Genome-wide transcriptome profiling of *Gossypium* spp. roots during early growth after infection with *Rotylenchulus reniformis*

Yonathan Tilahun*

Center for Molecular Biology, Alabama A&M University, Normal AL, 35762, USA

*Corresponding author: ytilahun@langston.edu

Abstract

Plant-nematode interactions have been studied extensively. The identification of genes expressed in root organs subsequent to nematode infection have been studied to a lower extent. Fewer still, potential resistance genes have been identified in cotton towards *Rotylenchulus reniformis* the reniform nematode (RN). Currently, there have been three cotton genomes completely sequenced. *Gossypium*, or cotton, genomes sequenced include: *Gossypium hirsutum* cv TM1, *Gossypium arboreum*, and *Gossypium raimondii*. This study imparts knowledge of differentially expressed genes (DEGs) after individual domesticated and wild cotton plants have been infected with approximately 50,000 juvenile RNs. After extraction of Total RNA from infected roots and subsequent next generation 454 pyrosequencing methods were employed, FastQC quality control measures were completed prior to Trimomatic-0.32 trimming from 33,788 reads, resulting in a significant decrease in the number of input reads (3454 input reads or DEGs). Putative descriptions were made for 634 input reads or DEGs, many of which were repeated multiple times as predicted proteins/enzymes or partials. The definition of 52 DEGs were made and enriched through the assignment of gene ontology (GO) terms that highlight categories of host plant genes for potential targets in future downstream transformation of cotton plants, especially those involved in strengthening systemic acquired resistance (SAR) or the hypersensitive response (HR) that occurs most often in plants when in defense against pathogen attack. This study reveals known protein/enzymes (52), unknown protein/enzymes (582) and unassigned DEGs (2820).

Keywords: *Gossypium hirsutum* cv.1218; *Gossypium hirsutum* cv. Coker; *Gossypium arboreum*; *Gossypium barbadense*; *Rotylenchulus reniformis*; cotton; reniform nematode; host-pathogen interaction; functional genomics; root transcriptome.

Abbreviations: GH1218, _G. hirsutum_ cv. – 1218; GHCoker, _G. hirsutum_ cv. – Coker; GA _G. arboreum_ A2 B7; GB _G. barbadense_ 713; cDNA_complementary Deoxyribonucleic acid; RN_reniform nematode; RNA_ribonucleic acid; TDFs_Transcript derived fragments; DEGs_Differentially expressed genes; BLAST_Basic Local Alignment Search Tool; B2GO_BLAST 2 Gene Ontology; EST_Expressed Sequence Tags; rRNA_ribosomal RNA; NBS_nucleotide binding site; SMART_Switching Mechanism At 5’ end of RNA Transcript; LD_Long distance; TAIR_The Arabidopsis Information Resource.

Introduction

Yield losses due to reniform nematode (RN, *Rotylenchulus reniformis*) make it the most damaging nematode pest of *Gossypium hirsutum* (Upland cotton) in the southeastern United States (Moore and Lawrence, 2012). Greater damage has been reported in the eastern half of the US cotton belt (Lawrence and McLean, 2001; Starr et al., 2005). Infestation by this pest has resulted in annual losses of approximately $130 M, with major impact in the states of Mississippi, Louisiana, and Alabama. Identification and integration of genetic resistance in cotton to RN is a more sustainable strategy but has not been fully attained in spite of recent exhaustive exploration (Li et al., 2015) and more work in diverse cotton cultivars is needed. This will result in more efficient marker-systems that will likely aid in development and release of nematode-resistant cotton cultivars with superior yield potential and high fiber quality. Thus far, commercial cultivars of cotton are susceptible to RNs (Usery et al., 2005; and Weaver et al., 2007; Sacks and Robinson, 2009). The young meristematic elongation zone of young tap and secondary roots are especially vulnerable areas of attack by RN. Genes that are expressed after nematode attack have been studied in several species in paired resistant/susceptible hosts. The *de novo* assembly of cotton transcriptome collected during the early stages of RN infection, (15 days post inoculation) was carried out. In this study, the formation of syncytia in cotton roots caused a reaction from the plant hosts. Urwin et al. (2000) identified that certain proteinase inhibitor genes impart resistance to reniform nematode, but such genes in cultivated and wild relatives of cotton have not yet been identified. Efforts continue to identify sources of resistance to RN in *Gossypium* (Muhammad and Jones, 1990; Robinson et al., 2004) and previously assumed resistant or tolerant germplasm stocks have been released (Nelson, 2009; Jones et al., 1988; Cook et al., 1997a; Cook et al., 1997b). However, the only known completely tolerant species is *G. longicalyx*. Other possible strategies have been made to introgress resistance genes of diploid species, such as *Gossypium arboreum*, into cultivated tetraploid cotton (Bell and Robinson, 2004; Robinson et al., 2004; Avila et al., 2003; Agudelo et al., 2005; Robinson et. al., 2007; Dighe et al., 2001). Among these, two genotypes of Upland cotton have been identified with useful levels of resistance to RN,
the BC\textsubscript{2} lines LONREN-1 and -2 (Robinson et al., 2007). Additionally, \textit{G. barbadense}, have also been an exception with some accessions having the ability to suppress RN populations by 70-90\% (Robinson et al., 2004; Starr et al., 2007). The leading cultivars of cotton, with no known resistance to RN, account for 90\% of the world’s annual cotton crop, underscoring the need for intensive exploration of genes. Through differentially expressed genes (DEGs) and comparison of tolerant versus susceptible species a valid approach for identification of genes, expressed transcripts or markers that play a key role in imparting tolerance are possible (Avila et al., 2003; Robinson et al., 2007). The genomes of three species have been sequenced (\textit{Gossypium hirsutum} cv TM1 (Udall et al., 2006; Chen et al., 2007; Li et al., 2015); \textit{Gossypium arboreum} (Li et al., 2014); and \textit{Gossypium raimondii} (Wang et al., 2012). The information provided by the sequenced genomes aids in cotton trait improvement, but the nature of genes that regulate resistance to RN remains elusive. This study aims to identify the existing variation using DEG profiles of four genotypes. They represent three cotton species (\textit{G. hirsutum}, \textit{G. arboreum}, and \textit{G. barbadense}) and two genotypes from the same species (\textit{G. hirsutum}) associated with RN infection. The goals were to: (1) Identify DEGs associated with the root transcriptome in cotton during early stages of infection; and (2) Determine DEGs that are shared among susceptible cultivars \textit{G. hirsutum} (\textit{G. hirsutum} cv. 1218 (GH1218), \textit{G. hirsutum} cv. Coker (GHCoker)), while also identifying DEGs in comparison to the relatively more tolerant genotypes of \textit{G. arboreum} cv. - A2 87 (GA) and \textit{G. barbadense} cv. - GB713 (GB) to RN infection.

Results

Differentially expressed genes (DEGs) after 15 days of RN infection

Approximately 9.5 Mb of raw sequences were generated for four cDNA libraries. This study investigated DEGs with a length of 30 nucleotides (nt) and higher of GH1218, GHCoker, GA, and GB (Table 1). Collectively this study identified 33,788 DEGs after quality control using FasQC of 33,798 DEG input reads. The application of Trimomatic-0.32 trimmed adaptor and poor quality sub-sequences from the reads eliminated 30,334 (Tables 1 and Table 2). The remaining sequences of 3,454 DEGS across all four libraries were isolated and used for further analysis. Trimomatic-0.32 (Bolger et al., 2014) provided the resulting numbers for GH1218 (662), GHCoker (752), GA (1034), and for GB (1006) (Table 1 and Table 2). The percentage of GH1218 that had BLAST hits equaled 39\% (261/662), 11\% (80/752) for GHCoker, 14\% (149/1034) for GA and 10\% (98/1006) for GB. The identified functional classifications for each library were assigned for the known genes obtained through the use of BLAST in the application Blast2GO Pro (B2G) (Conesa et al., 2005; Conesa and Götz, 2008; Götz et al., 2008; and Götz et al., 2011). There were a total of 582 individual DEG descriptions that were identified as having putative homology to known sequences for protein/enzymes using B2G (Supplementary Table 1). Finally, 52 annotation descriptions were identified and the specific DEGs belonging to each genotype were defined as annotated gene ontology (GO) terms using B2G (Supplementary Table 1). The numbers of DEGs that did not share homology (2,820) were not considered for further analysis (Table 2). There were 261 DEGs identified by BLAST and map hits as putative genes for GH1218 that revealed 6 individual descriptions for proteins/enzymes. GHCoker had 12 individual descriptions of proteins/enzymes with a total of 92 DEGs with BLAST and map hits. Sixteen individual descriptions were revealed for proteins/enzymes from a total of 165 DEGs that were derived from GA BLAST and map hits. The following are all shown in Table 2: 1) there were 98 DEGS with BLAST and map hits for GB that revealed 18 individual descriptions for proteins/enzymes; 2) four hundred-one DEGs were undefined by functional annotations for GH1218 in this study; six hundred sixty DEGs were undefined by functional annotation for GHCoker; eight hundred sixty-nine DEGs remained undefined by functional annotation for the GA cDNA library; finally, 890 DEGs were undefined functional annotations of GB (Table 2). The functions of the GH1218, GHCoker, GA, and GB TDFs revealed that the DEGs segregated to several categories. Response to stress was the most common function identified in all but the susceptible type GH1218 genotype after annotation. A comparison between Susceptible and Tolerant libraries, indicate that there are numerous target regions of DEGs that are associated with differences in response to stress induced by reniform nematode infection. Although the number of BLAST and map hits are similar for the DEGs associated with the Susceptible libraries (GH1218 and GHCoker), when compared to Tolerant (GA and GB), twice the number of segregation occurred for Tolerant DEGs than Susceptible in annotation. The comparison of DEGs in this study was conducted following the completion of the \textit{Gossypium hirsutum} (Chen et al., 2007; Li et. Al, 2015); \textit{Gossypium arboreum} (Li et. al., 2014), and \textit{Gossypium raimondii} (Wang et. Al., 2012) genomes, therefore, they were among the references to reveal gene homology.

In summary, this study undertook comparative analysis of cotton-nematode interaction between RN infection of GH1218, GHCoker, GB, and GA. It characterized the root transcriptome and the DEG pattern among four genotypes of cotton. GO analysis identified gene ontology terms that are in the set of genes expressed from membranes and the plasma membranes, and responses to water and stress. Thus, asking the question, what is the role that the infection by reniform nematode plays during the expression of these genes?

Discussion

Nematode infection induces wounding, especially during establishment of an appropriate feeding site. Gene expression changes have been documented with wound or defense responses (Gayhson and Fenoll, 2002; Williamson and Kumar, 2006). These responses are present in both compatible (Susceptible) and incompatible (Tolerant) reactions. Supplementary Table 1 reflects a good depiction of what may be occurring. Therefore, a possible explanation will be made utilizing the DEGs depicted in Supplementary Table 1. Response to stresses including water stress is highlighted in each but one of the genotypes under the GO Names list. This seems to indicate that all the species have the molecular signaling system that identifies stress, whether biotic or abiotic. Further analysis identifies that the susceptible genotypes (GH1218 and GHCoker) share the GO Names list identifiers of “mitochondrion” and “integral component of membrane.” The “integral component of membrane” is also identified for both Tolerant (GA and GB) genotypes. Integral components may be any number of numerous protein complexes that are imbedded partially or completely in the cellular membranes of the cells composing the root tissue of plants. Hydrolase activity was identified for the Susceptible GH1218 library (Supplementary Table 1). Hydrolase activity is known to be associated with diverse
processes and mechanisms of plant defense, starch metabolism and cell wall remodeling against pathogen attack (Garcia-Garrido et al., 2005; Grienenberger et al., 2010; Tyler et al., 2010). UDP-glucosyltransferase activity was also identified, in this case the Susceptible GHCoker library (Supplementary Table 1). UDP-glucosyltransferase activity is known as a plant enzyme that typically is used by cells in the transfer of reactions involving biotic or abiotic stress and defense responses in addition to xenobiotics such as herbicides (Li et al., 2001; Liu et al., 2015). The next components under the GO name list is mainly located in roots rather than in leaves, the inosine catabolic processes (Deng and Ashihara, 2010). Inosine is usually found in RNA and is included among degradation products that can be recycled for nucleotide synthesis. Plant ATP binding cassette (ABC) transporters move various substrates, using energy obtained by ATP hydrolysis, and are involved in several physiological functions, including detoxification of xenobiotics in addition to transport of hormones and secondary metabolites (Verrier et al., 2008; Yazaki et al., 2009; Kang et al., 2011; Shitan et al., 2015). They tend to also accumulate in high concentrations in root tissue (Shitan et al., 2015) compared to leaf or stem tissue. The DEG annotations GO name F: protein serine/threonine kinase activity refers to a receptor that interacts with other proteins. Wide ranges of processes are affected, including disease resistance to develop self- versus non-self-recognition (Goring and Walker, 2004; Afzal et al., 2008). Protein phosphorylation is among the GO names list that was revealed in a GHcoker (Supplementary Table 1). Mitogen-activated protein kinases (MAPKs) are the main components of protein phosphorylation where a cascade of extracellular signals is transduced to respond to pathogen attack (Sidonskaya et al., 2015).

The response to biotic and/or abiotic stress, described thus far, is associated with signal transduction or transport of reactions. The question; however, is what determines tolerance or host resistance? Although a comprehensive conclusion cannot be drawn from this study, the proposal is that the difference between susceptibility and tolerance/resistance is based on the ability of the cells of the plant to mount a reaction that results in programmed cell death (PCD). The difference that is observed between GH1218, GHcoker, GA, and GB could be the inability for GH1218 and GHcoker to mount a defense by the activation of PCD as effectively as GA and GB. How do GA and GB cause this to happen? It may be the activation of GA defense through elevating levels of heavy metals that increase generation of reactive oxygen species (ROS), such as superoxide free radicals (O$_2^•$), hydroxyl free radicals (OH$^•$), or non-free radical species such as singlet oxygen (O$_2^*$) and hydrogen peroxide (H$_2$O$_2$), in addition to cytotoxic compounds like methylglyoxal (MG). All of these cause oxidative stress by disturbing homeostasis through the disruption of pro- and anti-oxidants presence within the plant cells (Zengin and Munzuroglu, 2005; Hossain et al., 2012; Sytar et al., 2013; Emamverdian et al., 2015). Thus causing multiple deteriorative disorders among others; such as oxidation of protein and lipids, ion leakage, oxidative DNA attack, redox imbalance, and denaturation of cell structure and membrane. Ultimately, according to GO annotations the presence of metal ion binding results in the activation of programmed cell death (PCD) pathways (Rellán-Alvarez et al., 2006; Flora et al., 2008; Hatata and Abdel-Aal, 2008; Nagajyoti et al., 2010; Rascio and Navari-Izzo, 2011; Sharma et al., 2012). The defensive methods that GB uses are most likely similar to GA, in quicker detection, response, and in the mounting of defense through the PCD pathway similar to the use of the methods just discussed in addition to nucleic acid-binding proteins that positively and quantitatively regulate cotton resistance to reinfomnematode (Deng et al., 2015).

### Materials and Methods

#### Plant materials

#### Greenhouse bioassay

Genotypes were selected based on published reports of tolerance and susceptibility to RN and include tolerant species GA and GB (Udall et al., 2006); and two susceptible cultivars of GH1218 and GHcoker (GRIN: http://www.ars-grin.gov/). Seeds were germinated on moist germination paper in two days, and transplanted one per 150 cm$^3$ Conetainer® (Stuewe & Sons, Tangent, Oregon, USA) in autoclaved 1:1 clay:sand mixture. Six replicates were arranged randomly in a greenhouse with 14 h of light daylight and temperature maintained at 30°C at the Plant Science Research Center at Auburn University (Auburn, AL). Cultures of RN were procured by collecting samples of soil from various infested field locations throughout Alabama.
RN populations were increased in soil in greenhouse pot cultures with the host *Gossypium hirsutum* cotton cultivar PM 1218 BR. Sixty days later nematodes were extracted from the soil by combined gravity screening and sucrose centrifugal flotation (Jenkins, 1964). Each Container containing individual experimental susceptible or tolerant cotton plant was infected 10d after transplanting with approximately 50,000 vermiciform RN and additional reniform eggs present in the collected solution (~2000-3000 eggs), by injecting the nematode suspension with a 1ml pipette near the base of the plant. The root transcriptome of treated cotton were characterized to determine DEGs. Four cDNA libraries were constructed from root tissues 15 days post inoculation (DPI) with 50,000 vermiciform reniform nematodes per plant) for GH1218, GHCoker, GA, and GB. Test plants to confirm presence of female reniform nematodes with active synctia were included in the six replicates were also infected in similar fashion. On 15 DPI test plants were observed using a stereo-microscope after staining with aceticarmine. On 15 DPI roots were removed, washed with tap water immediately, flash-frozen in liquid nitrogen, placed in 50 ml conical tubes, and stored at -80°C.

**Total RNA extraction for infected and non-infected plants**

Total RNA was isolated from root tissue (3 g GH1218, 2 g GHCoker, 1 g GA, and 2 g GB) using a modified protocol (Wan and Wilkins, 1994). The quality and concentration of the RNA was assessed using Experion™ RNA StdSens Analysis Kit (Bio-Rad, Hercules, CA, USA) using a 12K DNA chip and Nanodrop 100 Spectrophotometer (NanoDrop, Wilmington, DE, USA). Electrophoresis of RNA extracted from RN-treated root tissue of each cotton genotype was conducted on the Experion™ (Bio-Rad Laboratories, Inc., Hercules, CA) electrophoresis unit, quality being assessed by both band and peak appearance. The resulting concentrations of total RNA obtained were: GH1218 (15.36 ng/µl), GHCoker (16.45 ng/µl), GA (13.35 ng/µl), and GB (15.18 ng/µl). Total RNA of the RN infected roots of GHCoker, GA, and GB were utilized to construct cDNA libraries using a SMART™ cDNA Library Construction kit (Clonetech, Mountain View, CA, USA) according to manufacturer protocol (www.clontech.com). These cDNA libraries were initially evaluated for their quality by gel electrophoresis. Similarly, they were also screened for their efficiency by isolating cDNA fragments from gel electrophoresis (QIAGenQuick Gel Extraction kit, Qiagen, Valencia, CA), followed by cloning (TOPO TA cloning kit, Invitrogen, Carlsbad, CA) and plasmid isolation (QIAGen Miniprep isolation Kit, Qiagen,Valencia,CA) and by sequencing methods using an ABI 3100 sequencer (Applied Biosystems, Foster City, CA). Sequence information derived from the screened samples reinforced the quality of the methods used through the identification of homologous sequences matching *Gossypium hirsutum*, *Gossypium arboreum*, and *Gossypium barbadense* verified by BLASTN analysis at the National Center for Biotechnology Information (NCBI:http://www.ncbi.nlm.nih.gov).

**cDNA Library preparation for 454 pyrosequencing and Transcriptome sequencing**

A cDNA library was prepared at concentrations of 7.5 µg in a 210 µL volume for GH1218, 5.2 µg in 201 µL volume for GHCoker, 4.4 µg in a 190 µL volume for GA, and 3.4 µg in a 281 µL volume for GB. Libraries were amplified with Long Distance PCR (LD PCR1 and LD PCR2) in order to achieve high yields of cDNA (4-5 µg). Amplified samples were purified using GenElute™ PCR Clean-Up kit (SIGMA-ALDRICH, St. Louis, MO) and electrophoresed on a 1% agarose gel to ensure that appropriate fragment size ranges were obtained. Concentrations of the cDNA libraries were assessed using a Nanodrop spectrophotometer. The final cDNA libraries were sent to the Advanced Center for Genomic Technologies (ACGT) at the University of Oklahoma (Norman, OK) for sequencing using a 454/Roche Genome Sequencer FLX (GS20 FLX) instrument (454 Life Sciences, Branford,CT). Over 9.5 Mb of raw sequence data was generated from the four cDNA libraries: GH1218, GHCiker, GA, and GB. Samples were analyzed at The Advanced Center for Genome Technology (ACGT) at the University of Oklahoma (Norman, OK) using Newbler (454 Life Sciences, Branford, CT), software that accompanied the 454 pyrosequencing instrument. Read quality was assessed further with FastQC (Andrews, 2010), followed by adaptor trimming with Trimmomatic-0.32 (Bolger et al., 2014).

**Sequence assembly and data analysis**

The flowgrams generated by a 454 GS 20 FLX instrument were read and produced raw reads using Newbler (454 Life Sciences, Branford, CT). Raw sequence data were quality controlled using FastQC and trimmed using Trimmomatic-0.32 for the following sequences: >Key_plus_tag TCAAGCAGTGGCCT; >SMART_IV AACAGTGGTT- ATCAACGCAGAGTGGCATCAG GCGGGG; >CDS_III ATTCGTAGGCCGAGGCGCGCCGACATG-TTTTTTTTTTTTT; >5prime_PCR_Primer AACAGTGGG-GTATCAGCGAGATC; and >5prime_PCR_Primer_frame AAGCGATGTGTAT. As these were initially sequences formatted as .fna and .qual or .sff, they were converted to .fastq formats through utilizing bioinformatics conversion software suites on iPlant Discovery Environment (Goff et al., 2011).

The identified putative functional classifications for each library were assigned for the known genes obtained through the use of BLAST in the application Blast2GO Pro (B2G) (Conesa et al., 2005; Conesa and Gótz, 2008; Gótz et al., 2008; and Gótz et al., 2011). The reference genome sequence for all sample sequences was *Gossypium hirsutum*. Blast2GO was then utilized to identify putative and functional classifications of the homologous sequences. Blast2GO was further utilized for subtraction hybridization to identify unique sequences from each sample genotypic data. The cDNA libraries were constructed from the RN infected root tissues. Therefore it is understood that RN associated sequences are likely integral to the raw data. Further excision of RN associate sequences indicated that pathogen DEGs existed among the cDNA libraries.

**Conclusion**

This study examined four cotton genotypes, 15 DPI, providing insight into cotton–reniform interactions. Comprehensive gene expression profiles of syncytial development significantly will advance our understanding of plant resistance to RN. This study provides many possible targets for full-length cDNA marker development for molecular based RN resistance strategies in cotton utilizing natural mechanisms of resistance. The mechanism of plant response to RN has practical significance for nematode control through the development of future resistant crop varieties. Also, this work provides a useful resource for the cotton research community by providing a DEG repository...
for further investigations. The defined DEGs can be used to strengthen additional downstream studies by providing iterative genes that may be utilized for transformation of plants. Furthermore, as more defined functions become apparent, the data accumulated in this study will strengthen the breadth of applications that will be made towards combating RN in cotton.

Acknowledgements

Acknowledgements go to the USDA-CSREES (2004-38814-15160, ALAX 011-206, ALAX 011-706), NSF Grant: 0703470, USDADE – Title III – HBGI for providing financial support, and Auburn University, Langston University, and Alabama A&M University for providing the property, and nurturing environment for such research to be completed.

References


Gossypium hirsutum


