CHARACTERIZATION OF COTTON LEAF miRNAs AND CHANGES IN miRNA EXPRESSION ASSOCIATED WITH HEAT STRESS IN COTTON AS DETERMINED BY DEEP SEQUENCING

Hongtao Hu
Narendra K. Singh
David Weaver
Auburn University
Auburn, AL
Shankar Pant
Marceline Egnin
Tuskegee University
Tuskegee, AL
Min Zhong
Chiachen Weng
Robert Locy
Auburn University
Auburn, AL

Abstract

Total RNA was extracted from the youngest fully expanded leaf on 6-week-old cotton seedlings of either genotype Delta Pineland 90 (DP90) or USDA accession TX337 that had either been heat stressed in a growth chamber at 42^{o}C for 4 or 24 hours (HS) or not heat-stressed (controls, grown at 25^{o}C). Total RNA was prepared from each sample, and a small RNA library was generated and sequenced using an Illumina pyrosequencer. The small RNAs were analyzed using a self-constructed pipeline involving shell scripts, the Bowtie assembler, a modified miRDeep script modified for use with plant miRNA precursors, and the Target-align mRNA target finder. In total we discovered 280 cotton leaf miRNAs that are derived from at least 191 precursor hairpins that were predicted from sequences found in cottonDB. These miRNAs have been partially characterized, and their mRNA targets predicted. Expression of both the miRNAs and their targets suggests that in cotton leaves, miRNAs can regulate temperature dependent gene expression, but other factors including tissue specific regulation of miRNAs impinge on the regulation of genes by heat stress making a complex picture of gene regulation in cotton leaves during thermal stress.

Introduction

Environmental stresses, particularly heat and drought stress, are major factors limiting cotton yield throughout the cotton belt in the US and worldwide. Cotton is a crop believed to be well adapted to hot and relatively "droughty" environments, but because the germplasm base in cotton appears to be relatively narrow, it is not clear that we have exploited the optimum of the *G. hirsutum* gene pool for heat and drought stress. Our goal is to develop a system for understanding, particularly those genes that impart heat tolerance to the cotton crop, and to genetically manipulate more heat and drought tolerant cotton plants using this information as we attempt to develop broader-based germplasm for cotton.

MicroRNAs (miRNAs) are a type of small regulatory RNA that is derived from primary transcripts that resemble mRNA primary transcripts, i.e. they contain a 5'-cap, a polyA tail, and typical introns and exons. Also miRNA primary transcripts form thermodynamically stable hair-pin structures (Meyer et al., 2008). Mature miRNAs are processed from these primary hairpin transcripts by an enzyme called DICER in conjunction with other proteins before leaving the nucleus of the cell. In the cytoplasm of the cell, mature miRNAs bind with sequence-specific, target messenger RNAs leading to specific target mRNA degradation, thus inactivating the mRNA, and silencing the expression of the specific gene(s).

Work on miRNAs in plants has rapidly advanced in recent years due to the advent of so called next generation or deep sequencing techniques. The work described here presents our progress to date on the identification of specific miRNAs found in cotton leaf, and summarizes changes in the expression of these miRNAs in response to heat stress. The key to our success in this regard was the bioinformatic application of a set of scripts referred to as the miRDeep program (Friedlander *et al.*, 2008). This program, with modification for the specific mechanism used to produce miRNAs in plants, has proven very effective in identifying miRNA precursors in species without a sequenced genome and with an extensive EST assembled collection.

Materials and Methods

Plant Material and Stress Treatment

Four- to six-week-old greenhouse-grown plants of DPL90 and TX337 were placed in a growth chamber, and given 4 and 20-hour heat treatments at 42 C. Leaf samples (first and second fully expanded leaves) were taken from *circa* 6 week old plants and samples at 0 hours of heat stress were taken as the unheated controls, while samples from 4 and 20 hours of heat stress were pooled and considered the heat-stressed samples.

RNA Extraction

Total RNA was separately extracted from the four samples using the hot borate method (Wan and Wilkins, 1994; Meisel *et al*., 2005).

cDNA library preparation & Deep Sequencing

cDNA was prepared from the total RNA from each of these 4 treatments as described above, and submitted for Roche 454-Flex MPSS sequencing at the University of South Carolina Environmental Genomics facility (ENGENCORE). The analysis of the raw sequence data was conducted according to Pant et al. (2011).

SmallRNA library preparation & Deep Sequencing

Small NA was prepared from the total RNA and the library was sequenced on an Illumina Sequencer at the Core Genomics facility at Virginia Commonwealth University.

Bioinformatic Analysis

Figure 1. Describes the pipeline we have constructed to analyze small sequence reads and determine mature and precursor miRNAs in cotton. Raw sequence reads (15,690,561 reads) 39 nucleotides in length in FASTQ format coming straight from the Illumina sequencer run, were processed into crude putative sRNA reads using a custom prepared shell script running in a Linux command window. This script removed the 3' and 5' sequencing adaptor sequences from the raw reads and removed all sequenced inserts smaller than 15 nucleotides from the dataset. This yielded 11,063,659 adaptor-trimmed total reads (see Table 1).

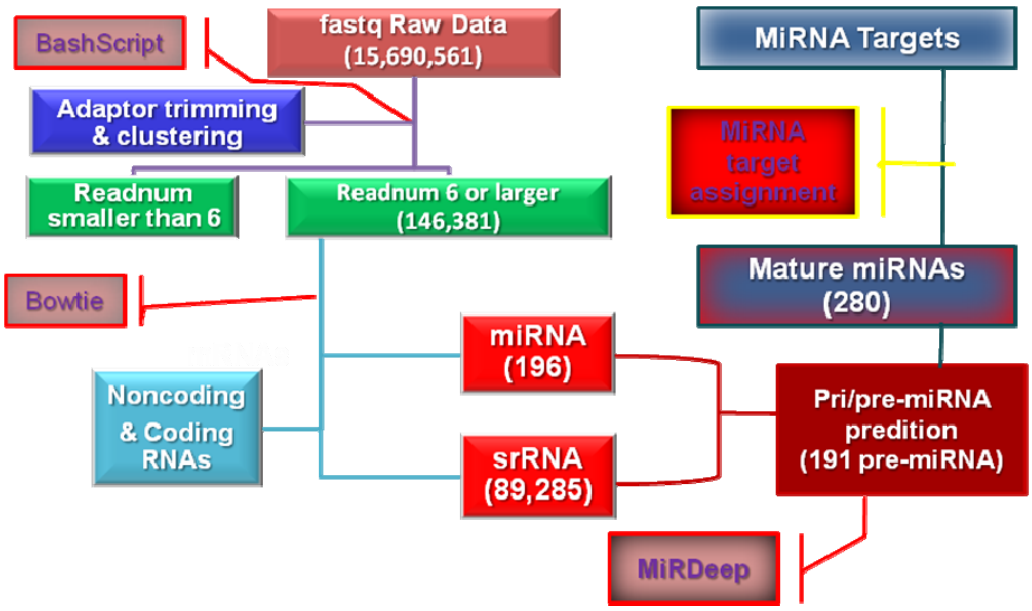

Figure 1. Flow diagram showing the processing of the raw sequence data reads from Illumina sequencing to obtain putative miRNA precursors, and putative sRNAs. Targets for the miRNAs were also bioinformatically estimated.

These reads were then clustered into a single file of unique sequences in each dataset, and the number of adaptor-trimmed reads clustering in each unique-read group were determined for each of the 4 datasets and for the total

dataset. Low abundance reads (readnum =< 5 were then filtered out yielding 146,381 unique groups that appeared in the datasets at least 5 times or more. Those sequences with a match to known miRNAs found in the miRBase database (Griffith-Jones, *et al.*, 2006; 2008) were determined using the Bowtie assembler (Langmead *et al.*, 2009). This amounted to 26,300 of the 146,381 total reads.

After removal of the noncoding RNAs (tRNAs, rRNAs, snoRNAs, etc) by assembly against the Rfam database (Gardner *et al.*, 2009; Griffith-Jones *et al.*, 2005) and removal of the coding RNA sequences by assembly against cottonDB (www.cottondb.org), the remaining srRNA reads along with those sequences hitting miRBase were submitted to miRDeep (Friedlander *et al*., 2008) to predict pri- and pre-miRNAs. This yielded 280 mature miRNAs produced from 191 precursors. The Target-align program (Xie & Zhang, 2010) was used to predict miRNA target genes found in the NCBI cotton unigene dataset.

Results and Discussion

Sequencing Results and Bioinformatic miRNA Discovery

After trimming adaptor adaptors, we obtained a total of 11,690,561 reads from a total 15,690,451 raw reads. After removal of those unique reads with reand numbers less than 5, 8,155,818 high abundance reads that clustered into 146,381 unique groups (see Table 1) remained. One hundred six or these miRNAs had an exact match to miRbase, and 26,300 unique reads were homologous to known miRNAs with up to 2 mismatches. After removal of the noncoding RNAs, and RNA reads matching to known cotton protein-coding genes, a total of 89,285 sequences remain that have been tentatively called putative sRNAs.

Table 1. Summary of bioinformatic Analysis.

Genotype	TX337 Control	TX337 Heat Stress	DP90 Control	DP90 Heat Stress	Combined
Total Sequence Reads	3,664,304	4,338,236	3,974,693	3,713,328	15,690,561
Primer-trimmed	2,516,100	3,272,435	2,806,175	2,468,949	11,063,659
Removal of low abundance reads	2,108,448	2,778,096	1,942,350	1,326,924	8,155,818
Unique putative sRNA sequences	87,313	76,500	45,029	29,152	146,381
Average reads / unique read	24.15	36.31	43.14	45.52	55.72
miRbase hits					
Exact match	86	74	83	71	106
No mismatch in sequence	1515	1723	688	523	2,724
1 mismatch	4434	5381	1826	1645	8,727
2 mismatch	7318	7934	3131	2581	14,849
Total putative miRbase hits	13,353	15,112	5,728	4,820	26,300
Non-coding RNA sequences	2,188	794	1,084	739	2,247
Protein coding reads	12,887	11,199	14,114	8,117	28,549
Candidate siRNAs	58,885	49,395	24,103	15,476	89,285

When the primer-trimmed reads were examined for length, approximately 5 million of the reads had a length of either 21 or 24 nucleotides (see Figure 2). When the length of the 146,381 unique groups was examined, it was determined that the 24 nucleotide putative sRNAs tend to be the majority of the total read groups, but they have lower abundance when compared to the 21 nucleotide RNAs where there are a smaller number of groups with higher average read numbers is apparent (see Figure 2).

When the miRDeep (Friedlander *et al*., 2008) scripts were applied to the 115,585 putative sRNAs hitting and not hitting miRbase, a total of 280 mature miRNAs were identified (Table 2). One hundred six of these identified exact sequence matches to other miRNAs in miRbase (conserved miRNAs), and eighteen of these are miRNAs already found in cotton, while 88 correspond to miRNAs found in other plants, but not yet identified in cotton. An additional 174 putative miRNAs were identified among the inexact matches to miRbase and the 89,285 other reads, on the basis of the 94 novel precursors predicted by miRDeep.

Table 2. miRNAs and Precursors Discovered using miRDeep.

Description	Number
Total mature miRNAs discovered to date	280
1. Conserved miRNAs (miRBase exact match)	106
Known cotton mature miRNAs	18
Mature miRNAs from other plants	88
2. Putative novel mature miRNAs (srRNA reads)	174
Total putative pre-miRNAs discovered to date	191
1. Pre-miRNAs from conserved miRNAs	97
2. Pre-miRNA from srRNA reads	94

Characterization of the putative miRNA precursors obtained from miRDeep Analysis

In order to characterize the putative precursors and miRNAs discovered using the miRDeep program in the absence of a complete genome sequence for *Gossypium hirsutum*, the minimum free energies (MFE) of the hairpins predicted for cotton were compared with the MFEs of the precursors for all miRNAs in miRbase (see panels A and B, Figure 3). Plant pre-miRNA usually has a minimum free energy (MFE) from -100 kcal/mol to -20 kcal/mol as shown in Figure 3. The cotton miRNAs discovered with miRDeep demonstrate similar MFEs except that the MFE distribution is skewed to slightly lower MFEs indicating that the predicted cotton hairpins are similar to those of other plant miRNA precursors.

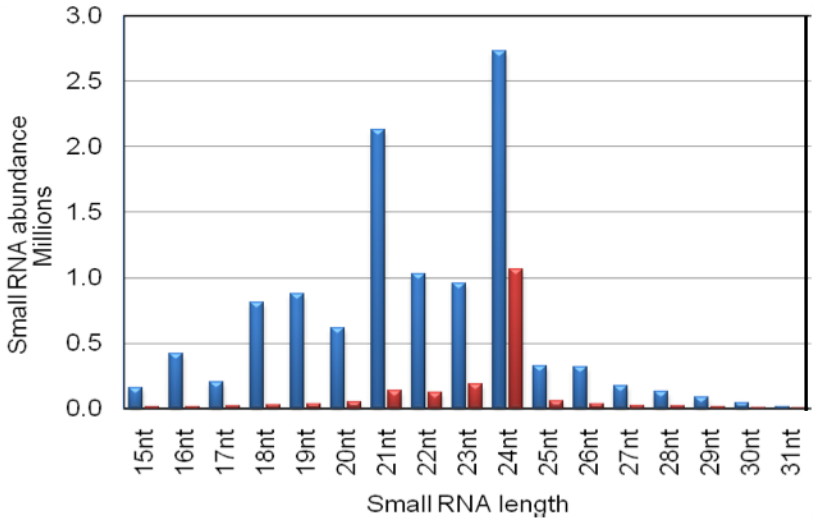

Figure 2. The distribution of the length of the inserts in the primer-trimmed reads (blue bars) is plotted to show that both 21 and 24 nucleotide long classes of putative sRNAs were the most abundant in the dataset. The distribution of the unique groups (red bars) is also shown for the same size classes.

Also the GU content of the predicted cotton miRNA precursors was also compared with the GU content of the reported precursors for all plant miRNAs found in miRBase (panels C and D, Figure 3). This figure shows that the GU content of the predicted cotton pre-miRNAs are similar in magnitude and distribution to those of other plant miRNA precursors. These facts demonstrate that the predicted cotton miRNA precursors are similar to those for plant miRNAs in general, and validates our bioinformatic approach to miRNA discovery.

Characterization of the heat-stress expression of cotton miRNAs.

Figure 4 shows the abundance (expression level) of each of the 106 conserved miRNAs from cotton leaves in control and during heat stress for both the heat tolerant genotype, TX337, and the less heat tolerant genotype, DP90. Our analysis shows that the abundance data for each of the 4 treatments is similar for all 280 of the miRNAs discovered in cotton leaves. Almost any expression pattern expected can be observed, i.e. one can find increases,

decreases, or no change in miRNA abundance in both DP90 and in TX337 in response to heat stress. Also the type of change observed in one genotype may be different from the type of change observed in the other genotype. These observations are typically much more prevalent for the lower abundance miRNAs than for those miRNAs more highly expressed. Consequently, the details of such analyses will require more careful subsequent analysis and experimental validation prior to being able to fully appreciate the relevance of the large number of changes observed. Additionally, it is unlikely that one can fully appreciate the changes taking place until all (or at least most) of the precursors for a given mature miRNA species are known. Some heat-induced changes in mature miRNAs may result from the heat-specific expression of a totally different miRNA gene/precursor than is expressed under other circumstances, and individual responses may vary from species to species.

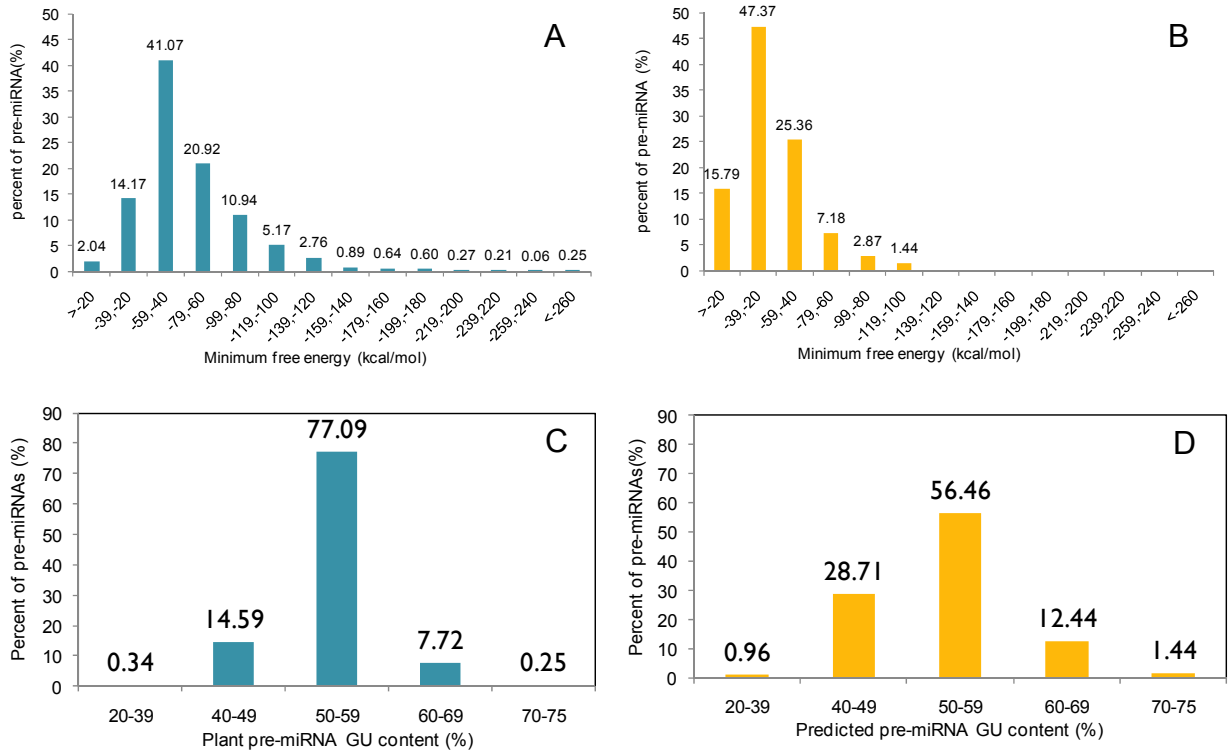

Figure 3. Comparison of the distribution of minimum free energy of the predicted hairpins for all plant miRNAs in miRbase (Panel A) with the distribution of the minimum free energy of the predicted hairpins for the cotton putative miRNA precursors discovered using miRDeep (Panel B). Comparison of the GU content distribution of plant pre-miRNAs found in miRbase (Panel C) with the GU content distribution of the predicted cotton pre-miRNAs (Panel D).

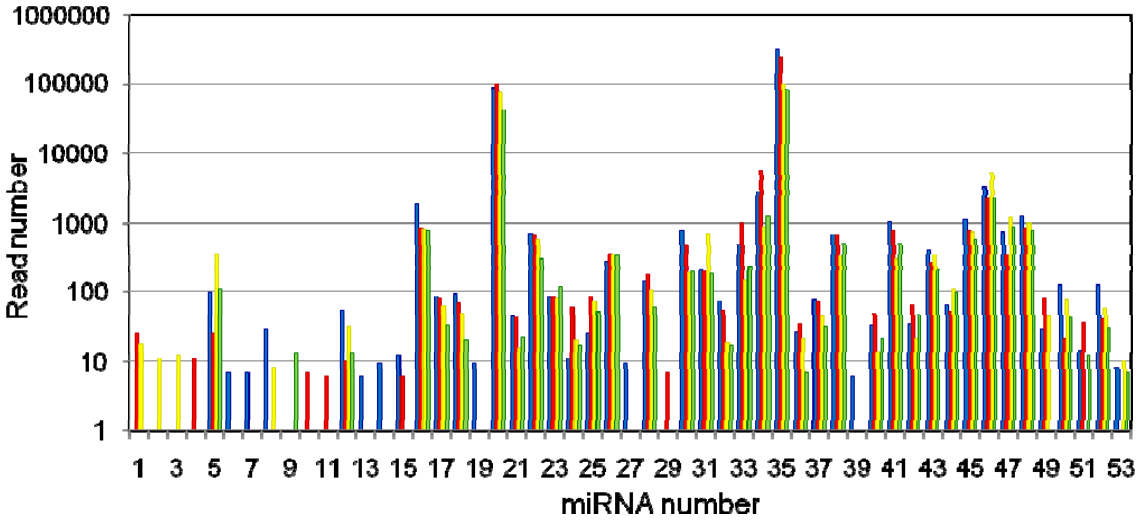

Figure 4A. Expression level of the 106 conserved miRNAs as a function of genotype and heat stress in each of the 4 experiments reported here. Figure 4A (above) shows miRNAs 1-53 while figure 4B (below) shows miRNAs 54-106. (Blue line) expression of genotype TX337 unstressed; (Red line) expression of genotype TX337 heat stressed at 4 and 24 hours; (Yellow line) expression of genotype DP90 unstressed; (Green line) expression of genotype TX337 heat stressed at 4 and 24 hours. The expression of a number of miRNAs increases with heat stress while others fall. This also can vary by genotype. MiR20 and miR35 are the most abundant miRNAs found in our dataset and are plotted on a separate scale.

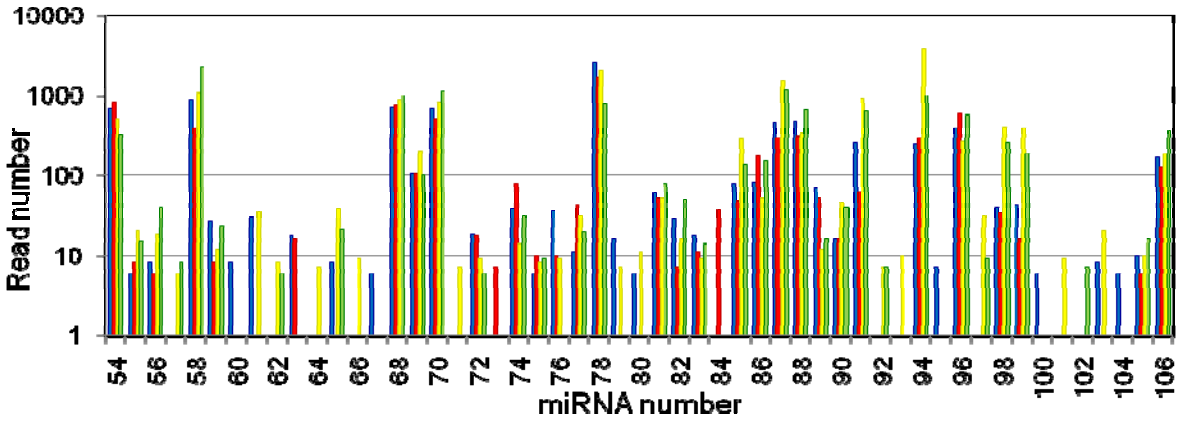

Figure 4B. Expression level of the 106 conserved miRNAs as a function of genotype and heat stress in each of the 4 experiments reported here. Figure 4A shows miRNAs 1-53 while figure 4B shows miRNAs 54-106. (Blue line) expression of genotype TX337 unstressed; (Red line) expression of genotype TX337 heat stressed at 4 and 24 hours; (Yellow line) expression of genotype DP90 unstressed; (Green line) expression of genotype TX337 heat stressed at 4 and 24 hours. The expression of a number of miRNAs increases with heat stress while others fall. This also can vary by genotype. MiR20 and miR35 are the most abundant miRNAs found in our dataset and are plotted on a separate scale.

Correspondence between miRNA Abundance and Gene Expression

When miRNAs were highly expressed (Figure 4, right side of graph), target gene expression was greatly reduced or eliminated regardless of the effect of heat stress on the miRNA levels. However, when miRNAs with lower abundances are considered (Figure 4, left side of graph), target gene expression varied greatly and rose to much higher values than those for the high-miRNA-abundance reads. These data support the hypothesis that miRNAs in general can mediate mRNA expression in a tissue. However, other factors are often involved in the control of gene expression. As shown previously, many of these miRNA genes are changing expression in response to heat stress,

and we are presently involved in a more detailed overall analysis and confirmation of the expression level in additional genotypes to identify specific relationships that may be of value in heat-stress improvement efforts.

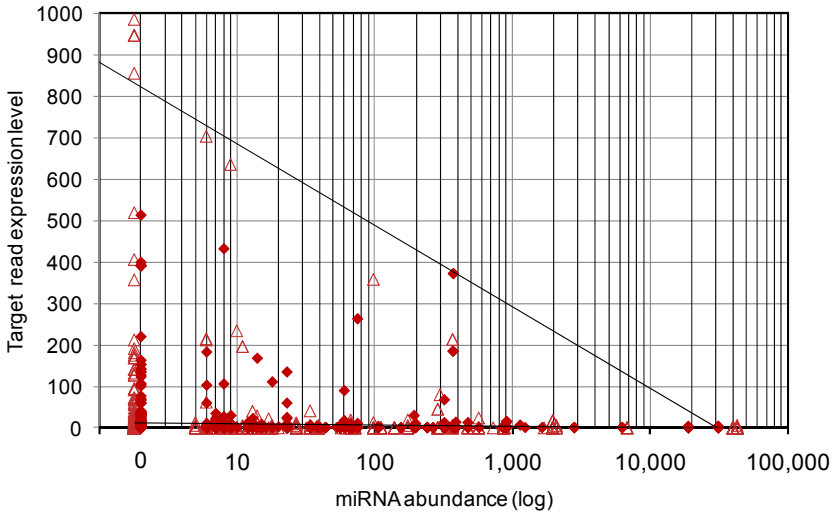

Figure 4. The steady state expression of target miRNAs is shown plotted against the miRNA abundance for the corresponding miRNA observed in the same total RNA samples.

Conclusions

1. We identified 280 mature miRNAs and 191 predicted pre-miRNAs in cotton using a pipeline specifically developed for species without a sequenced genome. This pipeline relies on the use of a modified version of the miRDeep program for predicting miRNA precursors.
2. The specific cotton miRNA precursors obtained using miRDeep have properties consistent with the precursors found in other plant species, and appear to be valid precursors, although specific validation remains to be completed.
3. Heat stress can lead to the up- or down regulation of specific sRNAs and miRNAs in significant ways. This likely has profound effects on mRNA stability and degradation during episodes of heat stress.
4. The heat-stress regulation of specific miRNAs in the leaves of more heat tolerant and less heat tolerant cotton genotypes can also be demonstrated. Further analysis is required to determine the significance and utility of such regulation can be fully appreciated.
5. It is difficult to infer the significance of heat stress regulation until a more complete complement of miRNAs and their targets have been established in cotton.
6. The prediction of miRNA targets (regulated genes) and their analysis suggests that steady-state levels of target mRNAs do not always vary in a way that suggests that miRNAs are the only factor regulating the expression of these genes. However, examples of apparent miRNA regulation of appropriate mRNAs can be observed.
7. The steady state level of target mRNAs for highly expressed miRNAs in cotton leaves are lower than the levels of target mRNAs for miRNAs that are expressed at low levels or not at all, but the this trend is variable and inconsistent at low miRNA expression levels. This suggests that factors other than miRNA expression play valid roles in temperature-dependent gene expression in cotton leaves. However, it is suggested that miRNAs may play an important role in cotton heat tolerance.

Acknowledgements

This work was supported by grants from the Alabama Cotton Commission, Cotton, Inc., the Alabama Agricultural Experiment Station, and the Alabama Land Grant Alliance.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Azhar, F.M., Z. Ali, M.M. Akhtar, A.A. Khan, and R. Trethowan. 2009. Genetic variability of heat tolerance, and its effect on yield and fibre quality traits in upland cotton (Gossypium hirsutum L.). Plant breeding 128:356-362.

Frederick, S.L. 2006. Thermotolerance in Gossypium hirsutum. M.S. Thesis, Auburn University, 69p.

Friedlander,M.R., W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, and N. Rajewsky. 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nature Biotechnol. 26:407-15.

Gardner, P.P., J. Daub, J.G. Tate, E.P. Nawrocki, D.L. Kolbe, S. Lindgreen, A.C. Wilkinson, R.D. Finn, S. Griffiths-Jones, S.R. Eddy and A. Bateman. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res. 37:D136-D140.

Griffiths-Jones, S, S. Moxon, M. Marshall, A. Khanna, S.R. Eddy, and A. Bateman. 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33:D121-D124.

Griffiths-Jones, S., R.J. Grocock, S. van Dongen, A. Bateman, A.J. Enright. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acid Res. 34:D140-D144.

Griffiths-Jones, S., H.K. Saini, S. van Dongen, A.J. Enright. 2008. miRBase: too.ls for microRNA genomics. Nucleic Acid Res. 36:D154-D158.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.1-R25.10.

Meisel, L., B. Fonseca, S. González, R. Baeza-Yates, V. Cambiazo, R. Campos, M. Gonzalez, A. Orellana, J. Retamales, and H. Silva. 2005. A Rapid and Efficient Method for Purifying High Quality Total RNA from Peaches (Prunus persica) for Functional Genomics Analyses. Biol. Res. 38:83-88.

Meyers, B.C., M.J. Axtell, B. Bartel, D.P. Bartel, D. Baulcombe, J.L. Bowman, X. Cao, J.C. Carrington, X. Chen, P.J. Green, S. Griffiths-Jones, S.E. Jacobsen, A.C. Mallory, R.A. Martienssen, R.S. Poethig, Y. Qi, H. Vaucheret, O. Voinnet, Y. Watanabe, D. Weigel, and J.K. Zhu. 2008. Criteria for annotation of plant MicroRNAs. Plant Cell 20: 3186-90.

Pant, S., N.K. Singh, D.B. Weaver, C. Weng, M. Zhong, and R.D. Locy. 2011. Changes in gene expression associated with heat stress in cotton as determined by deep sequencing of cDNA libraries. These Proceedings.

Parkhi, V., V. Kumar, G. Sunilkumar, LM. Campbell, N.K. Singh, and K.S. Rathore. 2009. Expression of apoplastically secreted tobacco osmotin in cotton confers drought tolerance. Molecular breeding. 23:625-639.

Pettigrew, W.T. 2004. Moisture deficit effects on cotton lint yield, yield components, and boll distribution. Agronomy Journal. 96:377-383.

Pettigrew, W.T. 2008. The Effect of Higher Temperatures on Cotton Lint Yield Production and Fiber Quality. Crop Science 48:278-285.

Rahman, M., I. Ullah, M. Ahsraf, J.M. Stewart, Y. Zafar. 2008. Genotypic variation for drought tolerance in cotton. Agronomy for Sustainable Development. 28:439-447.

Reddy, V.R., D.N. Baker, and H.F. Hodges. 1991. Temperature effect on cotton canopy growth, photosynthesis and respiration. Agron. J. 83:699–704.

Sridhar, J., G. Sowmiya, K. Sekar, and Z.A. Rafi. , PsRNA: a computing engine for the comparative identification of putative small RNA locations within intergenic regions. Genomics Proteomics Bioinformatics 8 127-34.

U.S. Department of Agriculture, Agricultural Research Service CottonDB at (www.cottondb.org)

Wan, C.Y., T.A. Wilkins. 1994. A Modified Hot Borate Method Significantly Enhances the Yield of High-Quality RNA from Cotton (*Gossypium hirsutum* L.). Analytical. Biochem. 223:7-12.

Xiaoqiu, H. and M. Anup. 1999. CAP3: A DNA Sequence Assembly Program. Genome Res. 9: 868-877.

Xie, F. and B. Zhang. 2010. Target-align: a tool for plant microRNA target identification. Bioinformatics 26:3002-3003.