7. SUMMARY

Abiotic stresses such as drought, salinity, heat and cold are controlled by quantitative traits. Several metabolic pathways and signaling molecules are involved in stress tolerance. Plants respond to their changing environment in a complex, integrated way that allows them to react to the specific set of conditions. Tolerance or resistance of drought stress in plant is not only very complex, but is also highly influenced by other environmental factors and by the developmental stages of the plant. The physiological responses of plants to water deficit includes, leaf wilting, reduction in leaf area, leaf abscission, stimulation of root growth by directing nutrients to the underground parts of the plants and molecular responses to avoid the damages to cell. Plants are more susceptible to drought during flowering and seed development, as plant’s resources are deviated to support root growth. These physiological changes in plants affect the yield of crops.

In the present investigation, an attempt was made to identify the drought responsive genes in pigeonpea, expressional quantification and their molecular interactions. The in silico soybean cDNA libraries from Unigene database of NCBI have been analyzed for identification of differentially expressed genes (DEGs) under drought stress using IDEG6 web statistical tool. The IDEG6 analysis of five normal leaf libraries and a drought stressed leaf library revealed a total of 11 up-regulated and 41 down-regulated genes. All the genes obtained from in silico study were further annotated for KEGG pathway database, GO annotation and protein-protein interaction studies (PPIs).

Pathway annotation of 52 genes in KEGG database revealed the involvement of 32 genes in signaling pathways of which 42% (13 genes) are related to photosynthesis, 29% (9 genes) to light harvesting/antenna molecules, 10% (4 genes) to carbon metabolism and remaining genes to miscellaneous pathways. In silico soybean leaf cDNA library analysis between drought stressed and unstressed has shown significant decrease in efficiency with respect to photosynthesis, light harvesting and carbon fixation pathways which ultimately reduces the crop yield.

AmiGO application of GO database was used to categorize DEGs into three main attributes as biological process, cellular component and molecular function based upon their unique identifiers. The biological process category
represents 62.5% of genes related to cellular process followed by 16.7% for response to stimulus. The cellular component category represents 42.9% of genes related to macromolecular complex followed by 30.6% in other intracellular organelles. The molecular function category represents major molecules to binding activity (64.3%).

The STRING v9.1 database and Cytoscape v2.8.2 software was used for protein-protein interaction studies. Four out of 11 up-regulated DEGs showed protein-protein interactions in STRING v9.1 database, namely, ADF3 (medium to highest confidence with scores 0.606 to 0.908), UGE5 (high and highest confidence with scores 0.911 to 0.996), APB (medium and high confidence with scores 0.434 to 0.648) and transcribed locus (Gma.7880) (medium and high confidence with scores 0.613 to 0.873). The ADF3 and UGE5 genes showed more interacting partners among the four up-regulated genes. All down-regulated differentially expressed genes showed protein-protein interactions except unigene Gma.11215. Photosynthesis and antenna molecule related proteins were associated with good protein interactions and showed highest confidence level (more than 0.900 score).

The PPIs obtained in STRING database were subjected to Cytoscape software to find out the hub proteins in 44 DEGs which revealed three major network modules. The major network consists of glyceraldehyde 3-phosphate dehydrogenase-A subunit 2 (GAPA-2) (Gma.31588) gene and other two modules consists less number of interactions. In total, the network has 45 bottleneck proteins with 137 nodes and 837 edges.

A whole genome-wide in silico survey of WRKY and NAC genes was conducted in pigeonpea. HMM profiles of WRKY (PF03106) and NAC (PF02365) domains downloaded from Protein family (Pfam; http://pfam.sanger.ac.uk/) was exploited against the amino acid sequences of pigeonpea draft genome using command line HMMER v3.0 (http://hmmr.janelia.org/). The result yielded a total of 97 WRKY and 96 NAC genes in pigeonpea draft genome.

Phylogenetic analysis of CcWRKY protein sequences clustered into 9 subfamilies and CcNAC protein sequences into 11 subfamilies using Neighbor-Joining (NJ) method of MEGA5. The phylogenetic classification of soybean and pigeonpea WRKY and NAC amino acid sequences formed eight sub-groups each. Also, the WRKY and NAC members from Arabidopsis and rice along with pigeonpea members were phylogenetically categorized. Inspection of the phylogenetic tree topology of WRKY or NAC proteins
revealed several pairs with a high degree of homology in the terminal nodes of each subfamily which helps in phylogeny based function prediction. Putative conserved motifs were predicted in CcWRKY and CcNAC proteins using MEME tool. Exon-intron and domain structures were displayed using GSDS. The individual CcWRKY or CcNAC genes illustrated by GSDS indicated similar pattern of exon-intron and domain organization in the end nodes of clusters formed by the phylogenetic analysis.

Homologous sequences of six up-regulated DEGs namely, ADF3, APB, ASR, DLP, LTP1 and UGE5 and, six CcNAC genes namely, CcNAC01, CcNAC27, CcNAC30, CcNAC72, CcNAC86.1 and CcNAC92 were used for primer synthesis and qRT-PCR was performed for their response to drought stress in pigeonpea leaf and root samples. The Asha (ICPL 87119) variety of pigeonpea was selected for the experiment and plants were raised in pots containing 2:1 proportion of coarse sand and clay, maintained under greenhouse condition. After 20 days of germination, water was withheld to induce water stress condition in the testing plants while the control pots were irrigated normally. The leaf RWC was monitored at every alternative day until it reached to 60% with visual stress symptoms in plants and harvested the tissue samples.

The qRT-PCR result revealed significant up-regulation of DLP (5.02 log2 fold) and down-regulation of APB (9.43 log2 fold) and LTP1 (18.81 log2 fold) in pigeonpea water stressed leaf sample compared to well watered leaf sample. No significant difference was observed in stressed root compared to the stressed leaf samples of pigeonpea except APB showing an up-regulation of 11.35 log2 fold change. In CcNAC genes, a significant up-regulation of CcNAC27 (log2 2.22 folds), CcNAC30 (log2 2.76 folds) and down-regulation of CcNAC30 (log2 11.64 folds) genes was observed in water stressed leaf sample compared to well watered leaf sample. An up-regulation of CcNAC01 gene (log2 0.24 folds) and down-regulation of remaining five genes was observed in water stressed root sample compared to water stressed leaf sample. Among the five down-regulated genes, CcNAC27 was found to be more repressed (log2 13.98 folds) in root sample.

The amino acid sequences of CcWRKY28 and CcNAC67 genes were subjected for protein modelling using Robetta server and molecular simulation was performed. The Ramachandran plot of CcWRKY28 protein showed 89.0% residues in most favoured regions, 10.5% residues in additional allowed regions, 0.3% residues in generously allowed regions and
0.3% residues in disallowed regions validated in PROCHECK of SAVES server. The PDBsum predicted secondary structure of modelled CcWRKY28 protein and revealed the presence of 3 sheets, 4 β-hairpins, 3 β-bulges, 9 strands, 8 helices, 2 helix-helix interacts, 45 β-turns and 7 γ-turns. The ProSA energy profile of modelled CcWRKY28 protein was calculated as Z score -5.13 showing negative values and average fluctuations revealed that the modelled protein is of comparable quality.

Similarly, the Ramachandran plot of modelled CcNAC67 protein showed 87.1% residues in most favoured regions, 11.5% residues in additional allowed regions, 0.7% residues in generously allowed regions and 0.7% residues in disallowed regions. The CcNAC67 protein revealed the presence of 3 sheets, 7 β-hairpins, 3 β-bulges, 13 strands, 6 helices, 1 helix-helix interact, 49 β-turns and 7 γ-turns. The ProSA energy profile of CcNAC67 protein was calculated as Z score -5.38 showing negative value and average fluctuations in plot of residue scores indicated that the modelled protein is of good quality.

The built CcWRKY28 and CcNAC67 protein models from Robetta server was performed for molecular dynamics (MD) study to compare the stability and to predict the flexibility of the protein structures. The MD and energy minimization were executed by Gromacs v. 4.5.5.

The RMSD (Root Mean Square Deviation) of all Ca atoms of CcWRKY28 proteins gradually increases pretty rapidly from 0.5 Å around 3.5 to 4.5 ns in the first part of the simulation and then reached a plateau, indicating equilibrium status of the protein. The model maintained high structural conservation within the range of 2500–3500 ps as observed by plotting RMSD values.

Similarly, the molecular dynamic simulation of CcNAC67 showed that the RMSD of all Ca atoms gradually increases pretty rapidly from 0.3 Å in the first part of the simulation around 3.8 to 5.0 ns and then reached a plateau, indicating the equilibrium status of the protein. The model maintained high structural conservation within the range of 3500–4500 ps as observed by plotting RMSD values as a function of the simulation time.

The modelled CcWRKY28 and CcNAC67 proteins were then subjected to docking analysis using Hex server with default parameters for protein-DNA interaction. The C-terminal domain of CcWRKY28 protein having WRKYQGK β-sheet residues (TRP-328, ARG-329, LYS-330, TYR-331, GLY-332, GLN-333, LYS-334) could be responsible for binding to 5'-TTGACC-3'
DNA of PDB structure 2LEX. In CcNAC67 protein, residues 102-112 forming turn, 285-295 forming β-sheet and 296-305 forming turn could be responsible for binding to 5'-TTGGAACACGC-3' DNA of PDB structure 3SWM.

Structural alignments of modelled proteins with PDB structures were carried out using PyMOL for their similarity. Five stranded anti-parallel β-sheets of C-terminal domain of CcWRKY28 are aligning with four stranded anti-parallel β-sheet domain of PDB structure 2LEX with RMS value of 1.703. Superimposition of three stranded anti-parallel β-sheets of CcNAC67 is aligning with domain region of PDB structure 3SWM chain 'A' which resulted the RMS value of 0.793. The alignment between modelled CcWRKY28 and CcNAC67 proteins showed very less multiple sequence alignment and structural homology. Though structural similarities are reported between WRKY and NAC domains, modelled proteins showed less structural similarity, however β-sheets of both proteins aligning at the same angle with RMS value of 21.855 revealed a similar arrangement in β-sheets in both the proteins.

The WRKY and NAC superfamily of transcription factors are one of the multigene families of transcriptional regulators in plants. The 2.0 kb regulatory region upstream of the coding or promoter regions of pigeonpea genes were searched for the presence of W-Box [(C/T)TGAC(T/C)] and NAC [CATGTG] elements obtained from PLACE database to know the binding sites. It revealed the presence of good number of W-Box [(C/T)TGAC(T/C)] and NAC [CATGTG] elements when subjected for BLAST analysis using local BLAST search in BioEdit tool. The presence of W-Box [(C/T)TGAC(T/C)] elements showed more number of hits compared to NAC [CATGTG] elements indicating the presence of more W-BOX elements on cis-regulatory elements of upstream regions of the pigeonpea genes.

The abiotic stress related genes, transcription factors and their regulatory networks in pigeonpea have immense importance, which can be further analyzed in detail and extrapolated for crop improvement.