



## Effect of *Mesorhizobium* in chickpea root nodulation using transcriptome profiling approach

S KAUR<sup>1</sup>, S SINGH<sup>2</sup> and G TANDON<sup>3</sup>

Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh 211 007

Received: 26 July 2017; Accepted: 29 August 2017

### ABSTRACT

Root nodule development is a key event in legume plants for nitrogen fixation. In this process, *Mesorhizobium* forms a symbiotic relationship with chickpea (legume) crop that can produce nodule to fix the environmental nitrogen for necessary vital activities. *Mesorhizobium* infected chickpea shows enhanced function at different stages. We studied high throughput sequencing (454/Roche) data of control and three stages, viz. early, middle and late stages of *Mesorhizobium* infection (Bioproject: PRJNA214031) during nodule formation to identify the key regulatory genes in nitrogen fixation process and long non-coding RNAs (lncRNAs) participating in *Mesorhizobium* inoculation of chickpea root nodule. *De novo* assembly of raw reads resulted in 38835 contigs with the N50 value of 1019 bp. Differential expression analysis of assembled transcripts showed that 5615 genes were expressed differentially, out of which 1696 genes were differentially expressed in early infection while 1338 and 2581 genes were expressed in middle stage and late stage of infection, respectively. On analysis of DEGs, it was found that 392 genes were common in all three stages, 1304 were present in both early and late infection while 946 genes were uniquely expressed in the middle stage of infection. It has been evident from studies that long non-coding RNAs participate in host responses to infection but many lncRNAs are still unknown at the level of *Mesorhizobium* inoculated root nodules of chickpea. The present study aimed at identification of the lncRNAs in chickpea root nodules and to understand the transcriptomic mechanism during *Mesorhizobium* inoculation. Findings of this study characterize differential expression of lncRNAs in *Mesorhizobium* infected chickpea root nodules and it further gives understanding of symbiotic association in both aspects of coding and non-coding regulation.

**Key words:** Chickpea, lncRNA, *Mesorhizobium*, NGS, Nodule, Transcriptome

Nitrogen is the major component of atmospheric air constituting approximate 78% of the total that plays a vital role in plant growth (Ghaly *et al.* 2015). The plant cannot use nitrogen directly due to less bioavailable nitrogen in the atmosphere. To fulfill the nitrogen demand of sustainable agriculture, nitrogenous fertilizer has been chemically produced and used (O'Hara *et al.* 1998). However chemical made fertilizer are not only expensive, but also adversely effect the environment. It contributes in the hike of greenhouse emission, water pollution, contamination and much more. In leguminous plants like chickpea, symbionts help in bioavailability by fixing atmospheric nitrogen (O'Hara *et al.* 1998). For this process legume plants have the special physiology of nodule in their root system and that provide host to the nitrogen-fixing microorganism. Chickpea forms a symbiotic relationship to *Mesorhizobium ciceri*, which is a nitrogen fixing, gram-negative, nodule

living bacteria (Nandasena *et al.* 2007).

Large scale transcriptomic studies like RNA sequencing and microarray in eukaryotic model genomes have identified that major portion of the eukaryotic genome is made up of non-protein coding transcripts (Liu *et al.* 2015, Li *et al.* 2014). Non protein coding classes are classified as tRNAs, rRNAs, snoRNAs and snRNA. tRNAs and rRNAs are well known for their role in protein synthesis, while in specific locations RNAs were modifies by snoRNA and snRNAs. Other than this, RNAs were classified on length basis, i.e. small non-coding RNA (20-30 nt), medium non-coding RNA (50-200 nt) and long non-coding RNA (above 200). All these RNAs have their specific functions out of which lncRNAs usually play a major role in splicing, gene activation and translation (Costa 2005). Next generation sequencing technologies have led to discovering lncRNAs in yeast and other eukaryotes (Bumgarner *et al.* 2009). Genome wide analysis discovered a vast amount of 50000 lncRNAs in the human genome (Iyer *et al.* 2015). Other than these very few studies had been carried out on plant, like rice (Zhang *et al.* 2014), cucumber (Hao *et al.* 2015), populus (Shuai *et al.* 2014) and maize (Li *et al.* 2014). It is evident from studies that lncRNAs play a role in stress

<sup>1</sup>Ph D Fellow (sukhii.deep@gmail.com), <sup>2</sup>Assistant Professor (satendra.singh@shiats.edu.in), <sup>3</sup>Ph D fellow (gitanjali.tandon@gmail.com), Department of Computational Biology and Bioinformatics.

response like heat and cold stress (Wunderlich *et al.* 2014, Heo and Sung 2011).

Chickpea genome is recently decoded but there is much more hidden information that needs to be explored. As chickpea is legume crop it contains a special physiological structure; nodule. Nodulation exhibit a symbiotic relationship for nitrogen fixation process. In our study, we not only identified the gene differentially expressed during different stages of *Mesorhizobium* inoculation to study the protein-coding genes, but also identified the lncRNAs in inoculation process in chickpea root nodule. There is no available information for inoculated chickpea during nitrogen fixation and similarly, there is no information available for coding or non-coding transcripts involvement and mechanism. This study will give an insight on the coding and non-coding transcripts involvement in the whole mechanism of root nodule formation for nitrogen fixing process.

#### MATERIALS AND METHODS

Single end reads were extracted from BioProject PRJNA214031 available at NCBI BioProject. This contains transcriptome sequencing data of cDNA extracted from root nodule of control and three stage of Mesorhizobial infection, viz. early, middle and late done using 454 GS FLX Titanium sequencer by National Institute of Plant Genome Research. SRA accession SRX330836 contains two runs, viz. SRR946983 and SRR946984 for root nodule reads of control state. Accession numbers SRX330835, SRX330827, and SRX330813 are for root nodule reads with early, middle and late Mesorhizobial inoculation, respectively.

Downloaded sequencing data were visualized by FastQC tool (Andrews 2010) followed by quality trimming using Trimmomatic (Bolger *et al.* 2014) recursively until clean reads were obtained. Filtered reads were checked for quality score, adaptor and GC content for all the three datasets. Preprocessed high quality reads were assembled using CLC Genomics workbench Version 7 (<https://www.qiagenbioinformatics.com/>). De novo assembly was done with set parameters of word size, bubble size, and minimum contig length.

Afterwards, the individual reads were mapped back to the assembly for differentially gene expression analysis in the early, middle and late stage of inoculation. Criteria like p-value cutoff, FDR, fold change cutoff were applied to select DEGs that are most expressed in *Mesorhizobium* inoculation. Filtered DEGs were compared with the non-redundant protein database using MPI-BLAST installed in ASHOKA (Advanced Supercomputing Hub for OMICS Knowledge in Agriculture). Gene ontology studies of assembled transcripts for describing biological processes, molecular functions, and cellular components, KEGG pathway analysis, and mapping was done by using BLAST2GO PRO (Conesa *et al.* 2005). Gene name and function were assigned by using top blast hit.

Contig has been filtered for removing sequences that can encode protein. For this, contigs were blasted against protein database after translating nucleotide to protein sequences to

use as query in Blastx. Blast2GO has been used to perform the blastx with e-value restriction of  $1 \times 10^{-5}$ . Higher e-value sequences were used for filtration of unannotated sequences and used for next step.

Coding potential assessment is used to identify the availability and potential of predicted ORFs. Three criteria have been applied for lncRNA detection in literature (Boltaña *et al.* 2016) (Khemka *et al.* 2016). (a) Coding potential assessment has been carried out using BLAST2GO plugin and that gives the information of coding and non-coding transcripts from the assembly. (b) Transcripts with less than 30 amino acid are chosen as per criteria of lncRNA i.e. <200 bp. (c) Conserved domain search has been performed using NCBI CDD to filter out any other instances that can code for protein. After all filtration, remaining transcripts were aligned with chickpea genome version 2 (Parween *et al.* 2015) that gives the information of all potential lncRNA. That has been reconfirmed by performing blastx for any protein-coding isoforms remains.

Each sample data with predicted lncRNAs were mapped to infer the expression of lncRNA at each stage of infection. Differentially expressed reads were identified by filtering out the reads with less than 2 fold change and p-value greater than 0.01. Remaining transcripts were differentially expressed in individual infection stage, viz. early, middle and late stage.

SSR mining was carried out using MISA tool (Thiel 2003.) MISA (MicroSAtellite Identification Tool) identifies the perfect and compound microsatellites based on a definite number of nucleotides bases. MISA identifies SSRs with motifs: mononucleotide, dinucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide with certain repeat units. In-house perl scripts were used to calculate and parse the motif type and repeat units. 5 sets of reverse forward primers were designed for di-hexanucleotide motif type along with their melting temperature and product size using primer3 script (Untergasser *et al.* 2012).

SNP mining has been carried out using bwa-samtools-bcftools. BWA is used for indexing mapping, samtools for alignment and bcftools for variant detection. Variant file obtained from bcftools were filtered out using snpEff (Cingolani *et al.* 2012a) and snpSift (Cingolani *et al.* 2012b). snpEff and snpSift is a toolbox for annotation of genetic variant and prediction of effect. VCF file were filtered out for quality score greater than 30 and depth greater than 8, followed by detection of homology or heterology.

To identify transcription factor, all DEGs were blastx against plant transcription factor database (PlnTFDB; <http://plntfdb.bio.uni-potsdam.de/v3.0/downloads.php>) (Riano-Pachon *et al.* 2007) with the e-value cutoff of  $1e-03$  (Kalra *et al.* 2013).

psRNATarget (Dai and Zhao 2011) was used to identify the miRNA targets for identified transcripts. psRNATarget uses parameters like length for complementary scoring of 20 bp with 200 top target genes for each small RNA. The tool uses 25 as allowed maximum energy to UPE (unpair the target site) and flanking length of 17. Enzyme code was

Table 1 Statistics of raw sequencing reads and reads after trimming

Samples	Control 1	Control 2	Early	Middle	Late
Total raw reads	330667	162169	180272	324289	525023
Total clean reads	327919	161004	178294	322068	521943
GC	44	44	42	43	42
Length	50-568	50-568	50-568	50-568	50-568
% trimming	0.63%	0.61%	1.54%	0.60%	0.58%

detected using enzyme code detection module of blast2GO.

## RESULTS AND DISCUSSION

To identify the expression and regulation of lncRNAs during *Mesorhizobium* inoculation, transcriptome data was collected at different stages, viz. early, middle and late stage (NIPGR; BioProject PRJNA214031). Quality evaluation and visualization of raw reads obtained from 454 GS FLX Titanium were performed using the FastQC tool (Andrews 2010). Trimmomatic (Bolger *et al.* 2014) was used to trim the low-quality reads and to filter out the adaptor content. After quality trimming 327919 and 161004 clean reads obtained for control sample with 44% GC content while 178294, 322068 and 521943 clean reads obtained for three-time interval of *Mesorhizobium* inoculation respectively (Table 1).

These high quality reads were assembled using CLC workbench Version 7. *De novo* assembly was carried out at set parameters of 21 word size and minimum contig length of 200. Total 38835 contigs were obtained with an average length of 769 bp and 1019 N50 value (Table 2).

Quality of assembly was assessed by aligning reads back to reference. Total 38841 mappings were obtained with our dataset. Total 449617 reads mapped to reference genome for control reads while 159260, 294298 and 463000 read mapped for early, middle and late infection samples, respectively.

Data was bifurcated for coding and non-coding analysis. Protein coding transcripts were removed. For this, all transcripts were blasted against protein database by translating nucleotide query to protein. e-value cut-off used for filtration was  $1 \times 10^{-5}$  and the coverage was also considered for further analysis.

Coding potential assessment tool has been used to identify and filter those transcripts that may be linked to coding transcript and false positively detected as non-coding in our study. CPAT (coding potential assessment tool) is available from blast2GO plugin that has been used to evaluate protein-coding transcripts and discarded from the dataset. After coding potential filter, the presence of ORF was identified, i.e. ORF > 200 and after that conserved domain analysis was performed for identification of associated conserved domain to filter out that also. After preprocessing, a total of 682 contigs passed all the filtration and represented potential lncRNAs in *Mesorhizobium* inoculated chickpea. After analytical filtering of this transcriptomic data, it was found that non-coding and coding both were differentially expressing in *Mesorhizobium* infection on chickpea roots

Table 2 Statistics of *de novo* assembled reads

Contig	38835
Total nucleotides	29847099
N75	572
N50	1019
N25	1687
Minimum length	200
Maximum length	6983
Average length	769

at various stages.

### Features of lncRNAs in stages of root nodulation and differential expression

lncRNA has been known as shorter RNAs that are less expressive than coding transcripts but showing differential expression in tissues and time interval sampling (Li *et al.* 2012). In our study time interval stage expression of lncRNAs was analyzed using transcriptomic approach. A total of 682 lncRNAs were detected while 614 were common in all. Specifically, 15, 14 and 29 lncRNAs were uniquely expressed in early, middle and late stages, respectively (Fig 1).

Out of all lncRNAs, 26, 27 and 37 were differentially expressing in early, middle and late stages, respectively. Out of these, 15 were uniquely expressing at the early stage while 14 and 29 in middle and late stage. 6, 3, and 1 lncRNAs were co expressing in early- middle, middle-late and late-early stages, respectively. miRNA target prediction analysis was performed using psRNATarget, on analysis Ca\_lncRNA\_24374 targeted on gme-miR8636,

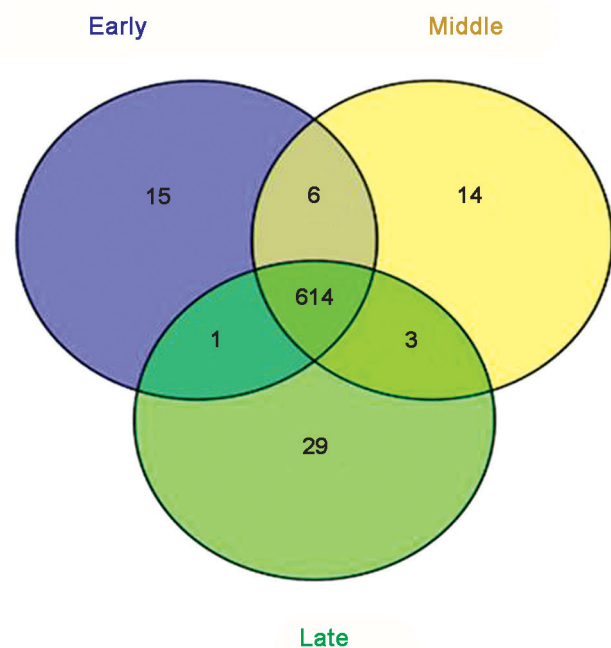


Fig 1 Venn diagram showing putative lncRNAs (684) distribution in all the three stages

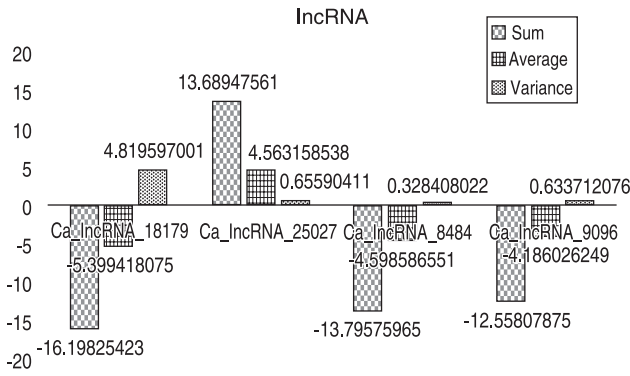


Fig 2 Statistical analysis of four commonly expressed lncRNA in all the three stages

Ca\_lncRNA\_2425 on gra-miR4403, Ca\_lncRNA\_967 on cre-miR906-5p from early stage and Ca\_lncRNA\_13671, Ca\_lncRNA\_2873, Ca\_lncRNA\_7546, Ca\_lncRNA\_8286 on ath-miR5021, vvi-miR2611, pti-miR5477 and tae-miR9672b respectively for late stage and Ca\_lncRNA\_23578 on osa-miR2122 for middle stage lncRNAs. lncRNAs that were expressed in all three stages are Ca\_lncRNA\_18179, Ca\_lncRNA\_25027, Ca\_lncRNA\_8484, Ca\_lncRNA\_9096. Statistical analysis shows the co-expression of all 4 lncRNA in nodulation stages (Fig 2).

Differential expression analysis of coding genes

Statistical analysis of DGE (differential gene expression) was performed using CLC Bio Workbench. Expression analysis of assembled transcripts showed that 5615 genes were expressed differentially, out of which 1696 genes were differentially expressed in early infection while 1338 and 2581 genes were expressed in middle stage and late stage of infection, respectively. DEGs of each stage were filtered out using P-value < 0.05 and 2 fold change cutoff. It was found that 392 genes were common in all three stages, 1304 were present in both early and late infection while 946 genes were uniquely expressed in the middle stage of infection.

Annotation of assembled transcripts was carried out

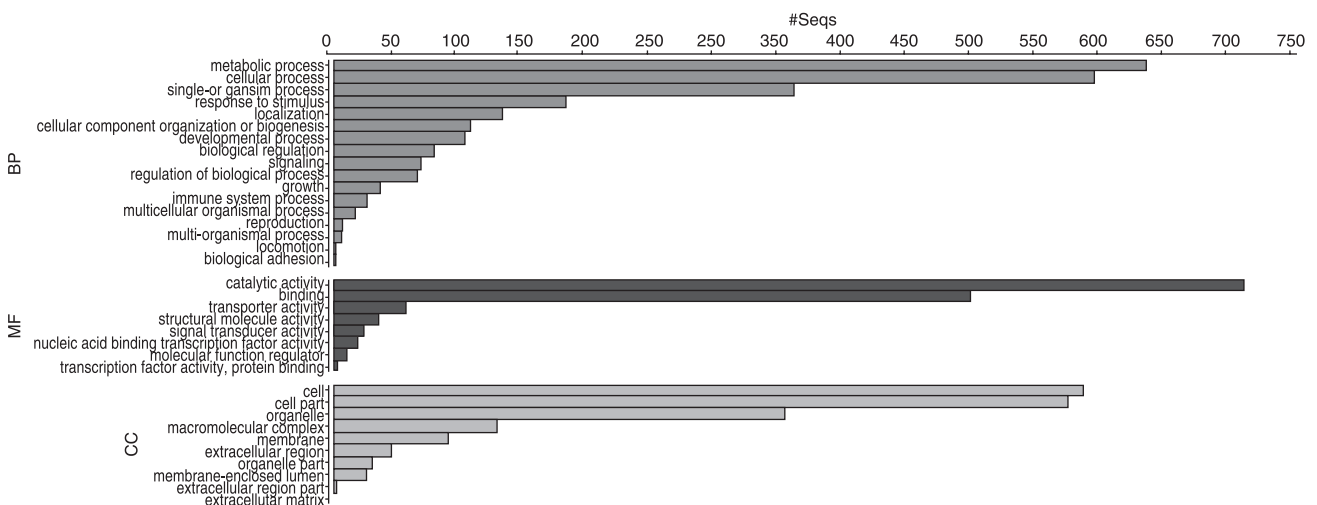


Fig 3 GO annotation distribution of chickpea expressed root genes in Mesorhizobium infection state

using blastx against the NCBI-protein nr database with e-value cutoff of 1e-3. Assembled transcripts were mapped against Pfam and InterPro for domains. Species analysis shows major hits from Glycine max followed by Medicago truncatula, and Glycine soja.

Protein sequence analysis and classification was done on DEGs. InterPro scan results show the involvement of protein family, viz. protein phosphatase 2c, kinesin like-protein, cytochrome p450, aspartic peptidase and Trichome birefringence-like family etc.

Protein phosphatase 2C also known as PP2C is the member of serine/threonine specific protein phosphatases. PP2C shows substrate specificity and its activity depend on Mn and Mg ions. Cytochrome P450 superfamily of heme-containing mono-oxygenases is significant for the biosynthesis hormones, fatty acids, and defensive compounds. In bacteria like Mesorhizobium, Cytochrome P450 is important for several metabolic processes, like biosynthesis of antibiotic. Kinesin is a microtubule-associated force-producing protein playing a role in organelle transport.

WD40 and LRR repeat motifs are an important site for protein-protein interaction, and proteins containing WD40 repeats are known to serve as platforms for the assembly of protein complexes. Proteins containing TPRs are involved in a variety of biological processes, such as cell cycle regulation, transcriptional control, and mitochondrial and peroxisomal protein transport. These repeats contain conserved domains. In our study, we found domain like P-loop containing nucleoside triphosphate hydrolase, protein kinase-like domain, serine/threonine/dual specificity protein kinase, catalytic domain and Armadillo-type fold.

All the DEGs were used for GO annotation and assigned with GO terms. Total 35 functional groups were distributed in three categories, viz. molecular function, biological process and cellular components. In molecular function category, molecular process, cellular process and single organism process are top GO terms. While catalytic and binding in biological process and cell, cell part and organelle in cellular components were mainly enriched (Fig 3).

### Molecular marker analysis of transcripts

We identified SSRs for all assembled transcripts. A total of 3530 SSRs were predicted, of which 1063 were mono-nucleotide, 779 di-nucleotide, 1336 tri-nucleotide, 49 tetra-nucleotide, 15 penta-nucleotide and 25 were hexa-nucleotide. 237 and 26 complex and compound SSRs were predicted, respectively. In addition to that, motif type, repeat number, length, starting and ending position of each SSR were analyzed. Total five pairs (reverse, forward) of primers was generated for each marker along with their melting temperature and product size. Parietal primer list of markers is available in supplementary Table 3.

Along with SSR mining, the detection of SNP and indels was done using bwa-samtools-vcf tools. On analysis, each sample against control, a total of 757, 970 and 1036 SNPs were obtained for early, middle and late, respectively. Similarly, 3484, 4322 and 4525 indels were detected for early, middle and late, respectively. A total of 643 SNPs and 3013 indels were common in all the three stages suggesting that they are not influenced by the *Mesorhizobial* stage of infection (supplementary Table 4).

### Transcription factor identification

DEGs were blasted against PlantTFDB using blastx with an e-value cutoff of 1e-03. On analysis, all the stage DEGs were searched against complete plant transcription factors. For early infection stage a total of 789 hits found, while 491 and 981 hits were found for middle and late stages of *Mesorhizobium* inoculation, respectively. Transcription factor families like MYB\_related, bHLH, WRKY, C3H, bZIP, kinase superfamily, ERF, MYB, C2H2, G2-like, B3, NAC, FAR1 family were found (Fig 4).

Previous studies used five legumes (*Cajanus cajan*, *Cicer arietinum*, *Glycine max*, *Lotus japonicus* and *Medicago truncatula*) transcription factor information and used it on chickpea nodule transcriptome and reported similar results (Kant *et al.* 2016) (Moreau *et al.* 2011).

### Enzyme detection and network analysis of the transcripts

We further performed intensive biochemical pathway analysis by assigning DEGs to KEGG database. A total of 135 KEGG involved in nodule formation and other key events of chickpea-*Mesorhizobium* infection were identified. Out of these, 124 KEGG pathways were associated with early inoculated samples, while 112 and 124 pathways in middle and late (supplementary Table 5). Common pathways that are highly associated in all three stages are purine metabolism (78, 79, 141 seqs), biosynthesis of antibiotics (61, 34, 91 seqs), starch and sucrose metabolism (58, 39, 77 seqs), thiamine metabolism (58, 69, 115 seqs), amino benzoate degradation (43, 35, 44 seqs), phenylpropanoid biosynthesis (43, 14, 38 seqs) and T cell receptor signaling pathway (22, 24, 31 seqs). Major enzyme class associated with pathways was hydrolases and transferases that play a major role in the reaction mechanism of pathways involved in chickpea *Mesorhizobium* inoculation.

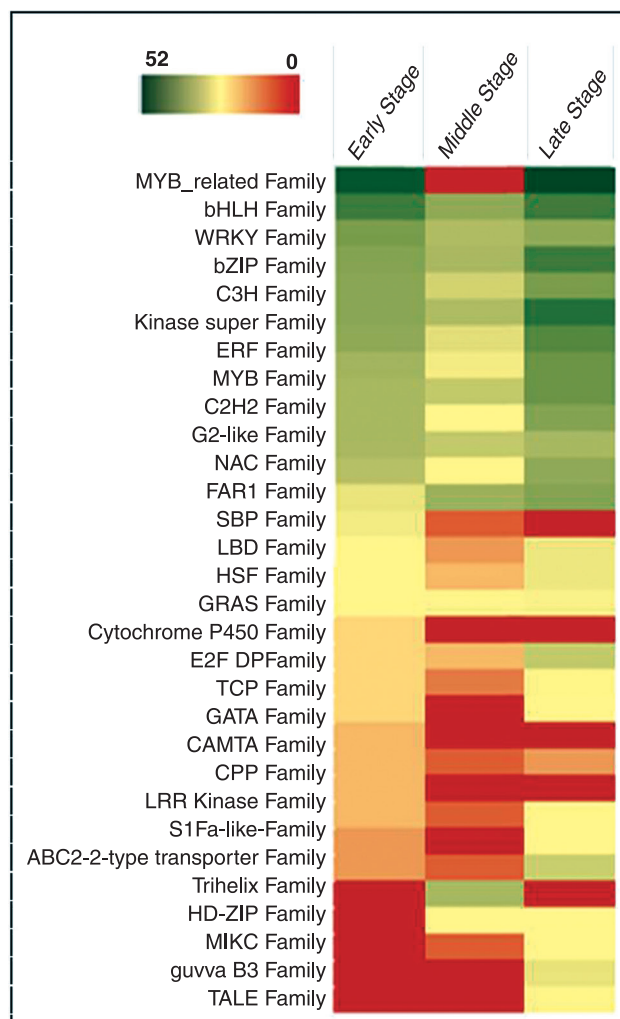


Fig 4 Differential expression patterns of TFs in various stages of nodule development.

### miRNA target prediction

A total of 2826, 2049 and 2049 miRNA target instances were found from DEGs of early (961), middle (723) and late stage (722). Out of these, 1973 were unique miRNA target detected for early stage and 1597, 1597 for middle and late stage respectively. In above all miR395, miR171, miR393, miR2630 abundantly present. Evidence have been found from previous studies that miR393 plays role in defense pathways by suppressing the negative regulators of immune responses (Navarro *et al.* 2006). Target for miR171 was majorly present in DEGs supported by previous studies that show miR171 as regulator of NSP2, i.e. key nodulation TF (Formey *et al.* 2016)

During nodule formation, symbiotic association of *Mesorhizobium* and chickpea express and change the immune system. Many studies have been published on the expression of protein-coding genes during inoculation or the virus plant interaction. But lncRNAs also reported as it plays role in regulation of cellular mechanism during the inoculation at the tissue level. We have studied that non-coding RNAs do play a major role in stress and infection

state that gives a clue that lncRNA can also be important (Boltaña *et al.* 2016), Pauli *et al.* 2012).

In chickpea itself, a recent study published showing the importance of lncRNAs during flower development in chickpea (Khemka *et al.* 2016). In our study, lncRNAs that are regulated and expressed during *Mesorhizobium* inoculated chickpea at the early, middle and late stage were studied using high throughput data. Transcriptomic data was preprocessed using lncRNA detection pipeline. As a result, a total of 682 lncRNA were detected, out of which 161 were targeted by chickpea miRNA (supplementary Table 6). Chickpea miRNAs are not available in miRBase so public available chickpea miRNA information has been collected from Jain *et al.* (2014). Comparative analysis of lncRNA in each stage shows that 15, 14 and 29 lncRNAs were unique in each stage. miRNA target prediction of these unique lncRNAs with all available miRBase miRNA shows 3 hit from early, 1 from middle and 4 from the late stage. For early stage Ca\_lncRNA\_24374, Ca\_lncRNA\_2425 and Ca\_lncRNA\_967 got hit with miR8636, miR4403, miR906-5p. Similarly, Ca\_lncRNA\_23578 matched to miR2122. Ca\_lncRNA\_13671, Ca\_lncRNA\_2873, Ca\_lncRNA\_7546, Ca\_lncRNA\_8286 matched to miR5021, miR2611, miR5477, miR9672b. miR8636 is reported in cotton plant and it supports the fibre elongation and similar process (Xue *et al.* 2013), miR4403 (Joshi *et al.* 2010), miR06-5p (Zhao *et al.* 2007), miR5021 (Borges *et al.* 2011), miR5477 (Huang *et al.* 2011), miR9672b (Wei *et al.* 2009) in *Glycine max*, green algae, *Arabidopsis thaliana*, *Phaeodactylum tricornutum* bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.). miR2122 has been reported in rice in heavy metals regulation mechanism (Huang *et al.* 2009), while miR2122 is already reported functional in root and nodule regulation in *Medicago truncatula*. Phylogenetically *Medicago* is proximal to chickpea and shares a common ancestor (Lelandais-Briere *et al.* 2009). All these findings of non-coding RNA and their homology with the other crops suggest the involvement of these lncRNAs in the nodule formation process and this gives a novel insight of long non-coding RNA in chickpea plant. Along with non-coding analysis, we analyze differentially expressed protein-coding genes and thoroughly studied that. Expression analysis of assembled transcripts showed that 5615 genes were expressed differentially, after annotation we found that Ankyrin repeat protein, prefoldin subunit 6, PREDICTED: uncharacterized protein LOC101497554, conserved protein is common in all stages while 18 common in early and middle stage, i.e. casein kinase I isoform delta-like, myb family transcription factor apl-like isoform x2, probable low-specificity l-threonine aldolase 1, caffeic acid o-methyltransferase, plant t10f18-100 protein, leukocyte receptor cluster member 8 homolog isoform x1, aminoacyl tRNA synthase complex-interacting multifunctional protein 1, wd repeat-containing protein 82-b, abscisic acid receptor pyl9-like, serine threonine-protein kinase ctrl1, arm repeat protein interacting with abf2-like, helicase protein mom1-like isoform x1, carotenoid (9)-cleavage

dioxygenase 1-like, short-chain alcohol dehydrogenase, mitochondrial import inner membrane translocase subunit tim17 tim22 tim23 family protein, dihydropyrimidine dehydrogenase, burp domain-containing protein 3-like, f-box protein pp2-b15-like. Similarly, 10 common in early and late stage hypothetical protein Smp\_000460, probable glutathione s-transferase para, duf581 family protein, cysteine proteinase, rna recognition motif, embryo-specific protein, dna glycosylase superfamily protein isoform 1, cytoplasmic membrane protein, auxin-responsive protein iaa14-like, duf1262 family protein and 20 common in middle and late stage abscisic acid-insensitive 5-like protein 2, agenet domain protein, polygalacturonase adpg2, gamma-glutamyltranspeptidase 1, snf1-related protein kinase catalytic subunit alpha kin10 isoform x1, eeig1 ehbp1 protein amino-terminal domain protein, tld domain-containing protein 1, common plant regulatory factor 1, PREDICTED: uncharacterized protein LOC101495764, auxin-induced 5ng4-like protein, dnaj homolog subfamily c member 21, protein flug, phosphoenolpyruvate carboxylase 4, glutathione s- amino-terminal domain protein, aba-responsive protein, s-adenosylmethionine synthase 3, udp-glycosyltransferase 76f1-like, beta-glucosidase 18-like, hypothetical protein glysoja\_035587, cysteine-rich repeat secretory protein 60-like. Molecular markers for each stage also listed.

Transcriptomic 454 sequencing reads of *Mesorhizobium* inoculated root tissues were obtained and after preprocessing used for *de novo* assembly by CLC workbench. Assembly generated 38835 contigs with average length of 769. This assembly used for further downstream analysis that addresses an important study that characterize differential expression of lncRNAs in *Mesorhizobium* infected chickpea root nodules and it further gives understanding of symbiotic association in both aspects of coding and non-coding regulation. Furthermore, potential SSRs and their primers, SNPs and indels, miRNA target, associated metabolic pathways, transcription factor identification was carried out that gives the way for further future genomics and proteomics researches on chickpea.

#### REFERENCES

- Alexandre A, Brigido C, Laranjo M, Rodrigues S and Oliveira S. 2009. Survey of chickpea rhizobia diversity in Portugal reveals the predominance of species distinct from *Mesorhizobium ciceris* and *Mesorhizobium mediterraneum*. *Microbial Ecology* **58**(4): 930–41.
- Allen E, Xie Z, Gustafson A M and Carrington J C. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**(2): 207–21.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
- Arfaoui A, Sifi B, Boudabous A, Hadrami I E and Chérif M. 2006. Identification of *Rhizobium* isolates possessing antagonistic activity against *Fusarium oxysporum* f. sp. *ciceris*, the causal agent of *Fusarium* wilt of chickpea. *Journal of Plant Pathology* **88**(1): 67–75.
- Bolger A M, Lohse M and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114–20.

- Boltaña S, Valenzuela-Miranda D, Aguilar A, Mackenzie S and Gallardo-Escárate C. 2016. Long noncoding RNAs (lncRNAs) dynamics evidence immuno modulation during ISAV-Infected Atlantic salmon (*Salmo salar*). *Scientific reports*, **6**: 22698
- Borges F, Pereira P A, Slotkin R K, Martienssen R A and Becker J D. 2011. MicroRNA activity in the *Arabidopsis* male germline. *Journal of Experimental Botany* **62**(5): 1611–20.
- Bumgarner S L, Dowell R D, Grisafi P, Gifford D K and Fink G R. 2009. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences* **106**(43): 18321–6.
- Cingolani P, Patel V M, Coon M, Nguyen T, Land S J, Ruden D M and Lu X. 2012. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics* **3**: 35.
- Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X and Ruden D M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**(2): 80–92.
- CLC Genomics Workbench 7 (<https://www.qiagenbioinformatics.com/>)
- Conesa A, Götz S, García-Gómez J M, Terol J, Talón M and Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18): 3674–6.
- Costa F F. 2005. Non-coding RNAs: new players in eukaryotic biology. *Gene* **357**(2): 83–94.
- Dai X and Zhao P X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* **39**(suppl\_2): W155–W159.
- Formey D, Martín-Rodríguez J Á, Leija A, Santana O, Quinto C, Cárdenas L and Hernández G. 2016. Regulation of small RNAs and corresponding targets in nod factor-induced *Phaseolus vulgaris* root hair cells. *International Journal of Molecular Sciences* **17**(6): 887.
- Ghaly A E and Ramakrishnan V V. 2015. Nitrogen sources and cycling in the ecosystem and its role in air, water and soil pollution: a critical review. *Journal of Pollution Effects and Control* **3**: 136.
- Hao Z, Fan C, Cheng T, Su Y, Wei Q and Li G. 2015. Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. *PLoS One* **10**(3): 0121800.
- Heo J B and Sung S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**(6013): 76–9.
- Howieson J and Ballard R. 2004. Optimising the legume symbiosis in stressful and competitive environments within southern Australia—some contemporary thoughts. *Soil Biology and Biochemistry* **36**(8): 1261–73.
- Huang A, He L and Wang G. 2011. Identification and characterization of microRNAs from *Phaeodactylum tricorutum* by high-throughput sequencing and bioinformatics analysis. *BMC Genomics* **12**(1): 337.
- Huang S Q, Peng J, Qiu C X and Yang Z M. 2009. Heavy metal-regulated new microRNAs from rice. *Journal of Inorganic Biochemistry* **103**(2): 282–7.
- Iyer M K, Niknafs Y S, Malik R, Singhal U, Sahu A, Hosono Y, Barrette T R, Prensner J R, Evans J R, Zhao S and Poliakov A. 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics* **47**(3): 199–208.
- Jain M, Chevala V N and Garg R. 2014. Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. *Journal of Experimental Botany* **65**(20): 5945–58.
- Jones-Rhoades M W and Bartel D P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell* **14**(6): 787–99.
- Joshi T, Yan Z, Libault M, Jeong D H, Park S, Green P J, Sherrier D J, Farmer A, May G, Meyers B C and Xu D. 2010. Prediction of novel miRNAs and associated target genes in *Glycine max*. *BMC Bioinformatics* **11**(1): S14.
- Kant C, Pradhan S and Bhatia S. 2016. Dissecting the root nodule transcriptome of chickpea (*Cicer arietinum* L.). *PLoS One* **11**(6): p.e0157908.
- Khemka N, Singh V K, Garg R and Jain M. 2016. Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Scientific Reports* **6**: p 33297.
- Laranjo M, Alexandre A and Oliveira S. 2014. Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus. *Microbiological Research* **169**(1): 2–17.
- Laranjo M, Alexandre A, Rivas R, Velázquez E, Young J P W and Oliveira S. 2008. Chickpea rhizobia symbiosis genes are highly conserved across multiple *Mesorhizobium* species. *FEMS Microbiology Ecology* **66**(2): 391–400.
- Laranjo M, Young J P W and Oliveira S. 2012. Multilocus sequence analysis reveals multiple symbiovars within *Mesorhizobium* species. *Systematic and Applied Microbiology* **35**(6): 359–67.
- Lelandais-Brière C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J and Crespi M. 2009. Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* **21**(9): 2780–96.
- Li L, Eichten S R, Shimizu R, Petsch K, Yeh C T, Wu W, Chetoor A M, Givan S A, Cole R A, Fowler J E and Evans M M. 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology* **15**(2): R40.
- Li T, Wang S, Wu R, Zhou X, Zhu D and Zhang Y. 2012. Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* **99**(5): 292–8.
- Liu X, Hao L, Li D, Zhu L and Hu S. 2015. Long non-coding RNAs and their biological roles in plants. *Genomics, Proteomics and Bioinformatics* **13**(3): 137–47.
- Moreau S, Verdenaud M, Ott T, Letort S, De Billy F, Niebel A, Gouzy J, de Carvalho-Niebel F and Gamas P. 2011. Transcription reprogramming during root nodule development in *Medicago truncatula*. *PLoS One* **6**(1): e16463.
- Nandasena K G, O'Hara G W, Tiwari R P, Willems A and Howieson J G. 2007. *Mesorhizobium ciceribiovarbiserrulae*, a novel biovar modulating the pasture legume *Biserrulapelecinus* L. *International Journal of Systematic and Evolutionary Microbiology* **57**(5): 1041–5.
- Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O and Jones J D. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312**(5772): 436–9.
- O'Hara G W. 1998. The role of nitrogen fixation in crop production. *Journal of Crop Production* **1**(2): 115–38.
- Parween S, Nawaz K, Roy R, Pole A K, Suresh B V, Misra G, Jain M, Yadav G, Parida S K, Tyagi A K and Bhatia S. 2015. An advanced draft genome assembly of a desi type chickpea

- (*Cicer arietinum* L.). *Scientific Reports* **5**: 12806
- Pauli A, Valen E, Lin M F, Garber M, Vastenhouw N L, Levin J Z, Fan L, Sandelin A, Rinn J L, Regev A and Schier A F. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research* **22**(3): 577–91.
- Postgate J. 1998. *Nitrogen Fixation*, 3rd ed. Cambridge University Press, Cambridge.
- Riaño-Pachón D M, Ruzicic S, Dreyer I and Mueller-Roeber B. 2007. PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics* **8**(1): 42.
- Rivas R, Laranjo M, Mateos P F, Oliveira S, Martínez Molina E and Velázquez E. 2007. Strains of *Mesorhizobium amorphae* and *Mesorhizobium tianshanense*, carrying symbiotic genes of common chickpea endosymbiotic species, constitute a novel biovar (*ciceri*) capable of nodulating *Cicer arietinum*. *Letters in Applied Microbiology* **44**(4): 412–8.
- Shuai P, Liang D, Tang S, Zhang Z, Ye C Y, Su Y, Xia X and Yin W. 2014. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *Journal of Experimental Botany* **65**(17): 4975–83.
- Smil V. 2000. *Cycles of Life*. Scientific American Library.
- Susan B. 2014. "The small but mighty chickpea". Phys.Org. Retrieved 8 October 2015.
- Thiel T. 2003. MISA—Microsatellite identification tool.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth B C, Remm M and Rozen S G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**(15): e115–e115.
- Verma J P, Yadav J, Tiwari K N and Kumar A. 2013. Effect of indigenous *Mesorhizobium* spp. and plant growth promoting rhizobacteria on yields and nutrients uptake of chickpea (*Cicer arietinum* L.) under sustainable agriculture. *Ecological Engineering* **51**: 282–6.
- Wei B, Cai T, Zhang R, Li A, Huo N, Li S, Gu Y Q, Vogel J, Jia J, Qi Y and Mao L. 2009. Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) Beauv. *Functional and Integrative Genomics* **9**(4): 499.
- Wunderlich M, Groß-Hardt R and Schöffl F. 2014. Heat shock factor HSFB2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. *Plant Molecular Biology* **85**(6): 541–50.
- Xue W, Wang Z, Du M, Liu Y and Liu J Y. 2013. Genome-wide analysis of small RNAs reveals eight fibre elongation-related and 257 novel microRNAs in elongating cotton fibre cells. *BMC Genomics* **14**(1): 629.
- Zhang Y C, Liao J Y, Li Z Y, Yu Y, Zhang J P, Li Q F, Qu L H, Shu W S and Chen Y Q. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biology* **15**(12): 512.
- Zhao T, Li G, Mi S, Li S, Hannon G J, Wang X J and Qi Y. 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes and Development* **21**(10): 1190–203.