Genomics xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identification and temporal expression analysis of conserved and novel microRNAs in Sorghum

Li Zhang [a,1], Yun Zheng [b,1], Guru Jagadeeswaran [a], Yongfang Li [a], Kanchana Gowdu [a], Ramanjulu Sunkar [a,*]

[a] Department of Biochemistry and Molecular Biology Oklahoma State University, Stillwater, OK 74078 USA
[b] Institute of Developmental Biology and Molecular Medicine and School of Life Sciences, Fudan University, Shanghai, 200433, China

ARTICLE INFO

Article history:
Received 28 March 2011
Accepted 17 August 2011
Available online xxxx

Keywords:
miRNAs
miRNA targets
Post-transcriptional gene regulation
Sorghum

ABSTRACT

Sweet Sorghum is largely grown for grain production but also recently emerged as one of the model feedstock plants for biofuel production. In plants, microRNA (miRNA)-guided gene regulation plays a key role in diverse biological processes, thus, their identification in different plant species is essential to understand post-transcriptional gene regulation. To identify miRNAs in Sorghum, we sequenced a small RNA library. Sequence analysis revealed the identity of 29 conserved miRNA families. Importantly, 13 novel miRNAs are identified, seven of which are conserved in closely related monocots. Temporal expression analysis of conserved and novel miRNAs indicated differential expression of several miRNAs. Approximately 125 genes that play diverse roles have been predicted as targets and a few targets were experimentally validated. These results provided insights into miRNA-controlled processes in Sorghum and also laid the foundation for manipulating miRNAs or their targets for improving biomass production and stress tolerance in Sorghum.

1. Introduction

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 [1–3]. Recently identified microRNAs (miRNAs) act as regulators of gene expression at the post-transcriptional level in higher eukaryotes [1,2,4]. In plants, miRNA genes are transcribed, and the resulting primary miRNA transcripts can adopt a hairpin-like structure, which is processed by the enzyme Dicer-like 1. Several other proteins such as hyponastic leaves 1 (a double-stranded RNA [dsRNA] binding protein), serrate (a C2H2 zinc-finger protein), and Dawdle (DDL), help the DCL-1 enzyme process the mature miRNA duplex (miRNA and miRNA*) from the hairpin-like structure. A methyl group is added to the 3′ ends of the miRNA duplex by the HEN1, a methyltransferase to stabilize the duplex [5]. The methylated miRNA:miRNA* duplex is then exported to the cytoplasm by HASTY5 (a plant ortholog of exportin 5) [6], where the miRNA is loaded into the RNA-induced Silencing Complex containing Argonaute protein. In plants, miRNAs show near-perfect complementarity to their target mRNAs. The interaction between miRNA and its mRNA target leads to the degradation of target mRNA and/or prevents the target mRNA translation [1,7].

Sorghum (*Sorghum bicolor* L. Moench), an African grass related to sugarcane and maize, is grown for food, feed, fiber and fuel [8]. As a C4 plant species, Sorghum has high photosynthetic efficiency for converting solar energy to biomass and high water use efficiency for growing in high-temperature and drought-prone areas and can be grown on poor and marginal lands [9,10]. 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 [11]. Sorghum is not used widely as food grain in the United States but recently attracted much attention as a versatile feedstock for large-scale bioenergy production. Despite its importance both as a grain producer and as an emerging biofuel crop species, we know little about the gene-regulatory processes controlling the plant architecture, biomass accumulation, nutrient uptake and assimilation and stress responses in Sorghum. An understanding of these processes could assist in designing biotechnological strategies aimed at improving grain or biomass production.

To date, small RNAs have been identified in rice, Arabidopsis, rice, *Populus*, *Physcomitrella*, *Medicago truncatula*, switchgrass, tomato and several other plant species [12–23]. These studies indicated the expression of lineage-specific and species-specific miRNAs besides the

Abbreviations: AGO-1, Argonaute-1; ARF, auxin response factor; AP2-like, apetala 2-like transcription factor; CBP, cap-binding protein; DCL1, Dicer like-1; HEN1, Hua Enhancer-1; Hyl1, Hyponastic leaves-1; miRNAs, microRNAs; RACE, rapid amplification of cDNA ends; SCL, scarecrow-like transcription factors; SPL, Squamosa promoter binding protein-like; tasiRNAs, trans-acting small interfering RNAs; TCP factors, Teosinte branched 1, Cycloidea, PCF (TCP)-domain protein family; TPM, transcripts per million.

* Corresponding author. Fax: +1 405 744 7799.
E-mail address: ramanjulu.sunkar@okstate.edu (R. Sunkar).
[1] Equally contributed to this work.

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.08.005

~21 well-conserved miRNA families. Thus far, some of the conserved miRNAs in Sorghum have been only predicted [8,24] and novel miRNAs have not been identified. Here, we report the identification of not only conserved miRNAs but also 13 novel miRNA families, 7 of which are conserved in closely related monocots. We also determined the temporal expression of conserved and novel miRNAs and predicted their mRNA targets, and validated a few targets in Sorghum.

2. Materials and methods

2.1. Construction of a small RNA library, sequencing and sequence analysis

We constructed a small RNA library from the 3-week-old sweet Sorghum (M81E) seedlings as previously described [15]. In brief, small RNAs of 18–28 nt were fractionated, isolated and ligated with the 5′ and 3′ RNA adapters. Then the ligated small RNAs with adapters were used for reverse transcription and subsequent PCR. The final PCR product was purified and sequenced by using 454-pyrosequencing technology.

Total RNA was isolated from the three-week-old Sorghum (M81E) seedlings using Trizol reagent. Small RNAs of the desired size range (18–24 nt) were gel-isolated (denaturing 15% polyacrylamide gel) and small RNA library was constructed as previously described [15]. In brief, 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 μl) of the final PCR product electrophoresed using 3% low-melting agarose gel along with a 25 bp DNA ladder. The final PCR product was purified and sequenced by using 454-pyrosequencing technology.

The adaptor sequences of the raw reads were trimmed, and the small RNAs in between the adaptors were extracted. The redundant sequences were eliminated, and unique small RNAs were counted. The unique small RNAs were aligned to RNAs in the databases Repbase (version 14.01, obtained from http://www.girinst.org) and the TIGR Sorghum Repeats DB (http://plantrepeats.plantbiology.msu.edu/downloads.html), known noncoding RNAs (e.g., rRNAs, tRNAs, snRNAs, snoRNAs obtained from http://www.sanger.ac.uk/Software/Rfam/ftp.shtml) and mRNAs of *S. bicolor* (http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, filtered models) obtained by NCBI BLASTN search. Small RNAs mapped to these non-coding RNAs were removed from the dataset. Then the small RNAs were mapped to the reported miRNAs in the miRBase (release 16, obtained from http://microrna.sanger.ac.uk/sequences/ftp.shtml). Small RNAs that were mapped to known miRNAs of *S. 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 *S. bicolor* (downloaded from http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, masked assembly) by use of BLASTN. Unique small RNAs with more than 10 genomic hits were removed from further analysis because these small RNAs might have been derived from repeat-rich loci. The flanking regions of the remaining genome-matched sequences were extracted, and the fold-back structures were predicted with use of the RNAfold program [25]. The small RNAs for which a fold-back structure could be predicted were considered potential new miRNAs. Most importantly, sequencing miRNA* is a prerequisite to name a new small RNA as a new miRNA in plants [26]. The miRNA* sequences were predicted on the basis of 2-nt overhangs at the 3′ end and then searched for the presence of such predicted miRNA* sequences in our small RNA library.

2.2. Nutrient-deprivation treatments

Sorghum seeds were germinated in wet paper towels and seedlings were transferred to a 96-well PCR plate (with holes) sitting on a container. These seedlings were grown in a controlled growth chamber (22 °C to 24 °C) with a 16-h photoperiod and 600 μmol m^{-2} s^{-1} light intensity. Seedlings were grown in the presence of control media for 2 weeks, and later transferred to media containing indicated sulfate levels (0.02, 0.2, or 2.0 mM) or without phosphate or without copper for 5 days. After 5 days of nutrient starvation, seedlings were separated into roots and shoots and frozen in liquid nitrogen, and RNA was extracted.

2.3. Collection of plant tissue and RNA isolation

We used diverse 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), which were collected and immediately froze them in liquid nitrogen for storage at −80 °C until use. Total RNA was isolated from different tissues with TRIzol (Invitrogen) and low-molecular weight RNA (LMW) fraction was isolated from the total RNA by precipitating high molecular weight RNAs with the use of PEG (5%) and NaCl (500 mM).

2.4. Small RNA blot analysis

LMW RNA (20 μg) was resolved on a 15% polyacrylamide gel containing 7 M urea in TBE buffer along with labeled 21–24 nt RNA markers. The size-fractionated small RNAs were then transferred to Hybond-N+ (Amersham) membranes, UV cross-linked and baked for 1 h at 80 °C. Blots were pre-hybridized for at least 1 h and hybridized at 38 °C with a labeled DNA oligonucleotide probe complementary to miRNA sequence. Blots were washed 3 times at 50 °C with washing buffer (2xSSC, 0.1% SDS) and autoradiographed using a phosphorimager. Membranes were stripped and re-probed with a labeled U6 (small nuclear RNA), which served as a loading control.

2.5. Bioinformatic prediction of miRNA targets and their validation

To predict potential targets for Sorghum miRNAs, the annotated Sorghum coding sequences were used for searching complementary sequences to the miRNAs (http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html, filtered models). To obtain the annotation for Sorghum genes, Sorghum coding sequences were aligned to the cDNA sequences of *Arabidopsis thaliana* (http://www.tair.org/, release 9) and rice (http://rice.plantbiology.msu.edu/data_download.shtml, release 6.1) by use of BLAST. The gene annotations of the best-matched Arabidopsis and rice cDNAs were treated as putative annotations for Sorghum genes. In predicting targets, we allowed a maximum of 3.5 mismatches (G:U pair is treated as half mismatch) between the miRNA and its target mRNA [27]. A modified RNA ligase-mediated rapid amplification of cDNA ends (5′RACE) assay was used to verify whether the predicted miRNA target was subjected to cleavage in vivo [14]. The final PCR products were gel-purified, cloned and sequenced.

3. Results and discussion

Sequencing small RNA populations has the potential to find novel miRNAs, besides identifying conserved miRNA homologs. We constructed and sequenced a small RNA library using RNA isolated from 3-week-old Sorghum seedlings. A total of 619,010 sequences of 18 to 26 nt were obtained after removing the adapter sequence (Table 1). The highest read abundance was found for 21- and 24-nt small RNAs, which is consistent with the size of the Dicer products

Please cite this article as: L. Zhang, et al., Genomics (2011), doi:10.1016/j.ygeno.2011.08.005

Table 1
Sequence analysis of Sorghum small RNA library.

Reads	Number of unique reads	Percent of unique reads	Number of total reads	Percent of total reads
Genome matching reads	359,936	77.45	492,803	79.61
Messenger RNA	58,397	12.56	109,087	17.62
miRBase (miRNAs and miRNA*s)	197	0.04	25,733	4.16
Rfam (tRNAs, rRNAs snRNAs and snoRNAs)	26,456	5.69	71,099	11.49
Repeats	24,574	5.29	55,954	9.04
Mitochondrion/chloroplast	10,638	2.29	18,342	2.96
Total	464,763		619,010	

(Fig. 1). 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 other peaks (Fig. 1), which is also consistent with several reports from plants [20–23]. From the total reads, we established a dataset of unique reads with their read counts (Table 1). This dataset was used to search for the degradation products from rRNA, tRNA, small nuclear RNA, small nucleolar RNAs and mRNAs, whose sequences were excluded from further analysis. The filtered reads were used to identify conserved miRNA homologs by mapping to miRBase. The remaining set of sequences with perfect matches to the Sorghum genome was used to identify novel miRNAs in Sorghum.

3.1. Conserved miRNAs in Sorghum

Approximately 21 miRNA families are highly conserved between dicotyledonous and monocotyledonous plants. Additionally, miR403 is conserved in dicotyledonous plants only, and miR437, miR444, miR528, miR1126, miR1318, and miR1436 are found only in monocots (miRBAse release version 16). A few other miRNAs such as miR529, miR530 and miR827 are conserved in some of the monocots as well as dicots, whereas miR158, miR391, miR2118, miR2119 and miR2199 are conserved among within specific dicot lineages [21,22]. Furthermore, several species-specific miRNAs have been also reported from Arabidopsis, rice, *M. truncatula*, *Physcomitrella* and Populus [12,17–20,22].

We identified 113 conserved miRNA homologs belonging to 31 distinct miRNA families (Table 2 and Supplemental Table 1) in Sorghum by processing the small RNA reads. With the exception of miR162 and miR399, we identified miRNA families that are broadly conserved between dicots and monocots (Table 2). miR399 is known to be abundantly expressed under phosphate-limiting conditions [28,29]. miR162 is highly conserved and targets DCL-1, which processes mature miRNAs

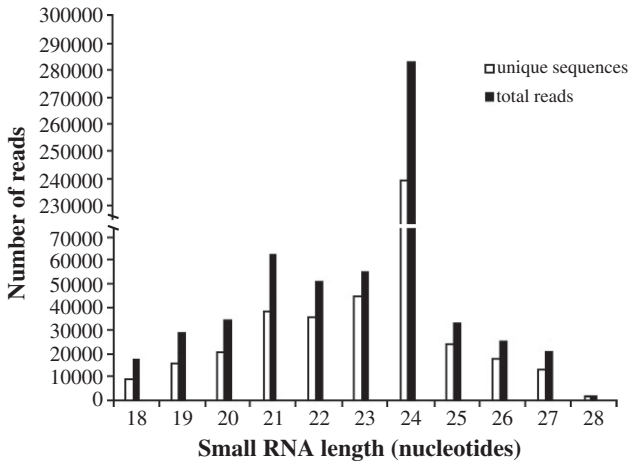


Fig. 1. Abundance of 18- to 27-nt small RNAs in the small RNA library.

Table 2
Conserved miRNA families, their frequency and predicted target gene families in Sorghum.

miRNA family	Normalized f requency(TPM)	Target gene family
miR156	87	SBP factors
miR159	1052	MYB factors
miR160	97	ARFs
miR164	27	NAC factors
miR165/166	3810	HD-ZIP factors
miR167	429	ARFs
miR168	305	Argonaute-1
miR169	352	NFY subunits
miR170/171	386	Scarecrow-like
miR172	32	AP2-like factors
miR319	10	TCP factors
miR390	2	TAS3 precursor
miR393	2	TIR1 homologs
miR394	31	F-Box proteins
miR395	48	APS and Sultr1
miR396	486	GRFs
miR397	23	Laccases
miR398	3	CSDs
miR408	23	Plantacyanin
miR437	2	
miR444	488	MADS-box factors
miR528	2	
miR529	6	
miR530	6	
miR827	55	
miR894	310	
miR1126	2	
miR1318	8	
miR1436	2	
miR2118	12	
miR2910	764	
tasiRNA3a	10	ARFs
tasiRNA3b	73	ARFs

from their precursors. Bioinformatics analysis suggested the conservation of miR162 in Sorghum [8], but the absence of miR162 sequence in the library implies that it is expressed at extremely low levels in seedlings. We have performed small RNA blot analysis using 20 μg low molecular weight RNA from different Sorghum tissues but unable to detect signal for miR162. Non-recovery of miR162 sequences in the small RNA library and inability to detect its expression using small RNA blot analysis (data not shown) in diverse tissues suggest that the expression abundance is extremely low in Sorghum. Consistent with these observations, miR162 is also expressed at extremely low levels in rice seedlings [15]. We also identified miR529, miR530, miR827, miR437, miR444, miR1126, miR1318 and miR1436 in Sorghum (Table 2), which have been reported only from monocot species. The fold-back structures for all these Sorghum miRNA precursors could be predicted using mfold [25] (Supplemental Fig. 1). Interestingly, miR894 homolog was found in Sorghum small RNA library (Table 2 and Supplemental Table 1). Thus far, miR894 has been reported only from *Physcomitrella patens*, a moss, and homologs were not found in Arabidopsis, rice, *Populus*, grapevine, *M. truncatula*, maize or other plant species for which high-throughput sequencing of small RNA populations have been performed (miRBAse release version 16).

The abundance of different miRNA families can be inferred from their frequency in the library. The read abundance for each of these miRNA families varied highly; the miR166 family is the most abundantly expressed, followed by miR159, miR444, miR396 and miR167. By contrast, the least-abundant expression was found for miR393, miR437, miR1126 and miR1436 (Supplemental Table 1). Similarly, miR390 was represented by least number of reads in the library. Interestingly, the frequency of TAS3-siRNAs is much greater than that of miR390 in Sorghum seedlings (Supplemental Table 1). TAS3-siRNA generation depends on miR390, because miR390-directed cleavage on

Please cite this article as: L. Zhang, et al., Genomics (2011), doi:10.1016/j.ygeno.2011.08.005

TAS3 precursors sets the stage for converting it into dsRNA and subsequently processing TAS3-siRNAs in a phased manner [22,30].

Most conserved miRNA families are represented by the same sequence or a slight variation in the nucleotide sequence at multiple loci in plants. Each of these loci appears to differ in expression, which could confer a tissue- or cell-specific expression of different members within a miRNA family. Therefore, the locus that is highly expressed must be assessed. The expression from each of these loci can be assessed by the frequency of their appearance in the library, provided that these members vary in at least 1 nt. On the basis of a slight variation in nucleotide sequence, miR169 and miR444 are the largest miRNA families and are represented by 14 and 12 members, respectively in Sorghum (Supplemental Table 1). The next largest are miR166 (11 members), miR160, miR167 and miR168 (9 members each), and the miR56, miR159 and miR167 families (8 members each); several families such as miR319, miR390, miR393, miR398, miR437, miR529 and miR530 are represented by a single member. 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, the miR166a-3 variant is the most abundantly expressed (1226 TPM; transcripts per million), whereas miR166f-2 is the least expressed (10 TPM) in seedlings; similarly, miR159a is represented by 331 TPM, and miR159c is represented by 3TPM; miR167C-2 by 129 TPM and miR167h by 2 TPM (Supplemental Table 1). Eight of the miR168 members showed almost a similar level of expression as represented by their frequency (Supplemental Table 1). Interestingly, eight miR168 variants have been identified in the small RNA library suggesting that miR168 family may be represented by eight loci in Sorghum. miR168 is represented by fewer loci (1 or 2) in Arabidopsis, rice and others; therefore, it appears that miR168 in Sorghum underwent not only additional duplications but also diverged in its sequence.

Some of the highly conserved miRNAs are induced when specific nutrients are deprived, which suggests that miRNAs play an important role in nutrient homeostasis [3,31]. miR395 is induced under sulfate deficiency and miR399 under phosphate deficiency in *Arabidopsis*, *Brassica* sp. and *M. truncatula* [22,32–34], although such regulation has not been observed in switchgrass [23]. Albeit at a low frequency, we found miR395 reads in our library generated from 3-week-old seedlings grown on hydroponic medium containing an optimal level of sulfate levels in the nutrient media (Table 2 and Supplemental Table 1). The recovery of miR395 reads from Sorghum seedlings suggests a basal expression of miR395 under normal conditions. However, we could not recover reads for miR399, which is induced under phosphate deficiency.

3.2. Expression analysis of conserved miRNAs

Temporal miRNA expression analysis in rice, Arabidopsis, *M. truncatula* and switchgrass indicated that some miRNAs expressed only in certain cell types or tissues, or certain developmental stages [13,22,23,35–37]. The expression of 15 conserved miRNA families was analyzed in 8 different Sorghum 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 with seed setting was initiated) (Fig. 2). Several miRNAs, miR172, miR156, miR319, miR159, miR529, miR164, miR160, miR167 and miR169, showed the highest expression in inflorescence relative to other tissues (Fig. 2). As well, the expression levels varied highly between mature and young inflorescence tissues: miR156, miR159, miR160, miR164, miR319 and miR172 were expressed at low levels in young inflorescence, whereas miR396 was expressed at high levels in mature inflorescence. Also, miR156 and miR159 expression levels greatly varied between the leaves from young plants or adult plants: low

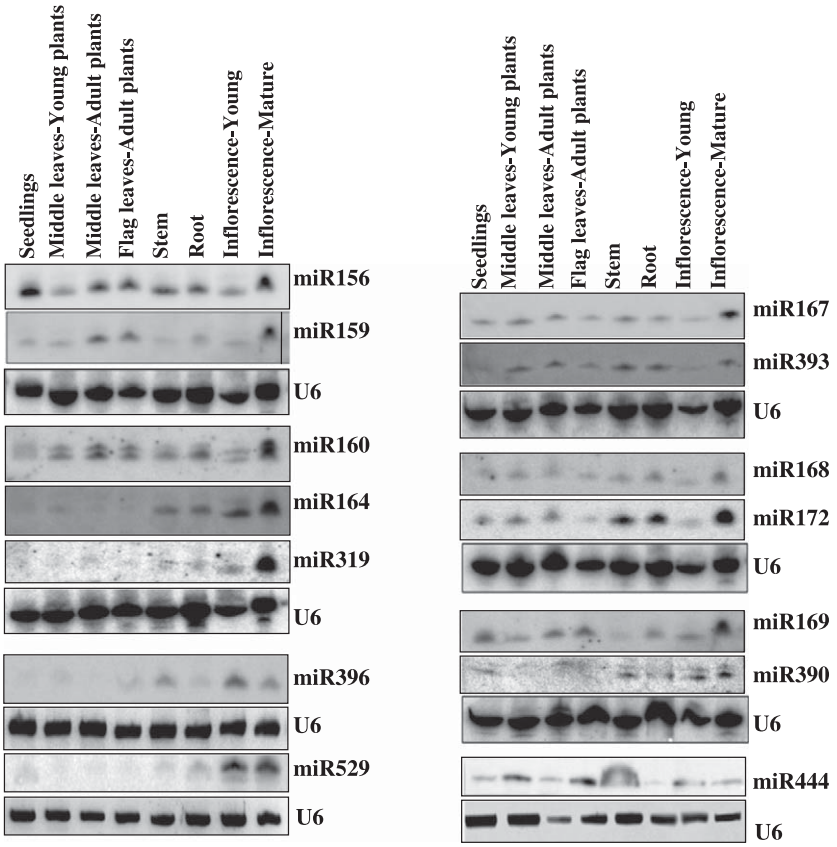


Fig. 2. Expression patterns of conserved miRNAs in different tissues. Blots were stripped and re-probed with ^{32}P-end-labeled oligonucleotides complementary to the U6 probe, which served as a loading control.

Please cite this article as: L. Zhang, et al., Genomics (2011), doi:10.1016/j.ygeno.2011.08.005

expression in leaves from young plants but high expression in leaves from adult plants (Fig. 2). miR529 had almost a similar level of abundance in both young and mature inflorescence tissues (Fig. 2).

Both miR168 and miR167 showed almost uniform expression in all tissues examined, with the exception that mature inflorescence had relatively high level of expression whereas the young inflorescence had the least expression (Fig. 2). In Sorghum seedlings, miR156 is abundantly expressed; miR167, miR169, miR172, miR390 and miR444 are moderately expressed; miR393 and miR529 are expressed at extremely low abundance and miR164, miR396 and miR319 could not be detected (i.e., below the detection limit). With the exception of miR169 and miR396 (miR169, miR396 and miR444), which had low expression levels in stem and root, the remaining 13 miRNAs did not differ in their abundance in roots and stems.

3.3. miRNAs induced under nutrient deprivation

It is known that miR395, miR399 and miR398 in Arabidopsis and *M. truncatula* are induced in response to sulfate-, phosphate- and copper-deprived conditions, respectively [22,28,33,38]. In addition, the expression of miR397 and miR408 is elevated in response to copper deficiency in *Brassica* and *Arabidopsis* [32–34]. By contrast, miR395 and miR399 are constitutively expressed in plants grown with 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 [23]. To analyze the response of miR395, miR397, miR398, miR399 and miR408 in Sorghum, 3-week-old seedlings were exposed to sulfate, phosphate or copper-deprived conditions. miR395 was induced in shoots and roots in response to low sulfate (Fig. 3A). miR399 was induced during phosphate deficiency in both roots and shoots (Fig. 3B). miR397, miR398 and miR408 were induced in shoots and roots in response to copper deficiency (Fig. 3C).

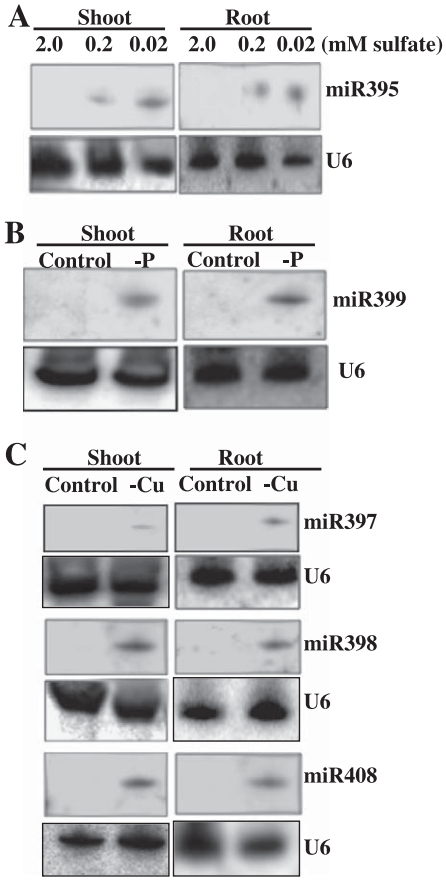


Fig. 3. miRNAs induced in response to growth medium with deprivation of nutrients. (a–c) Small RNA blots of 20-μg low-molecular-weight RNA isolated from Sorghum grown continuously in the same growth medium (control) or transferred to medium without (A) phosphate, (B) sulfate or (C) copper. Blots were rehybridized with the U6, which served as a loading control.

3.4. Prediction of miRNA targets and their validation

A total of 100 genes were predicted as targets for conserved miRNAs (Table 2 and Supplemental Table 2). 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 7 Squamosa promoter binding transcription factors; miR159, 4 MYB transcription factors; miR160, 6 auxin response factors; miR164, 6 No Apical Meristem-containing transcription factors; miR165/166, 8 Homeobox transcription factors; miR167, 3 auxin response factors; miR168, 4 piwi-domain-containing Argonaute proteins; miR169, 2 nuclear transcription factor Y subunit; miR170/171, 7 SCARECREW (GRAS domain-containing) transcription factors; miR172, 4 Apetala-2-related transcription factors; miR319, 5 TCP/MYB factors miR396, 6 growth-regulating transcription factors; and miR444, 5 MADS box transcription factors (Table 2 and Supplemental Table 2). Other predicted targets include F-box containing proteins (miR394), a sulfate transporter and an ATP sulfurylase (miR395), 11 laccases (miR397), 2 Cu/Zn superoxide dismutases (miR398), 4 inorganic phosphate transporters (miR399), 1 ubiquitin-activating enzyme (miR399), 5 plastocyanin-domain containing proteins (miR408), and 2 laccases (miR408) (Table 2 and Supplemental Table 2). Interestingly, AGO-1 homologs underwent multiple duplications in monocots, and 4 of these homologs are targeted by miR168 in rice [39]. Similarly, miR168 family has been predicted to target 4 AGO-1 homologs in Sorghum. Tas3siRNAs have been predicted to target 4 auxin response factors, like in other plant species. These target genes are likely to be involved in wide variety of physiological processes including growth and development, nutrient acquisition, translocation and assimilation and biotic and abiotic stress responses in Sorghum.

mRNAs that are not targeted by the miRNAs will be intact, with a polyA tail at the 3′end and a cap structure at the 5′end. However, mRNAs, which are subjected to miRNA-guided cleavage, are sliced between 10th and 11th nt of the complementary region. Such cleaved products can be verified by modified 5′-RACE assay [14,40]. Using this assay, ATP sulfurylase targeted by miR395 (Fig. 4A), laccase (multicopper oxidase) targeted by miR397 (Fig. 4B), Cu/Zn superoxide dismutase targeted by miR398 (Fig. 4C) and phosphate transporter targeted by miR399 (Fig. 4D) were confirmed as targets by mapping the respective miRNA-directed cleavages in vivo.

3.5. Identification, characterization and expression analysis of novel miRNAs in Sorghum

Recent deep-sequencing efforts in *Arabidopsis*, rice, *M. truncatula* and several other plant species led to the identification of several novel miRNAs, which are either conserved in closely related species (lineage-specific) or not conserved but species-specific miRNAs [15–17,20,22,41]. Given the plant small RNA populations include not only miRNAs but also endogenous siRNAs, which outnumbers the miRNA population both in number and diversity, identifying such novel non-conserved miRNAs requires their detection in *dcl-1* mutant which is unavailable for many plant species. Because of this situation and that some of the siRNAs in the miRBase have been mis-annotated as miRNAs, the plant small RNA community has established a set of criteria for correct annotation of miRNAs, particularly “non-conserved miRNAs” in plants, which includes the sequencing of an miRNA* read in the library [26]. Our sequence analysis revealed 13 candidates as potential novel miRNAs based on the predicted fold-back structures (Fig. 5), but only 9 small RNAs (sbi-MIR5564a, sbi-MIR397, sbi-MIR5566, sbi-

Please cite this article as: L. Zhang, et al., Genomics (2011), doi:10.1016/j.ygeno.2011.08.005

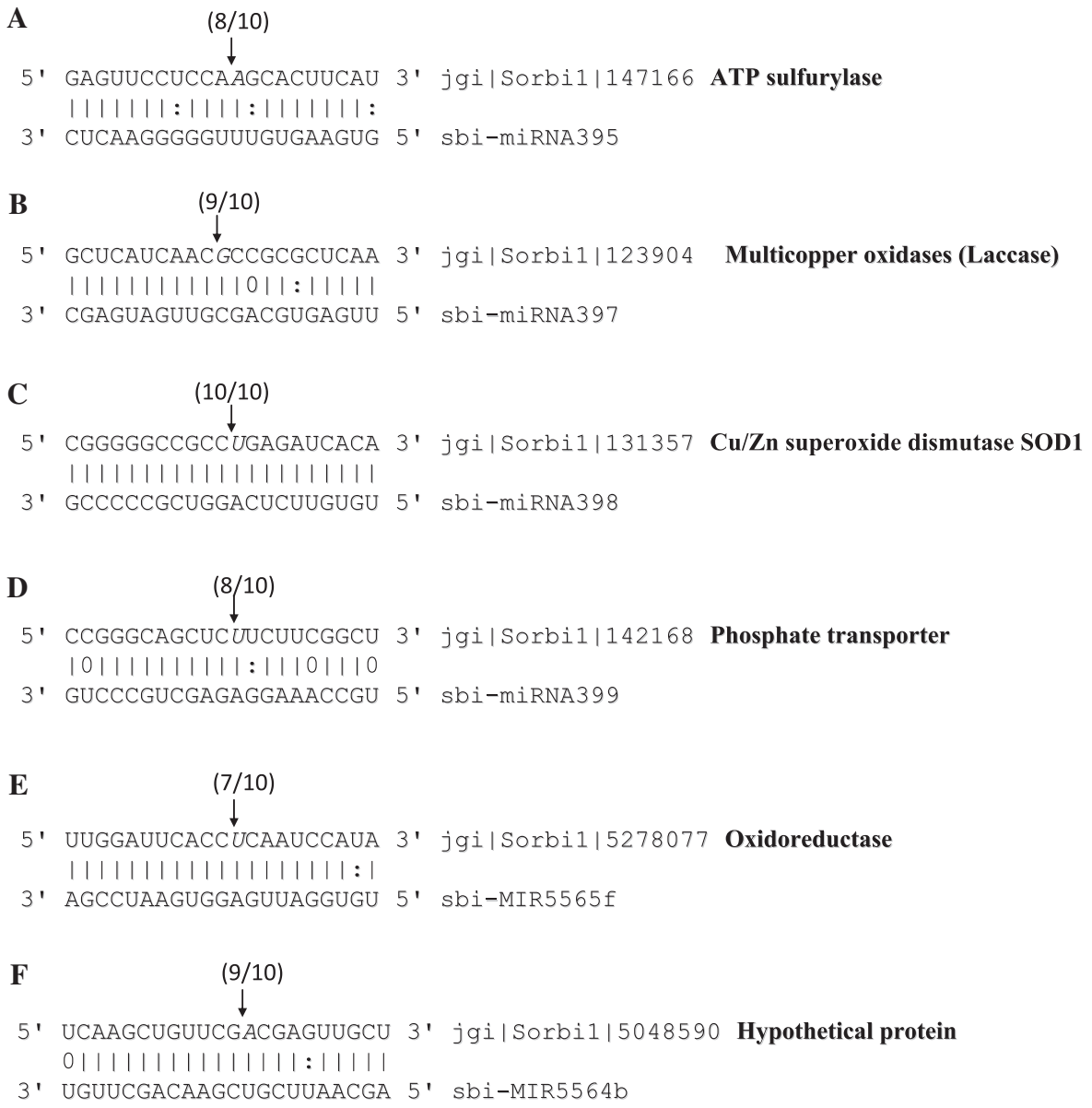


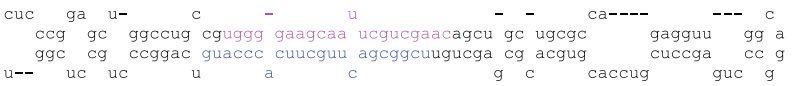
Fig. 4. Validation of miRNA-guided cleavages on mRNA targets in Sorghum. In each alignment, the perfect matches, G-U wobbles and mismatches are represented by straight lines, colons and circles, respectively. The site of miRNA-guided cleavage on the target mRNA is shown by arrow and the fraction of cloned products that terminated at the predicted cleavage site are shown in parentheses.

MIR5568, sbi-MIR5387b, sbi-MIR5569, sbi-MIR5570, sbi-MIR5565 and sbi-MIR5565d) are supported by the miRNA*, and the remaining 5 did not have a corresponding star-sequence in the library (Table 3). Limited sequencing depth (about 619,000 reads in this study relative to several millions of reads in other studies) may be a reason for not sequencing miRNA* read for some of the novel low-abundant miRNAs in Sorghum. For 4 cases (sbi-MIR5564b, sbi-MIR5565e, sbi-MIR5567 and sbi-MIR5565c), which appeared to be novel miRNAs but did not have corresponding miRNA* sequences in the library, we conducted miRNA* expression analysis with small RNA blot analysis. Because the abundance of miRNA* species is usually lower than that of miRNA species, we used relatively higher amounts of RNA (50 μg of low-molecular-weight RNA) for detecting miRNA* species, as compared with 20 μg used for detecting the expression of novel miRNAs. Small RNA blot analysis revealed the miRNA* expression for 4 novel mRNAs (sbi-MIR5564b, sbi-MIR5565e, sbi-MIR5567 and sbi-5565c) but not for surrounding 21 nt probes (data not shown). The conservation of novel miRNAs in closely related species could be additional supporting evidence for the annotation of novel small RNAs as "novel miRNAs" [22]. To examine the conservation of these novel miRNAs in closely related monocots, we performed BLAST searches against the NCBI EST database. Surprisingly, 7 of these novel miRNA sequences are conserved at least in one another monocot (sugarcane, maize, wheat or switchgrass), which suggests that some of these could be designated monocot-specific or lineage-specific miRNAs (Supplemental Fig. 2). Fold-back structures for the novel miRNA precursors from sugarcane, wheat or maize could be predicted from their precursor sequences (Supplemental Fig. 2). On the basis of miRNA* reads in the library, detection of miRNA* expression by small RNA blot analysis and/or their conservation in related monocots, we annotated 13 small RNAs as "novel miRNAs" in Sorghum. Seven novel miRNAs are

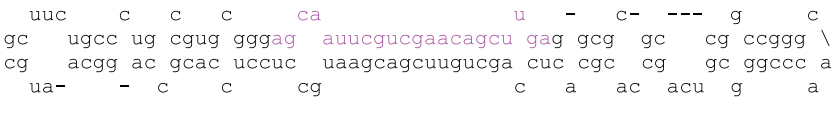
Fig. 5. Predicted fold-back structures for the novel miRNAs using their precursors. The sequences colored in pink and blue represent the cloned mature miRNA sequence and miRNA* sequences, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: L. Zhang, et al., Genomics (2011), doi:10.1016/j.ygeno.2011.08.005

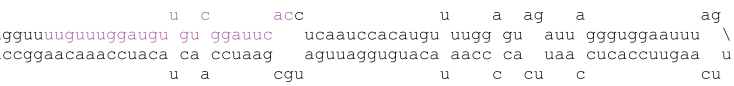
sbi-MIR5564a



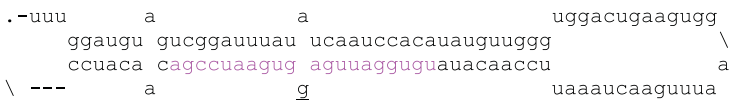
sbi-MIR5564b



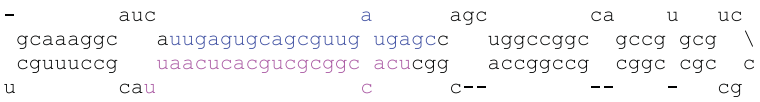
sbi-MIR5565e



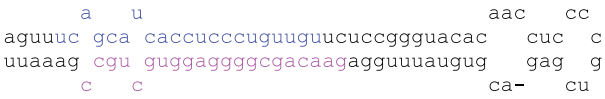
sbi-MIR5565f



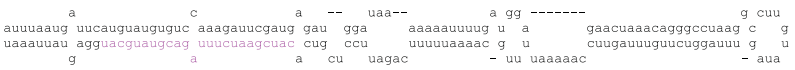
sbi-MIR397



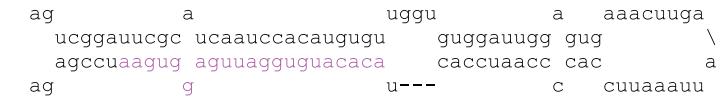
sbi-MIR5566



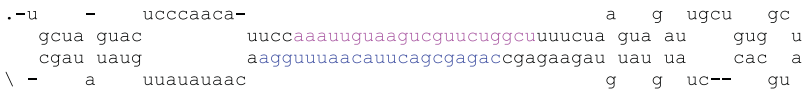
sbi-MIR5567



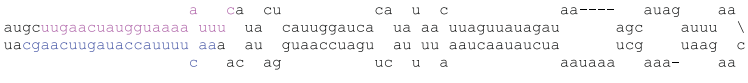
sbi-MIR5565c



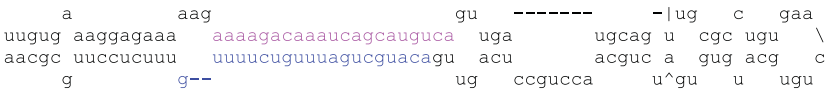
sbi-MIR5568



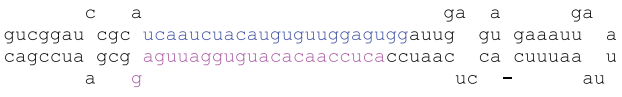
sbi-MIR5569



sbi-MIR5570



sbi-MIR5565



sbi-MIR5565d

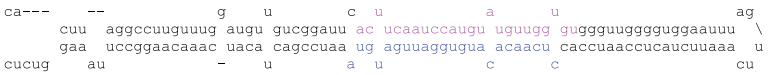


Table 3
Identified novel miRNAs in Sorghum on the basis of expression/detection of miRNA* or their conservation in related monocots. (NB: small RNA blot analysis).

miRNA	miRNA sequence	miRNA frequency	NB	miRNA* frequency	NB	Conservation
sbi-MIR5564a	UGGGGAAGCAAUUCGUCGAAC	1472	+	5		Sugarcane
sbi-MIR397	UCACCGGCGCUGCACUCAAUU	15	+	14		Switchgrass
sbi-MIR5566	GAACAGCGGGGAGGUGCUGCC	2	+	3		
sbi-MIR5568	AAAUUGUAAGUCGUUCUGGCU	7		1		Wheat
sbi-MIR5387b	CGUGGCUCUGACCGGUGCUAAA	1		1		
sbi-MIR5569	UUGAACUAUGGUAAAAAUUUC	1		1		
sbi-MIR5570	AAAAGACAAAUCAGCAUGUCA	1		1		
sbi-MIR5565	ACUCCAACACAUGUGGAUUGAG	7		1		Sugarcane, wheat
sbi-MIR5565d	ACUUCAAUCCAUGUAUGUUGGU	1		1		
sbi-MIR5564b	AGCAAUUCGUCGAACAGCUUGU	130	+	0	+	
sbi-MIR5565e	UUGUUUGGAUGUUGUCGGAUUC	43	+	0	+	Wheat, maize
sbi-MIR5567	CAUCGAAUCUUUAGACGUAUGC	55	+	0	+	
sbi-MIR5565c	ACACAUGUGGAUUGAGGUGAA	154	+	0	+	Wheat, sugarcane
sbi-MIR5565f	UGUGGAUUGAGGUGAAUCCGA	34	+	0		Wheat

conserved at least in one another monocot plant species, which strongly suggests their annotation as novel lineage-specific miRNAs. Because the homologs for the remaining 7 miRNAs (sbi-MIR5566, sbi-MIR5564b, sbi-MIR5567, sbi-MIR5387b, sbi-MIR5569, sbi-MIR5570 and sbi-MIR5565d) could not be found in other monocots, these could be annotated as Sorghum-specific miRNAs. Interestingly, the frequency (reads count) of a few novel miRNAs (sbi-MIR5564a, sbi-MIR5564b and sbi-MIR5565c) is substantially higher and even greater than that of several conserved miRNAs (Tables 2 and 3). Recently evolved miRNAs show inaccurate processing of the precursor, resulting in a somewhat heterogeneous population of mature miRNAs. We observed such heterogeneous population of miRNAs and miRNA-stars for several novel miRNAs in Sorghum (Supplemental Fig. 3).

We also analyzed the expression of these novel miRNAs in several tissues. We could detect abundant and ubiquitous expression of sbi-MIR5564a in all tissues analyzed (Fig. 6). By contrast, most other novel miRNAs showed low and uniform expression in all tissues, with the exception of sbi-MIR5566 and sbi-MIR5564b, which showed high expression in root tissue relative to other tissues.

We also predicted 25 genes as potential targets for the 13 novel miRNAs (Supplemental Table 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 (sbi-MIR5565), SNARE protein syntaxin (sbi-MIR5565d), SAM decarboxylase (sbi-MIR5570), putative receptor kinases (sbi-MIR5565f) and oxidoreductase and arabinogalactan protein (sbi-MIR5565e) are also among the predicted targets for the novel miRNAs in Sorghum (Supplemental Table 3). Two of these novel miRNA targets (an oxidoreducatse targeted by novel miR-sbi-MIR5565f and a hypothetical protein targeted by novel miR-S91856) have been validated in Sorghum (Fig. 4E and F).

Our current knowledge of the regulatory roles of miRNAs and their targets point to their 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 [1,2]. 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 [3,31]. Furthermore, overexpression of miR156 causes a delay in flowering, initiates leaves faster and also causes a severe decrease of apical dominance in *Arabidopsis* [27]. The combination of these traits leads to a 10-fold higher total leaf number in transgenic plants than in wild-type plants. Similarly, miR169 plays a role in drought tolerance

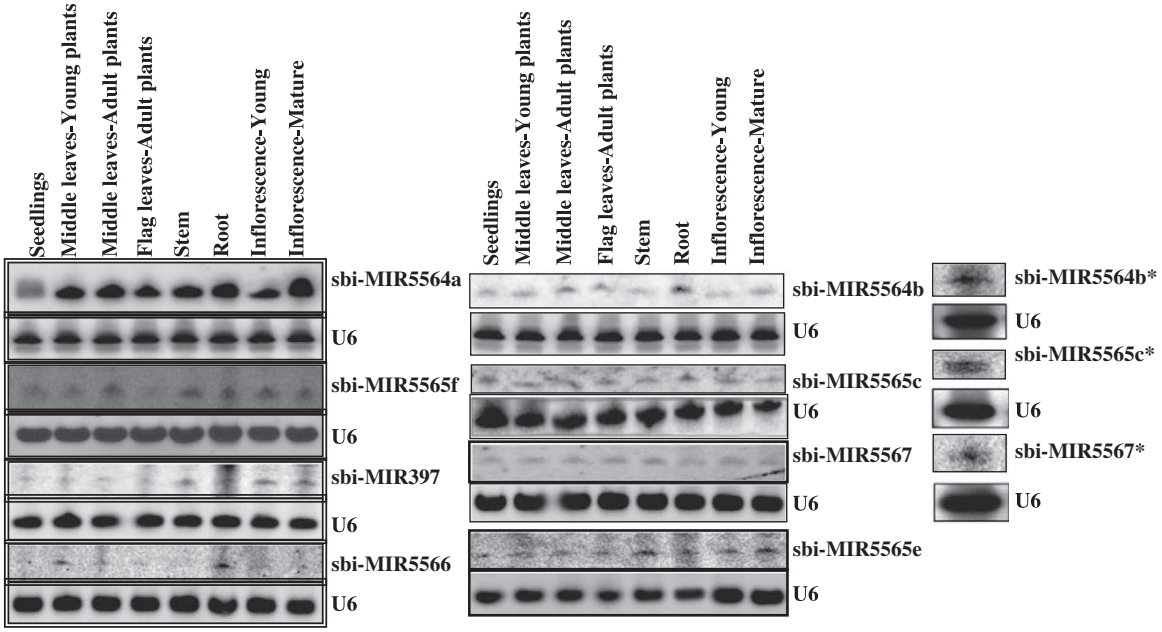


Fig. 6. Expression patterns of newly identified miRNAs in Sorghum as determined by small RNA blot analysis. (a) 20 μg low-molecular-weight RNA was used for detection of the miRNA and (b) 50 μg low-molecular-weight RNA was used to detect the expression of miRNA* reads.

in *Arabidopsis* [42] and suspected to play similar role in Sorghum. Thus, finding miRNAs in biofuel plants has implications 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 genetic engineering of Sorghum with improved quality traits. These results have 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 cereal crop as well as biofuel plant species.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.08.005.

Acknowledgments

Support for this research was provided by the Oklahoma Agricultural Experiment Station in the form of a Team Initiative Proposal and in part by the NSF EPSCoR award EPS0814361 to R.S. and a start-up grant of Fudan University and 10ZR1403000 of STCSM to YZ. Sweet Sorghum (M81E) seeds were kind gift from Dr. Gopal Kakani, Oklahoma State University.

References

[1] A.C. Mallory, H. Vaucheret, Functions of microRNAs and related small RNAs in plants, Nat. Genet. 38 (2006) S31–S36 Suppl.
[2] M.W. Jones-Rhoades, D.P. Bartel, B. Bartel, MicroRNAs and their regulatory roles in plants, Annu. Rev. Plant Biol. 57 (2006) 19–53.
[3] R. Sunkar, MicroRNAs with macro-effects on plant stress responses, Semin. Cell Dev. Biol. 21 (2010) 805–811.
[4] V. Ramachandran, X. Chen, Degradation of microRNAs by a family of exoribonucleases in Arabidopsis, Science 321 (2008) 1490–1492.
[5] B. Yu, Z. Yang, J. Li, S. Minakhina, M. Yang, R.W. Padgett, R. Steward, X. Chen, Methylation as a crucial step in plant microRNA biogenesis, Science 307 (2005) 932–935.
[6] M.Y. Park, G. Wu, A. Gonzalez-Sulser, H. Vaucheret, R.S. Poethig, Nuclear processing and export of microRNAs in Arabidopsis, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 3691–3696.
[7] P. Brodersen, O. Voinnet, Revisiting the principles of microRNA target recognition and mode of action, Nat. Rev. Mol. Cell Biol. 10 (2009) 141–148.
[8] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, et al., The Sorghum bicolor genome and the diversification of grasses, Nature 457 (2009) 551–556.
[9] T. Sasaki, B.A. Antonio, Plant genomics: Sorghum in sequence, Nature 457 (2009) 547–548.
[10] M.L. Wang, C. Zhu, N.A. Barkley, Z. Chen, J.E. Erpelding, S.C. Murray, M.R. Tuinstra, T. Tesso, G.A. Pederson, J. Yu, Genetic diversity and population structure analysis of accessions in the US historic sweet Sorghum collection, Theor. Appl. Genet. 120 (2009) 13–23.
[11] Z. Xin, M.L. Wang, N.A. Barkley, G. Burow, C. Franks, G. Pederson, J. Burke, Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population, BMC Plant Biol. 8 (2008) 103.
[12] T. Arazi, M. Talmor-Neiman, R. Stav, M. Riese, P. Huijser, D.C. Baulcombe, Cloning and characterization of micro-RNAs from moss, Plant J. 43 (2005) 837–848.
[13] R. Sunkar, J.K. Zhu, Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis, Plant Cell 16 (2004) 2001–2019.
[14] R. Sunkar, T. Girke, P.K. Jain, J.K. Zhu, Cloning and characterization of microRNAs from rice, Plant Cell 17 (2005) 1397–1411.
[15] R. Sunkar, X. Zhou, Y. Zheng, W. Zhnag, J.K. Zhu, Identification of novel and candidate miRNAs in rice by high throughput sequencing, BMC Plant Biol. 8 (2008) 25.
[16] C. Lu, D.H. Jeong, K. Kulkarni, M. Pillay, K. Nobuta, R. German, S.R. Thatcher, C. Maher, L. Zhang, D. Ware, B. Liu, X. Cao, B.C. Meyers, P.J. Green, Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs), Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 4951–4956.
[17] C. Lu, K. Kulkarni, F.F. Souret, R. MuthuValliappan, S.S. Tej, R.S. Poethig, I.R. Henderson, S.E. Jacobsen, W. Wang, P.J. Green, B.C. Meyers, MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant, Genome Res. 16 (2006) 1276–1288.
[18] Q.H. Zhu, A. Spriggs, L. Matthew, L. Fan, G. Kennedy, F. Gubler, C. Helliwell, A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains, Genome Res. 18 (2008) 1456–1465.
[19] M.J. Axtell, J.A. Snyder, D.P. Bartel, Common functions for diverse small RNAs of land plants, Plant Cell 19 (2007) 1750–1769.
[20] N. Fahlgren, M.D. Howell, K.D. Kasschau, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, T.F. Law, S.R. Grant, J.L. Dangl, J.C. Carrington, High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes, PLoS One 2 (2007) e219.
[21] R. Sunkar, G. Jagadeeswaran, In silico identification of conserved microRNAs in large number of diverse plant species, BMC Plant Biol. 8 (2008) 37.
[22] G. Jagadeeswaran, Y. Zheng, Y.F. Li, L.I. Shukla, J. Matts, P. Hoyt, S.L. Macmil, G.B. Wiley, B. Roe, W. Zhang, R. Sunkar, Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families, New Phytol. 184 (2009) 85–98.
[23] J. Matts, G. Jagadeeswaran, B.A. Roe, R. Sunkar, Identification of microRNAs and their targets in switchgrass, a model biofuel plant species, J. Plant Physiol. 167 (2010) 896–904.
[24] J.A. Bedell, M.A. Budiman, A. Nunberg, R.W. Citek, D. Robbins, J. Jones, E. Flick, T. Rholfing, J. Fries, K. Bradford, J. McMenamy, M. Smith, H. Holeman, B.A. Roe, G. Wiley, I.F. Korf, P.D. Rabinowicz, N. Lakey, W.R. McCombie, J.A. Jeddeloh, R.A. Martienssen, Sorghum genome sequencing by methylation filtration, PLoS Biol. 3 (2005) e13.
[25] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
[26] B.C. Meyers, M.J. Axtell, B. Bartel, D.P. Bartel, et al., Criteria for annotation of plant MicroRNAs, Plant Cell 20 (2008) 3186–3190.
[27] R. Schwab, J.F. Palatnik, M. Riester, C. Schommer, M. Schmid, D. Weigel, Specific effects of microRNAs on the plant transcriptome, Dev. Cell 8 (2005) 517–527.
[28] H. Fujii, T.J. Chiou, S.I. Lin, K. Aung, J.K. Zhu, A miRNA involved in phosphate-starvation response in Arabidopsis, Curr. Biol. 15 (2005) 2038–2043.
[29] T.J. Chiou, K. Aung, S.I. Lin, C.C. Wu, S.F. Chiang, C.L. Su, Regulation of phosphate homeostasis by MicroRNA in Arabidopsis, Plant Cell 18 (2006) 412–421.
[30] E. Allen, Z. Xie, A.M. Gustafson, J.C. Carrington, microRNA-directed phasing during trans-acting siRNA biogenesis in plants, Cell 121 (2005) 207–221.
[31] R. Sunkar, R.V. Chinnusamy, J. Zhu, J.K. Zhu, Small RNAs as big players in plant abiotic stress responses and nutrient deprivation, Trends Plant Sci. 12 (2007) 301–309.
[32] H. Yamasaki, S.E. Abdel-Ghany, C.M. Cohu, Y. Kobayashi, T. Shikanai, M. Pilon, Regulation of copper homeostasis by micro-RNA in Arabidopsis, J. Biol. Chem. 282 (2007) 16369–16378.
[33] S.E. Abdel-Ghany, M. Pilon, MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis, J. Biol. Chem. 283 (2008) 15932–15945.
[34] A. Buhtz, F. Springer, L. Chappell, D.C. Baulcombe, J. Kehr, Identification and characterization of small RNAs from the phloem of *Brassica napus*, Plant J. 53 (2008) 739–749.
[35] X. Chen, A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development, Science 303 (2004) 2022–2025.
[36] J.P. Combier, F. Frugier, F. de Billy, A. Boualem, F. El-Yahyaoui, S. Moreau, T. Vernie, T. Ott, P. Gamas, M. Crespi, A. Niebel, MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*, Genes Dev. 20 (2006) 3084–3088.
[37] A. Boualem, P. Laporte, M. Jovanovic, C. Laffont, J. Plet, J.P. Combier, A. Niebel, M. Crespi, F. Frugier, MicroRNA166 controls root and nodule development in *Medicago truncatula*, Plant J. 54 (2008) 876–887.
[38] M.W. Jones-Rhoades, D.P. Bartel, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA, Mol. Cell 14 (2004) 787–799.
[39] Y.F. Li, Y. Zheng, C. Addo-Quaye, L. Zhang, A. Saini, G. Jagadeeswaran, M.J. Axtell, W. Zhang, R. Sunkar, Transcriptome-wide identification of microRNA targets in rice, Plant J. 62 (2010) 742–759.
[40] C. Llave, Z. Xie, K.D. Kasschau, J.C. Carrington, Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis microRNA, Science 297 (2002) 2053–2056.
[41] R. Rajagopalan, H. Vaucheret, J. Trejo, D.P. Bartel, A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*, Genes Dev. 20 (2006) 3407–3425.
[42] W.X. Li, Y. Oono, J. Zhu, X.J. He, J.M. Wu, K. Iida, X.Y. Lu, X. Cui, H. Jin, J.K. Zhu, The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and post-transcriptionally to promote drought resistance, Plant Cell 20 (2008) 2238–2251.