Original Research Article

In silico Identification and Characterization of Conserved miRNAs and their Targets in Pigeon pea (Cajanus cajan L.) Expressed Sequence Tags

Akula Dinesh¹ and More Yogeshwar²

¹Department of Genetics and Plant Breeding, UAS, Raichur, 584102, India
²Department of Plant Pathology; PDKV, Akola, Maharashtra, India
*Corresponding author

ABSTRACT

MicroRNAs (miRNAs) are highly conserved class of short endogenous non-coding small RNA molecules of about 18–22 nucleotide in length. MicroRNAs negatively regulate the gene expression by degrading target mRNA. In the present study, comparative genomic based approach was used to identify and characterize new conserved microRNAs in “orphan legume crop” pigeon pea (Cajanus cajan L.) using expressed sequence tags (ESTs) by in silico method. A total of 4,621 uniquely previously reported microRNAs were used for homology search in 25,576 ESTs of pigeon pea for identifying conserved miRNAs. The results upon stringent selection found five conserved miRNAs namely cca-miR6483, cca-miR5219, cca-miR393a, cca-miR395a, cca-miR169b belonging to five different families. The target analysis through psRNATarget server found 27 mRNA targets which code for eukaryotic translation initiation factor 6 (EIF-6)-like protein, apyrase-like protein, ferric reductase, ATP sulfurylase, CCAAT-box transcription factor complex WHAP12, Lipoygenase-9, Transport inhibitor response 1, and MYB transcription factor MYB102 which play an important role in response to both biotic and abiotic stresses. Our results have laid the foundation for further research on miRNAs, which will lead to understand the gene silencing mechanisms at post transcriptional level for various stresses.

Keywords
MicroRNA, Pigeon pea, EST, Gene expression, In silico.

Introduction

Pigeon pea (Cajanus cajan L.) is an often cross pollinated legume crop (2n = 2x = 22) with genome size 808 Mb conferring 48,680 genes. In India, pigeon pea is cultivated in an area of 3.88 mha contributing to total production of 3.29 mt with average productivity of 849 kg/ha (Agricultural statistics at a glance, 2014). Pigeon pea serve has a multipurpose crop with unique benefits like firewood, fence, thatch and making baskets from its byproducts apart from economic yield. The decomposition of fallen leaves will enrich soil with nutrients and also symbiotic nitrogen fixation increases fertility (Varshney et al., 2010). The nutritional benefits include 45 % dietary fibers, 23 % protein, 7 % calcium, trace amounts of thiamin, riboflavin, and niacin. Despite of many advantages, pigeon pea has remained “orphan legume crop” with less genetic improvement (Varshney et al., 2009). Recognizing its importance, substantial amount of genomic resources have been generated, largely owing to the efforts of...
Indo-US Agricultural Knowledge Initiative (AKI), NSF and GCP funded projects and its genome sequence has been drafted (Varshney et al., 2010; Dutta et al., 2011; Bohra et al., 2011). To exploit advantages from pigeon pea, the knowledge about the genetic basis of yield, quality and stress tolerance is important for genetic improvement. Under such circumstances microRNAs (miRNA) are one of the major players in controlling biotic and abiotic stress responses in plants (Schwab et al., 2005). The miRNAs are approximately 21-nucleotide (nt) noncoding small RNAs that are encoded by MIR genes located in the intergenic (between protein-coding genes) or intragenic regions (within protein-coding genes) on the chromosomes which play critical roles in gene regulation at the post-transcriptional level (Schwab et al., 2005; Unver et al., 2009; Yang et al., 2012).

The molecular mechanism of miRNA-mediated gene expression involves perfect or near perfect complement with targeted mRNA sequences, and then degrade targeted mRNAs or repress mRNA translation (Bartel et al., 2004). The miRNAs are therefore negatively regulate the gene expression (Voinnet et al., 2009). A wide range of miRNAs have been discovered in model crops like arabidopsis, rice and maize which code for several developmental programs, such as root initiation and development, vascular development, leaf morphogenesis and polarity as well as floral differentiation (Marin et al., 2010; Donner et al., 2009; Mallory et al., 2004; Chuck et al., 2008).

Reviewing the important of miRNAs in plant development, many methods were developed to identify the miRNAs among them computational method was effective, quick and less time consuming. The biogenesis of miRNAs suggests that it is possible to find miRNAs by searching expressed sequence tags (ESTs) with known miRNAs. This method provided avenue for identification of conserved miRNAs based on comparative genomics in many species like Allium sativum, Camellia sinensis, Zea mays, Glycine max and Gossypium herbacium (Panda et al., 2014; Das and Mondal, 2010; Zhang et al., 2006a; Liu et al., 2011; Boopathi and Pathmanaban, 2012).

As mentioned above, pigeon pea an “orphan legume crop” despite of many advantages with less information was registered about miRNAs. This article mainly provides genetic basis for different responses in pigeon pea to stress and enhance the genetics of the crop. The identification of new miRNAs was done through in silico mining in expressed sequence tags (ESTs). Our efforts resulted in identification of five novel microRNAs in pigeon pea.

Materials and Methods

Collection of miRNA sequences and ESTs of pigeon pea

To search potential new miRNA in pigeon pea expressed sequence tags (ESTs), the sequences of previously known plant mature miRNA sequences from viridaeplantae kingdom were downloaded from miRBase release 21 (http://www.mirbase.org/ftps.html). The total count of plant mature miRNA sequences was around 8486 which includes lots of similarities and duplicates.

The redundancy in miRNA sequences were removed by cd-hit online sever with Sequence identity cut-off of 1 which means 100% similarity between the sequences. This resulted in 4621 unique miRNA sequences. These miRNA sequences were BLASTed to assembled EST sequences of pigeon pea (Cajanus cajan L.). The EST sequences (25,576 as on 2014) were downloaded from NCBI’s dbEST database (http://www.ncbi.nlm.nih.gov/).
Processing of pigeon pea ESTs in to contigs and singletons

All downloaded ESTs are partial sequences of a gene which consist of wide range of contamination like, polyA tails, repetitive regions, vector sequences and more over many of the ESTs are duplicates. All redundant and poor quality sequences were subjected to EGassembler online server, which processes the data in series of sequence cleaning, repeat masking, vector masking, organelle masking and finally sequence assembly by using CAP3 to give contigs (1,378) and singletons (14,814) with all parameter kept default (Masoudi-Nejad et al., 2006). These contigs and singletons represent the non redundant part of downloaded ESTs.

Precursor miRNA prediction

The processed unique miRNA reference set was used for homology search in pigeon pea contigs and singletons using BLASTn option in BioEdit software version 7.2.5 (Hall et al., 1999). The blast parameters like e-value was kept 1, word match size 11 and match-mismatch score (1,-4) and filtered low complexity regions. All those pigeon pea EST sequences which are having query coverage of 95 to 100% and with mismatch less than 2 with miRNA sequences were selected and forwarded for further analysis to remove protein coding sequences among the identified ESTs. Since all the miRNA genes are non protein coding (Lee et al., 1993), BLASTx online server was used to remove the protein coding sequences through blasting of precursor’s to non redundant (NR) protein data base of pigeon pea. The non coding sequences were selected based on criteria, E value less than e^{-5} along with the identity percent less than 25 (Dehury et al., 2013). The workflow for identification and characterization of new miRNAs were presented in figure 1.

The leftover non coding sequences were used for prediction of precursor miRNA using Zuker folding algorithm in MFOLD software version 3.2 (http://mfoldrnaalbanyedu/?q=mfold/rna-folding-form) with all default parameters (Zuker, 2003). The precursor sequences were searched at 50 nucleotides upstream or downstream from the location of mature miRNAs with an increment of 10 nucleotides. A stringent selection criteria was followed to select novel miRNA (Zhang et al., 2005). The selection criteria in order to identify the appropriate hair pin structure includes 1) The minimum length of the pre-miRNAs to be 60 nt, 2) The pre-miRNA should fold into an appropriate stem loop hairpin secondary structure, 3) The miRNA/miRNA* duplex i.e., the mature miRNA sequence and its opposite miRNA strand (miRNA*) should not have more than 7 nt mismatches (Das and Mondal, 2010) 4) The mature miRNA sequence should be placed in one arm of the hairpin structure, 5) The A+U content should be within 30–70% and 6) Predicted secondary structure should have higher minimal folding free energy index (MFEI)

$$\text{MEFI} = \frac{[(\text{MEF/length of the RNA sequence}) \times 100]}{(\text{G+C}) \%}$$

Target prediction

The newly identified miRNA sequences were submitted to psRNA Target tool for target prediction (http://plantgrnnobleorg/psRNA Target/) by specifying search on Glycine max (soybean), unigene, DFCI Gene Index (GMGI), version 16.

The earlier reports have found that pigeon pea genome has high synteny relation with soybean genome (Varshney et al., 2011). All the parameters were kept constant except expected value 3, to have good number of targets.
Functional annotation of target proteins

The genome annotation of the identified targets was done by using QuickGo (http://www.webiacc.uk/ QuickGO) tool. Furthermore, three important components such as biological process, cellular component and molecular function associated with each GO term were retrieved. The predicted miRNAs were named in accordance with miRBase (Griffiths-Jones et al., 2006). The mature sequences are designated ‘miR’, and the precursor hairpins are labeled as ‘mir’ with the prefix. In case of Cajanus cajan it will be cca-miR395a for homologue of sly-miR395a.

Results and Discussion

In silico identification of novel miRNA

A total of 25,576 ESTs of pigeon pea (Cajanus cajan) were assembled to form contigs (1,378) and singletons (14,814) which were used for identification of new miRNAs based on homology relationship with previously reported non redundant mature miRNAs. The BLASTn programme with specific parameters (see materials and method) led to detection of 7 ESTs (4 singletons and 3 contigs) showing conserved miRNAs sequence with mismatch less than 2 nucleotide (nt) with previous miRNAs. The BLASTx operation these 7 ESTs resulted in retrieving 5 ESTs sequences which do not code for any protein. These 5 EST sequences were analyzed for presence of characteristic secondary hairpin structure. All the five ESTs fold into appropriate secondary structure and proved to be new miRNAs in pigeon pea (Fig. 1). The size of the newly identified miRNAs was in range of 20 nt to 22 nt which are considered as ideal length for typical miRNA (Zhang et al., 2006a). The A+U content ranged from 49 % to 69 % which is in agreement with Zhang et al., (2005) with an average of 58 % and similarly the G+C content ranged from 30 % to 50 % with an average of 41 % (Table 1). The length of the precursor miRNAs ranged from 93 to 194 nt with an average of 133 nt. The newly identified precursor miRNAs have minimum folding free energies (MFE) ranging from -74 to -162 kcal/mol which are not in consistence with folding energies of tRNA –275 kcal/mol and rRNA –33 kcal/ mol (Barozai et al., 2008). The Minimal folding free energy index (MFEI) and MFE have been considered as significant features that distinguish miRNAs from other non coding RNAs (rRNA, tRNA, mRNA). The MFEI of newly identified precursor miRNAs ranged from 0.97 to 5.2 which is significantly higher than that of tRNAs (0.64), rRNAs (0.59) and mRNAs (0.62–0.66) with an average of 2.4 proving that these newly identified miRNAs in pigeon pea are likely to be actual miRNAs than any other kind of non-coding RNA (Zhang et al., 2006b). All of above findings and analysis indicated that these five small RNAs were probably new miRNAs. The hairpin structure of newly identified miRNAs was presented in figure 2.

Computation based identification of putative target of new miRNAs

Result from psRNATarget server showed that the newly identified miRNAs have 27 target mRNAs, which code for different proteins involved in metabolism, response to stress, transcriptional regulation, signal transduction, growth, development, sulfate assimilation, protein lipoylation, coenzyme and, oxidoreductase activity. Our results found that the cca-miR6483 has highest number of targets ie., 10 genes followed by cca-miR395a with 7 targets genes, cca-miR5219 (2 target genes), cca-miR169b (4 target genes) and cca-miR393a (4 target genes). In the present study, it was observed that, one miRNA has more than one target mRNA, which were in
consistence with the reports of Boopathi and Pathmanaban (2012). The majority of targets genes includes enzymes 42.1 % having important role in assimilation, resistance mechanisms and biological functions followed by genes for proteins 26.3 %, transcription factors 21.1%, and transporters 10.5 % (Fig. 3). Our results were in consistent with the studies of Nodine and Bartel (2010) and Das and Mondal (2010) who predicted that miRNAs regulates the functional genes in plants that were involved in various physiological processes, leaf morphogenesis, stress responses and signal transduction. Many previous reports revealed that miRNAs of plants also regulates transcription factors (Lu and Yang, 2010). The newly identified miRNAs have following targets proteins, Eukaryotic translation initiation factor 6 (EIF-6)-like protein, Apyrase-like protein, Protein kinase GhCLK1, Extensin, Ferric reductase, ATP sulfurylase, Dihydroflavonol-4-reductase, CCAAT-box transcription factor complex WHAP12, Lipoygenase-9, Transport inhibitor response 1, 50S ribosomal protein L20 and MYB transcription factor MYB102. The functional annotation of target miRNAs is presented in table 2.

Table.1 Details of newly identified miRNAs in pigeon pea

<table>
<thead>
<tr>
<th>New miRNA</th>
<th>Source</th>
<th>Mature miRNA</th>
<th>Homologue miRNA</th>
<th>L (nt)</th>
<th>LM (nt)</th>
<th>NM (nt)</th>
<th>E-value</th>
<th>(A+U) %</th>
<th>(G+C) %</th>
<th>MFE (ΔG, kcal/mol)</th>
<th>MFEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>cca-miR393a</td>
<td>Contig1 015</td>
<td>CATCCAAAGGGA TCGCATTG</td>
<td>ppe-miR393a</td>
<td>3'</td>
<td>116</td>
<td>20</td>
<td>0</td>
<td>6.00E-06</td>
<td>49.1</td>
<td>50.9</td>
<td>-90</td>
</tr>
<tr>
<td>cca-miR169b</td>
<td>Contig1 261</td>
<td>TGAGCCAAAGGAT GGATTGCC</td>
<td>vvi-miR169b</td>
<td>5'</td>
<td>165</td>
<td>20</td>
<td>1</td>
<td>0.006</td>
<td>59.4</td>
<td>40.6</td>
<td>-161</td>
</tr>
<tr>
<td>cca-miR395a</td>
<td>gi</td>
<td>25173 9378</td>
<td>TGAAGTGTTTGG AGGAACTCC</td>
<td>sly-miR395a</td>
<td>3'</td>
<td>93</td>
<td>21</td>
<td>1</td>
<td>0.012</td>
<td>63.4</td>
<td>36.6</td>
</tr>
<tr>
<td>cca-miR5219</td>
<td>gi</td>
<td>25173 9202</td>
<td>TCATGGAATTCA GCTGCTGCA</td>
<td>mtr-miR5219</td>
<td>3'</td>
<td>194</td>
<td>22</td>
<td>1</td>
<td>0.003</td>
<td>53</td>
<td>46.9</td>
</tr>
<tr>
<td>cca-miR6483</td>
<td>gi</td>
<td>28445 6793</td>
<td>TATTTGGAATT TTCAGGATC</td>
<td>hbr-miR6483</td>
<td>5'</td>
<td>101</td>
<td>22</td>
<td>0</td>
<td>3.00E-06</td>
<td>69.3</td>
<td>36.6</td>
</tr>
</tbody>
</table>

The novel identified miRNAs were characterized in terms of L = Location of miRNA; PL = precursor miRNA length; LM = mature sequence length; NM = number of mismatches; MFE = minimal folding free energies; MFEI = minimal folding free energy index.
Table 2 Prediction of miRNA target genes and their functional annotation

<table>
<thead>
<tr>
<th>miRNA ID</th>
<th>Target Acc</th>
<th>Target description</th>
<th>Target function</th>
<th>Biological process</th>
</tr>
</thead>
<tbody>
<tr>
<td>cca-miR6483</td>
<td>TC436145</td>
<td>Eukaryotic translation initiation factor 6 (EIF-6)-like</td>
<td>Translation initiation factor</td>
<td>Mature ribosome assembly</td>
</tr>
<tr>
<td></td>
<td>TC465796</td>
<td>protein</td>
<td>Ribosomal large subunit binding</td>
<td>Translation initiation</td>
</tr>
<tr>
<td></td>
<td>BE440256</td>
<td>Apyrase-like protein</td>
<td>Hydrolase activity</td>
<td>Metabolic process</td>
</tr>
<tr>
<td></td>
<td>TC452083</td>
<td>Protein kinase GhCLK1</td>
<td>Transferase activity</td>
<td>Protein phosphorylation</td>
</tr>
<tr>
<td></td>
<td>TC452253</td>
<td>Extensin</td>
<td>Structural constituent of cell wall</td>
<td>Plant-type cell wall organization</td>
</tr>
<tr>
<td></td>
<td>TC473325</td>
<td>Ferric reductase</td>
<td>Oxidoreductase activity</td>
<td>Oxidation-reduction process</td>
</tr>
<tr>
<td>cca-miR5219</td>
<td>TC421771</td>
<td>60S ribosomal protein L9</td>
<td>Structural constituent of ribosome</td>
<td>Translational elongation</td>
</tr>
<tr>
<td></td>
<td>TC423861</td>
<td>Peptidyl-prolyl cis-trans isomerase</td>
<td>Isomerase activity</td>
<td>Protein peptidyl-prolyl isomerization</td>
</tr>
<tr>
<td>cca-miR395a</td>
<td>BG789910</td>
<td>ATP sulfurylase</td>
<td>Sulfate adenyllytransferase (ATP) activity</td>
<td>Sulfate assimilation</td>
</tr>
<tr>
<td></td>
<td>TC432008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TC421817</td>
<td>Plastidial lipoyltransferase 2</td>
<td>Transferase activity</td>
<td>Protein lipoylation</td>
</tr>
<tr>
<td></td>
<td>GD787823</td>
<td>Dihydroflavonol-4-reductase</td>
<td>Coenzyme binding</td>
<td>Metabolic process</td>
</tr>
<tr>
<td></td>
<td>BI426387</td>
<td>Cytochrome P450 82A2</td>
<td>Oxidoreductase activity</td>
<td>Oxidation-reduction process</td>
</tr>
<tr>
<td>cca-miR169b</td>
<td>TC474864</td>
<td>CCAAT-box transcription factor complex WHAP12</td>
<td>Sequence-specific DNA binding transcription factor</td>
<td>Regulation of transcription</td>
</tr>
<tr>
<td></td>
<td>TC485068</td>
<td>Lipoyltransferase-9</td>
<td>Linoleate 13S-lipoyltransferase activity,</td>
<td>Oxidation-reduction process</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Oxidoreductase activity</td>
<td></td>
</tr>
<tr>
<td>cca-miR393a</td>
<td>TC431817</td>
<td>Auxin-responsive factor TIR1 protein</td>
<td>Inositol hexakisphosphate binding</td>
<td>Auxin-activated signaling</td>
</tr>
<tr>
<td></td>
<td>TC429100</td>
<td>50S ribosomal protein L20</td>
<td>Structural constituent of ribosome Rna binding</td>
<td>Translation</td>
</tr>
<tr>
<td></td>
<td>TC432236</td>
<td>MYB102</td>
<td>Sequence-specific DNA binding transcription factor</td>
<td>Biotic and abiotic stress</td>
</tr>
</tbody>
</table>
Fig. 1 Workflow for identification and characterization of new miRNAs and target genes in pigeon pea using ESTs

Fig. 2 Secondary hairpin structure of newly identified miRNA in pigeon pea

A) cca-miR169b, B) cca-miR5219, C) cca-miR6384, D) cca-miR393a, E) cca-miR395a. The pink color bar on the stem loop structure of miRNA indicates the mature miRNA sequences.
The cca-miR6483 targets multiple genes, it includes Eukaryotic translation initiation factor 6 (EIF-6)-like protein which has translation initiation factor activity, apyrase like protein, extensin and ferric reductase. In Arabidopsis and Rice EIF-6 has important role in embryogenesis (Kato et al., 2010). The enzyme apyrase play important role in plant nutrition and nodulation. In soybean, cell wall protein extensin composed of hydroxyproline-rich glycoproteins helps in cell wall metabolism and another enzyme ferric reductase regulated iron homeostasis in plants, thus helps in preventing Fe deficiency (Hong et al., 1994; O’Rourke et al., 2007).

The miRNA cca-miR395a targets ATP sulfurylase that involves in sulfur assimilation and improves nutrient content of the plant, Dihydroflavonol-4-reductase enzyme that controls seed coat and flower colour in soybean (Herrmann et al., 2014; Yan et al., 2014). The miRNA cca-miR5219 suppresses the expression of 60S ribosomal protein L9 and peptidyl-prolyl cis-trans isomerase which help in translation elongation and protein peptidyl-prolyl isomerization respectively. The cca-miR169b negatively regulates transcription factors like CCAAT-box transcription factor complex WHAP12, nuclear transcription factor Y subunit A-3. The suppression of cca-miR169b leads to expression of Lipoxygenase-9, an important enzyme that confers resistance against the plant parasitic nematode, Heterodera glycines in soybean (Klink et al., 2009). cca-miR393a regulates transport inhibitor response 1(TIR1), involved in auxin-activated signaling pathway. The tir1 mutants leads to abnormal hypocotyl elongation and lateral root formation (Ruegger et al., 1998). The miRNA cca-miR393a also regulates MYB102, which express under wide range of biotic and abiotic responses (Vos et al., 2006).

In conclusion we identified five new miRNAs in EST of pigeon pea having 27 mRNA targets, most of the them involved in stress responses, development, physiological process, protein phosphorylations and other metabolic processes. This finding from our study leads to further investigation of miRNAs functions and regulatory mechanism under wide range of biotic and abiotic stress which leads to crop improvement of Pigeon pea “an orphan legume crop”
References


Unver, T., Covert, D. N. and Budak, H. 2009. Review of current methodological approaches for characterizing microRNAs in plants. *Int. J. Plant*
Genomics, 262-463.


How to cite this article: