CHANGES IN GENE EXPRESSION ASSOCIATED WITH HEAT STRESS IN COTTON AS DETERMINED BY DEEP SEQUENCING OF CDNA LIBRARIES

Shankar Pant
Narendra K. Singh
David Weaver
Chiachen Weng
Zhong Min
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 C for 4 or 24 hours (HS) or not heat-stressed (controls, grown at 25 C). cDNA libraries were developed from the 4 total RNA preparations, and these libraries were submitted for deep sequencing using a Roche 454-Flex sequencer at the ENGENCOR Sequencing facility at the University of South Carolina. An analysis of the results of this sequencing describing the expression level of genes that are up- or down- regulated by heat stress in these 2 genotypes of cotton is presented here. The specific genes up or down regulated have been clustered into groups of common functions, and subsequent analysis indicates that the number of genes differentially expressed during heat stress at 42 C in cotton leaves is larger than expected involving may physiological processes.

Introduction

Environmental stresses, particularly heat and drought stress, are major factors limiting cotton yield throughout the cotton belt in the US and worldwide (Ashar, et al., 2009; Parkhi, et al., 2009: Pedigrew, 2004, 2008; Rahman, 2008; Reddy et al., 1991). 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 (Van Esbroeck & Bowman,1998; Cole *et al.*, 2009), 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.

We have previously screened the 1780 accessions of the USDA Cotton Germplasm collection for heat tolerance using chlorophyll fluorescence and a growth chamber protocol that we developed for this purpose (Frederick, 2006). We have identified 7 accessions with apparently improved heat tolerance compared to the commercial variety DP90 that anecdotally was considered a heat and drought tolerant cultivar. We have made crosses with 5 of these accessions with DP90, and this year we advanced progeny to the F2 generation.

In order to adapt to changing environments, plants differentially express various genes under stressful conditions that presumably assist them with survival of a stressful environment. In this study we are attempting to identify those genes whose expression changes in response to heat stress (up- or down-regulated), and here we report our progress on the use of next generation DNA sequencing to establish heat-stress-mediated changes in the expression of various genes in cultivar DP90 and in accession TX337, one of the elite accession we identified from our germplasm search.

Materials and Methods

Plant Material and Stress Treatment

Four- to six-week-old greenhouse-grown plants of DP90 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).

Results and Discussion

Sequencing Results

The four data sets were combined, and a total of 242,683 primary reads amounting to over 7,000,000 bases (see Table 1) were obtained. These reads were assembled into leaf contigs using the CAP3 assembler (Huang & Madan, 1999). This yielded a total of 19,243 contigs which have been used for subsequent analysis. A total of 157 contaminating (prokaryotic and vector) sequences were removed, from the dataset, and 12,986 low abundance reads (found in only 1 of the datasets with readnumbers less than 5) were set aside for future use, leaving 6,099 contigs which we used as the primary leaf contigs set. Both the full (19,243 ESTs) and higher abundance (6,099) parts of the dataset have been used for various aspects of subsequent analysis.

Table 1. Results of the sequencing of the 4 cDNA libraries.

Dataset (Tissue and treatment)	Total Base pairs	Total Reads	Assembled Contigs
DP90 – control (25 C)	1,174,076	31,888	4,241
DP90 – Heat Stress	1,827,782	58,551	6,889
TX337 – control (25 C)	1,574,318	75,969	5,808
TX337 – Heat Stress	2,429,379	76,285	9,123
Combined Total	7,005,555	242,693	19,243

Heat Stress Regulates Gene Expression in Cotton Leaves

The overall expression of EST sequences in cotton leaves in response to heat stress is shown in Table 2. Note that nearly 32% and 22% of the contigs were up- and down-regulated, respectively in both the DP90 and TX337 genotypes. It is noteworthy that a relatively low percentage of the genes were mutually up-regulated (7.6%) or down-regulated (3.2%) in both genotypes. Although not apparent from Table 2, a rather large number of ESTs demonstrate up-regulation in one genotype while showing down-regulation in the alternative genotype. This is particularly apparent in the low abundance reads and may be related to legitimate genotypic differences in mRNA expression or to the apparently low quality of parts of the dataset, particularly in the low abundance reads. This is under further study and emphasizes the importance of expression verification of results obtained from sequencing necessitating cautious interpretation at this preliminary stage. However, the data do serve to establish a baseline for sequences expressed in cotton leaves in control and during heat stress, and we have prepared replicated samples that we will be further analyzing using quantitative PCR to obtain a clearer picture of gene expression during heat stress in cotton leaves.

Table 2. Overall expression of contigs from cotton leaves in response to heat stress.

Genotype/Sample	Up-regulated [contig number (%)]	Down-regulated [contig number (%)
DP90	1949 (32.0%)	1431 (23.5%)
TX337	2004 (32.9%)	1166 (19.1%)
Both	461 (7.6%)	195 (3.2%)
Total (6099)	3492 (57.3%)	2402 (39.4%)

Functional annotation

In order to assign function to the 6099 ESTs, BLASTX (Altschul et al., 1990) searches against all protein sequences in the GenBank nr database at NCBI (http://www.ncbi.nlm.nih.gov) were preformed, and the best hit for each sequence from this BLASTX search was retained. We also performed BLASTN (Megablast) (Altschul et al., 1990) searches against the USDA Cotton Database (www.cottondb.org) to obtain all EST sequences hit by our contigs. The best 15 hits from that search were assembled into the longest contig(s) possible using the CAP3 assembler (Huang & Madan, 1999), and these contigs were used to conduct BLASTX searches against the Viridiplantae protein sequences in the GenBank database at NCBI. The highest scoring informative hit that functionally identified the protein putatively coded for by the contigs was selected. (Note the highest scoring informative hit was the hit

with highest score that informed up of potential function and had a hit score not less than 80% of the highest scoring protein and no higher scoring informative protein was observed.) This table of hits was then used to manually cluster the 6099 ESTs, into 4 categories of regulatory gene contigs and 4 categories of effector gene contigs as well as undefined or unhit sequences. These 8 main groups were further sub divided into 22 sub categories according to the analysis below shown in Table 3.

Functional Categorization of contigs

Contigs with similar function were categorized into 22 sub categories as shown in Table 3. The functions were divided into 2 major groups for further investigation, i.e. regulatory-type genes, and genes coding for effector-type proteins. The effector proteins were considered proteins of an essentially non-regulatory nature that had a defined biochemical, transport, or cytoskeletal function. Note that of the 6099 high abundance contigs we worked with,

TABLE 3. Categorization of contigs into various functional groups

Functional category/group	Total	Category Total
REGULATORY GENE CONTIGS -		2124
A. Transcription related		772
1. Transcription factors	455	
2. Histones & other nuclear acting proteins	317	
B. Translation regulation		337
3. Ribosomal proteins	252	
4. Translation factors	85	
C. Signaling		770
5. Heat Shock Proteins/Factors	156	
6. Oxidative/redox signaling	121	
7. Signal Transduction – protein kinases, protein phosphatases, receptors	493	
D. Protein Degradation		245
8. Proteasome/protease	119	
9. Ubiquitin	126	
EFFECTOR PROTEIN CONTIGS –		1614
E. Metabolic Gene Contigs		1081
10. Cytosolic	768	
11. Chloroplast/Photosynthesis	240	
12. Mitochondrial/TCA Cycle/Respiration	73	
F. Membrane transport		293
G. Cytoskeleton – actin, myosin, tubulin		54
H. Cell wall		39
I. Stress-related Proteins		147
13. Cytochrome P-450-related	45	
14. Pathogenesis-related	15	
15. Osmotic stress-related	14	
16. Salinity stress-related	13	
17. Cold stress-related	7	
18. Heavy metal/metalothienin	18	
17. Ethylene/senescence/PCD	8	
20. Miscellaneous	27	
Other genes – (Function undefined)		2361
21. No hits to Viridiplantae in database	927	
22. Miscellaneous or function unknown	1434	
Total		6099

1434 of them hit proteins of undefined function, although there were sequences found in plants analogous to these contigs. There were an additional 927 contigs that did not hit any protein sequence in the GenBank or protein database. These sequences likely coded for other noncoding RNAs and will not be further considered here.

Contigs coding for regulatory proteins (2124) as a group constituted 36.3% of the regulatory contigs obtained that coded for transcription-related regulators such as transcription factors, DNA binding proteins, Histones, and other chromosomal proteins involved in regulation of gene expression and chromatin structure (see Table 3 and Figure 1).

An additional 36.2% of the regulatory contigs coded for signaling related genes. Note that in addition to protein kinases and phosphatases and receptors, we have included heat-shock -related and oxidative stress signaling proteins in this group as well. These proteins might be characterized as stress-related effector proteins but typically have both a stress-related and regulatory role. Translation and protein degradation-related signaling proteins account for the remaining 27.5% of the regulatory proteins.

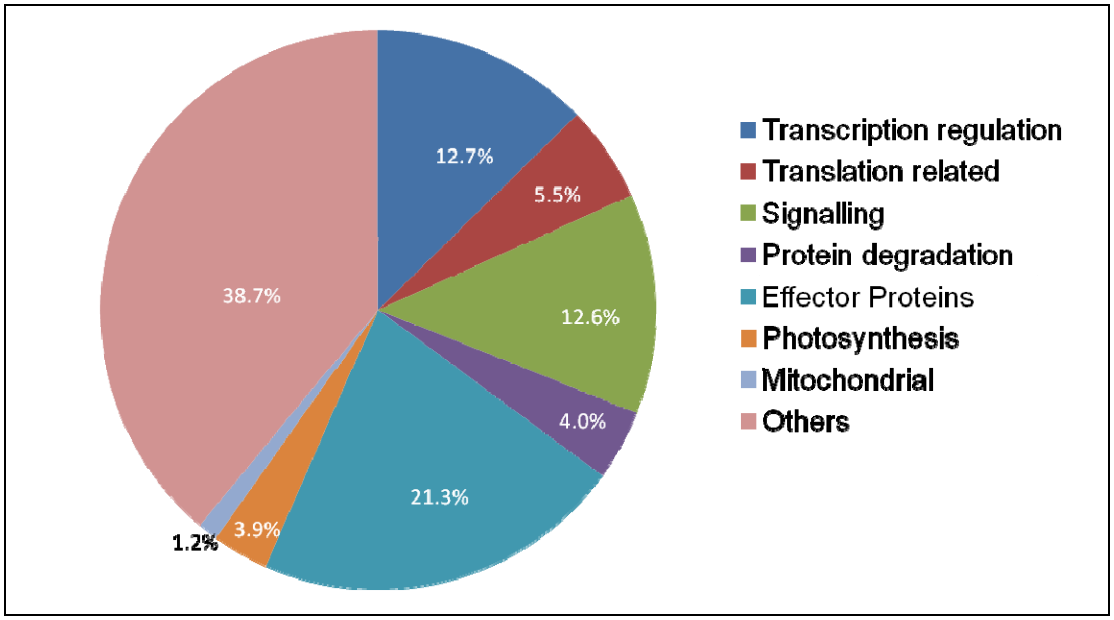

Figure 1. Pie Chart showing the percentage of processed contigs found in the clustered, grouped categories of Table 3. The percent of total contigs determined to cluster to into a group of contigs corresponding to Regulatory Genes (Table 3) including: transcription factors, translation-related genes, signaling genes, and genes involved in protein degradation is separately shown. Collectively these regulatory gene contigs account for 34.8% of the total contigs. Effector proteins found in the cytosol, chloroplasts and mitochondria account for a total of 26.4% of the proteins. The largest fraction of contigs correspond to the group of proteins consisting of hypothetical protein, proteins of unknown function, and proteins with no corresponding sequence in the NCBI GenBank database.

Proteins without a regulatory role assigned to metabolic functions including photosynthesis, and respiration, to membrane transport, to cell wall and cytoskeletal functions, and other stress-related proteins (see Table 3) make up the effector-protein group (Figure 1). The largest number of contigs overall are the metabolic group. They constitute approximately 2/3 of the effector protein contigs, and 17.7% of all contigs. The other significant effector protein functional groups are the membrane transport proteins (4.8% of all contigs) and the stress protein group (2.4% of all contigs).

Expression of cotton leaf ESTs during transient heat stress.

Using the readnumbers of each EST sequence as determined from the assembly of the reads into contigs, the expression of each contig was normalized to reads per 100,000 total reads in each of the 4 dataset, and EST expression was compared by sub category for both the less heat tolerant DP90 genotype and for the more heat tolerant TX337 accession. Figure 2, panel A shows that for all 9 subcategories of regulatory gene ESTs, expression is up-regulated to a greater extent in TX337 than in DP90. Transcription factors and signaling proteins are the two subgroups that contain the greatest number of up-regulated genes in both genotypes. It is also of note that heat-shock protein contigs are up-regulated to a greater extent in TX337 and down-regulated to a great extent in DP90. Ribosomal protein contigs and nuclear factor genes are most highly down-regulated in TX337 and DP90 respectively. In general for all categories of regulatory genes, a greater number of regulatory genes were up-regulated than were down-regulated by heat stress. However, the number of contigs showing up- or down-regulation is less than the number of contigs that did not change.

For effector genes (see Figure 3), more metabolic contigs were up-regulated in DP90 than in TX337, while the reverse was true for down-regulated genes. A greater number of photosynthesis-related metabolic contigs were also up-regulated in DP90 than in TX337. Other than these changes the number of up- and down-regulated contigs for the other effector genes tended to be similar for DP90 and TX337 and was consistent with the number of contigs in each category. The unknown protein contigs and unidentified contigs are also shown in Figure 3. There are a large number of contigs represented in these 2 categories and there appear to be significant differences between DP90 and TX337. It will apparently be important to identify more of these genes to determine the basis of these differences.

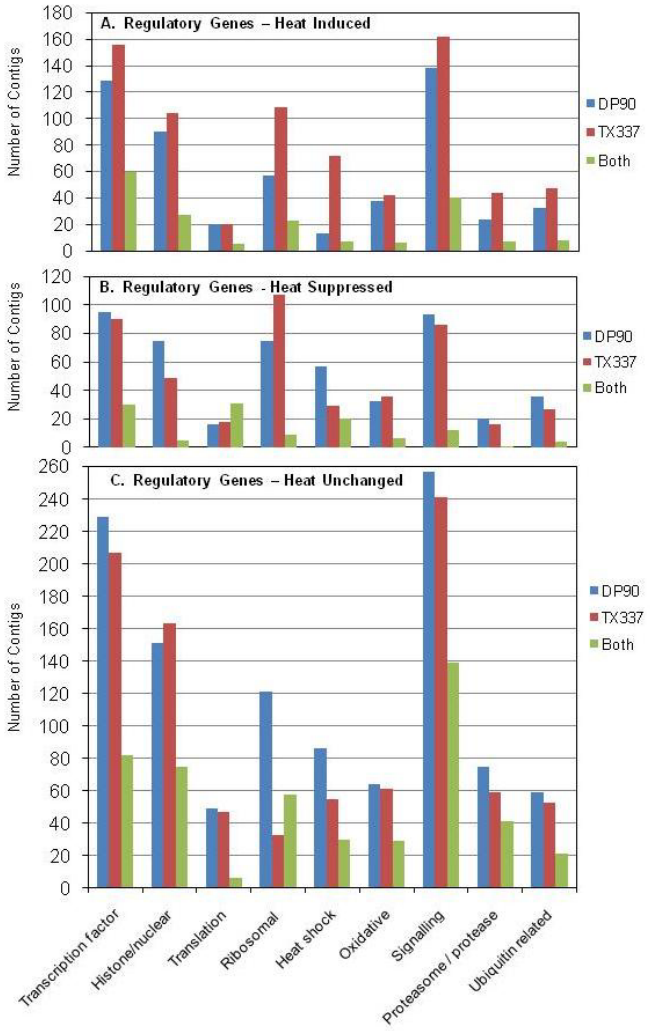

Figure 2. Panel A. The number of contigs up-regulated by heat stress in genotypes DP90 and TX337 or both in each of the regulatory gene categories is shown. Up-regulated means expression increases in the heat stress samples by greater than 4-fold. Panel B. The number of genes down-regulated by heat stress in genotypes DP90 and TX337 or both in each of the regulatory gene categories is shown. Down-regulated means expression decreases in the heat stress samples by greater than 4-fold. Panel C. The number of genes whose expression is unchanged by heat stress in genotypes DP90 and TX337 or both in each of the regulatory gene categories is shown.

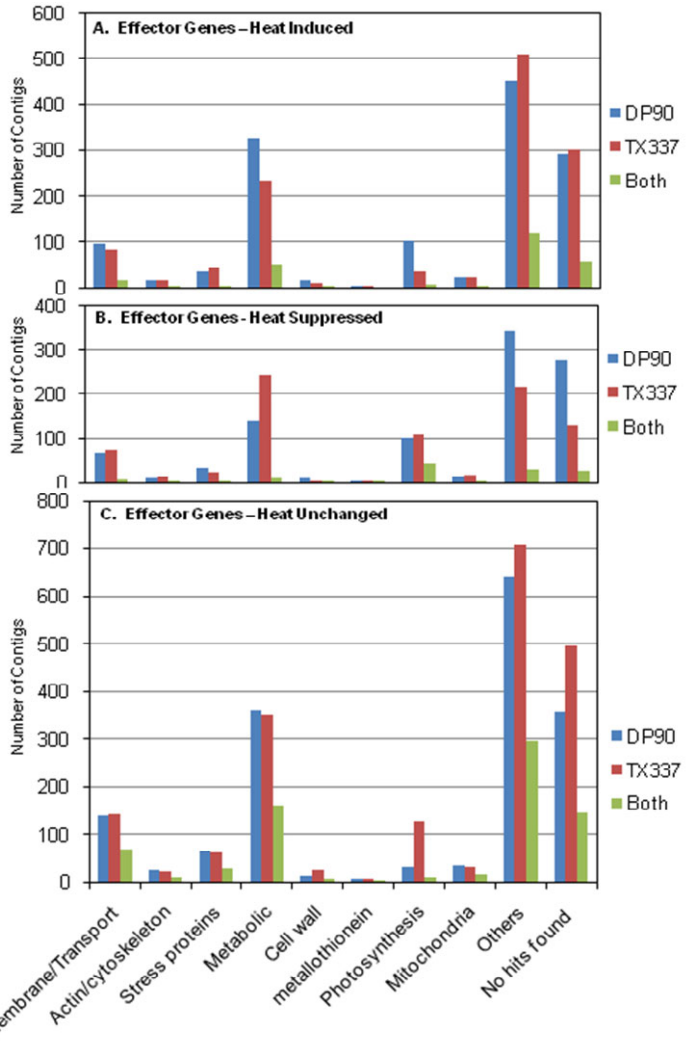

Figure 3. Panel A. The number of genes up-regulated by heat stress in genotypes DP90 and TX337 or both in each of the effector gene categories is shown. Up-regulated means expression increases in the heat stress samples by greater than 4-fold. Panel B. The number of genes down-regulated by heat stress in genotypes DP90 and TX337 or both in each of the effector gene categories is shown. Down-regulated means expression decreases in the heat stress samples by greater than 4-fold. Panel C. The number of genes whose expression is unchanged by heat stress in genotypes DP90 and TX337 or both in each of the effector gene categories is shown

Conclusions

1. Deep sequencing of RNA from cotton leaves during heat stress has generated a set of data that demonstrate changes in gene expression in a large number of genes during a 4 hour or 20 hour heat stress in cotton leaves.
2. The number of contigs sequences that demonstrated changes (up or down) in gene expression tended to be greater in the heat stressed, more thermotolerant accession (TX337), than in the more highly bred, domesticated variety (DP90).
3. The number of transcriptional and translational regulatory genes and signaling genes expressed in cotton leaves in response to heat is larger than expected, and can vary between genotypes and with heat stress.
4. Candidate individual genes whose expression during heat stress may be characteristic of the thermo tolerance of a genotype can be identified.
5. Expression of these genes is presently being verified by RT-PCR, and will be investigated in a wider range of thermo tolerance germplasm as well.

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.

Cole, C.B., D.T. Bowman, F.M. Bourland, W.D. Caldwell, B.T. Campbell, D.E. Fraser, and D.B. Weaver. 2009. Impact of Heterozygosity and Heterogeneity on Cotton Lint Yield Stability. Crop Science. 49:1577-1585.

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

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.

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.

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

Van Esbroeck, G. and D.T. Bowman. 1998. COTTON IMPROVEMENT: Cotton Germplasm Diversity and Its Importance to Cultivar Development. J. Cotton Sci. 2:121-129.

Wan CY, Wilkins TA. 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