuucggaccaggcu 5′
```

miR167

```
jgi|Sorbi1|5061512   5′ uagaucaggcuggcagcuugu 3′
                        |||||||0||||||||||00
miRNA167             3′ aucuaguacgaccgucgaagu 5′
```

miR168

```
jgi|Sorbi1|4995896   5′ uucccgagcugcaccaagccc 3′
                        0||||||||||||||||||00
miRNA168             3′ cagggcuagacgugguucgcu 5′
```

miR169

```
jgi|Sorbi1|5101414   5′ cuggcaacucauccuuggcuu 3′
                        0:|||||0||||||||||||0
miRNA169             3′ agccguucaguaggaaccgac 5′
```

miR171

```
jgi|Sorbi1|5235905   5′ gauauuggcgcggcucaauca 3′
                        |||||||||:||||||||||
miRNA171             3′ cuauaaccgugccgaguuagu 5′

jgi|Sorbi1|5233945   5′ gauauuggcgcggcucaauca 3′
                        |||||||||:||||||||||
miRNA171             3′ cuauaaccgugccgaguuagu 5′

jgi|Sorbi1|4987016   5′ gauauuggcgcggcucaauca 3′
                        |||||||||:||||||||||
miRNA171             3′ cuauaaccgugccgaguuagu 5′
```

miR172

```
jgi|Sorbi1|5287190   5′ ugcagcaucaucacgauucc 3′
                        ||||||||||||0|||||0
miRNA172             3′ acgucguaguaguucuaaga 5′
```

miR319

```
jgi|Sorbi1|5228446   5′ aggggacccuucaguccaa 3′
                        0||:|0||||||||||||||
miRNA319             3′ cccucgugggaagucagguu 5′

jgi|Sorbi1|5257392   5′ aggggacccuucaguccaa 3′
                        0||:|0||||||||||||||
miRNA319             3′ cccucgugggaagucagguu 5′
```

miR390

```
jgi|Sorbi1|127730    5′ gguuc-auuccuccugaucuu 3′
                        ||:0|0||:||||||||0|||
miRNA390             3′ ccgcgauagggaggacucgaa 5′
```

miR393

```
jgi|Sorbi1|5082622   5′ agacaaugcgaucccuuugga 3′
                        0:0||||||||||||||||||
miRNA393             3′ cuaguuacgcuagggaaaccu 5′
```

miR394

```
jgi|Sorbi1|4983693   5′ ggagguggacagaaugccaa 3′
                        ||||||||||||||||||||
miRNA394             3′ ccuccaccugucuuacgguu 5′

jgi|Sorbi1|5234325   5′ ggagguggacagaaugaagu 3′
                        ||||||||||||||||00:0
miRNA394             3′ ccuccaccugucuuacgguu 5′
```

miR395

```
jgi|Sorbi1|147166   5′ gaguuccuccaagcacuucau 3′
                       |||||||:||||:|||||||:
miRNA395            3′ cucaaggggguuugugaagug 5′
```

miR397

```
jgi|Sorbi1|123904   5′ gcucaucaacgccgcgcucaa 3′
                       ||||||||||||0||:|||||
miRNA397            3′ cgaguaguugcgacgugaguu 5′

jgi|Sorbi1|146423   5′ gcucaucaacgccgcacucaa 3′
                       ||||||||||||0|||||||
miRNA397            3′ cgaguaguugcgacgugaguu 5′

jgi|Sorbi1|5109791   5′ caucaucaacgcugcgcucaa 3′
                        |0||||||||||||:|||||
miRNA397             3′ cgaguaguugcgacgugaguu 5′
```

miR398

```
jgi|Sorbi1|131357   5′ cgggggccgccugagaucaca 3′
                       |||||||0:|||||||0||||
miRNA398            3′ gccccccgcuggacucuugugu 5′
```

miR399

```
jgi|Sorbi1|142168   5′ ccgggcagcucuucuucggcu 3′
                       |0|||||||||:|||0|||0
miRNA399            3′ gucccgucgagaggaaaccgu 5′
```

miR408

```
jgi|Sorbi1|5257619   5′ cucagggaagaggcggugcaa 3′
                        0:|||||||||||:|||||0
miRNA408             3′ cggucccuucuccgucacguc 5′
```

miR444

```
jgi|Sorbi1|144632   5′ aagcuugaggcaacaacugca 3′
                       ||||||||||||0||||||||
miRNA444            3′ uucgaacuccgucguugacgu 5′

jgi|Sorbi1|147544   5′ aggc-ugaaggagcaacugca 3′
                       |:||0|||0|0|||||||||
miRNA444            3′ uucgaacuccgucguugacgu 5′
```

Appendix 2. Alignments of the predicted targets and the conserved miRNAs. The top sequences are the predicted mRNA targets in the 5'-3' direction. The bottom sequences are their corresponding miRNAs in the 3'-5' direction. Matches, mismatches and G-U wobbles are indicated with straight line, circle and colons.

s412459

```
jgi|Sorbi1|5108736   5′ uguucgacgaauugcuuccgc3′
                        ||||||||||||||||||00
s412459              3′ acaagcugcuuaacgaagggu 5′

jgi|Sorbi1|5021956   5′ uguucgacgaauugcuucacc 3′
                        ||||||||||||||||||0|0
s412459              3′ acaagcugcuuaacgaagggu 5′

jgi|Sorbi1|4975996   5′ uguucgacgaauugcuucagc 3′
                        ||||||||||||||||||000
s412459              3′ acaagcugcuuaacgaagggu 5′

jgi|Sorbi1|5268653   5′ uguucgacgaauugcuucacc 3′
                        ||||||||||||||||||||0
s412459              3′ acaagcugcuuaacgaagggu 5′
```

```
jgi|Sorbi1|5221943   5′ uguucgaugaauugcuuccuc 3′
                        ||||||||||||||||||:0
s412459              3′ acaagcugcuuaacgaagggu 5′
```

s91586

```
jgi|Sorbi1|5108736   5′ ucaagcuguucgacgaauugcu 3′
                        0||||||||||||||||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5260497   5′ ucaagcuguucgacgaauugcu 3′
                        0||||||||||||||||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5268653   5′ ucaagcuguucgacgaauugcu 3′
                        0||||||||||||||||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5048590   5′ ucaagcuguucgacgaguugcu 3′
                        0|||||||||||||||:|||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5234440   5′ ucaagcuguuugacgaauugcu 3′
                        0|||||||||:||||||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5268646   5′ ucaagcuguucgacggauugcu 3′
                        0|||||||||||||:||||||
S91586               3′ uguucgacaagcugcuuaacga 5′


jgi|Sorbi1|5268643   5′ ucaagcuguucgacggauugcu 3′
                        0|||||||||||||:||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5268631   5′ ucaagcuguucgaugaauugcu 3′
                        0||||||||||||:|||||||
S91586               3′ uguucgacaagcugcuuaacga 5′

jgi|Sorbi1|5215956   5′ ugaagcuguucgacgaaaugcu 3′
                        0|||||||||||||||0||||
S91586               3′ uguucgacaagcugcuuaacga 5′
```

s449185

```
jgi|Sorbi1|5239195   5′ gugaaucugacaacauccaaacaa 3′
                        |||||||:|||||||||||||||
S449185              3′ cacuuaggcuguuguagguuuguu 5′

jgi|Sorbi1|5278077   5′ gugaaucugacaacauccaaacaa 3′
                        |||||||:|||||||||||||||
S449185              3′ cacuuaggcuguuguagguuuguu 5′

jgi|Sorbi1|5277399   5′ auaaauccgacaacauccaaauaa 3′
                        0|0||||||||||||||||:||
S449185              3′ cacuuaggcuguuguagguuuguu 5′

jgi|Sorbi1|5290360   5′ gugaauccgacuauauccaaacaa 3′
                        ||||||||||0|:||||||||||
S449185              3′ cacuuaggcuguuguagguuuguu 5′
```

s418541

```
jgi|Sorbi1|5278077   5′ uuggauucaccucaauccaua 3′
                        ||||||||||||||||||:|
S418541              3′ agccuaaguggaguuaggugu 5′
```

s252721

```
jgi|Sorbi1|124111    5′ agc-gcaucuccccgcugucc 3′
                        0||0|||:||||||||||0|
s252721              3′ ccgucguggaggggcgacaag 5′
```

s197538

```
jgi|Sorbi1|149258   5′ augcauacgucuaaagauucgaug 3′
                       |||||||||||||||||||||||
s197538             3′ uacguaugcagauuucuaagcuac 5′
```

s13121

```
jgi|Sorbi1|5046708   5′ agccagaacgacuuacaauuu 3′
                        ||||||||||||||||||||
s13121               3′ ucggucuugcugaauguuaaa 5′
```

s438157

```
jgi|Sorbi1|147278   5′ gacguuuuu-ccaugguugaa 3′
                       ||0:|||||0||||:|||0||
s438157             3′ cuuuaaaaaugguaucaaguu 5′
```

s71509

```
jgi|Sorbi1|4982924   5' aucaauccacauguauuagagu 3'
                        0|||||||||||||0||0||||
s71509               3' gaguuagguguacacaaccuca 5'
```

s76707

```
jgi|Sorbi1|147110   5′ acaauaacacauauggauuggagu 3′
                       |||0|||||0|:|||||||:|||
S76707              3′ ugugguuguauguaccuaacuuca 5′
```

Appendix 3. Alignments of the predicted targets and the novel miRNAs. The top sequences are the predicted mRNA targets in the 5’-3’ direction. The bottom sequences are their corresponding miRNAs in the 3’-5’ direction. Matches, mismatches and G-U wobbles are indicated with straight line, circle and colons.

VITA

LI ZHANG

Candidate for the Degree of

Master of Science

Thesis: CLONING AND CHARACTERIZATION OF MICRORNAS IN SORGHUM

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Master of Science in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in July, 2010.

Completed the requirements for the Bachelor of Science in Biotechnology at Sichuan University, Chengdu, Sichuan, China in 2004

Completed the requirements for the Master of Science in Microbiology at Sichuan University, Chengdu, Sichuan, China in 2007.

Experience:
10/2005-06/2006: Research assistant. I participated in the preliminary study in antibacterial-related proteins from Silurus meriaionalis

Professional Memberships: None

Name: LI ZHANG Date of Degree: July, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: Cloning and Characterization of microRNAs from Sorghum

Pages in Study: 62 Candidate for the Degree of Master of Science

Major Field: Biochemistry and Molecular Biology

Scope and Method of Study:

In plants, gene regulation guided by the microRNAs (miRNAs) plays a key role in normal growth and development, nutrient homeostasis and stress tolerance. MiRNAs post-transcriptionally regulate gene expression either by causing degradation or attenuating the expression of RNA targets. Thus, identification of miRNAs is as important as protein-coding genes. Sweet sorghum, a drought tolerant crop, is largely grown for grain production in northeast Africa and for fodder production in southern plains of the United States. Presently, sorghum has emerged as one of the model plants for biofuel production. To identify miRNAs expressed in Sorghum, we generated a small RNA library.

Findings and Conclusions:

Sequence analysis revealed the expression of 28 conserved miRNA families in Sorghum. Additionally, 14 novel miRNAs were found of which seven are conserved at least in one another monocot. Expression analysis indicated the diffrential expression of several conserved miRNAs in different tissues. Furthermore, small RNA blot analysis indicated that miRNAs such as miR395 and miR399 are induced under sulfate-, and phosphate-deprived conditions, respectively. We also predicted more than 100 potential targets for miRNAs and some of them were validated using modified 5'-rapid amplification of cDNA ends (RACE) assay. These findings suggest that a large number of conserved and novel miRNAs are encoded in the Sorghum genome.

ADVISER'S APPROVAL: Dr. Ramanjulu Sunkar