Journal of Applied Horticulture, 17(1): 12-17, 2015

Computational prediction and characterization of miRNA from coconut leaf transcriptome

S. Naganeeswaran, T.P. Fayas, K.E. Rachana and M.K. Rajesh*

*Bioinformatics Centre, Division of Crop Improvement, Central Plantation Crops Research Institute, P.O. Kudlu, Kasaragod 671124, Kerala, India. *E-mail: mkraju.cpcri@gmail.com*

Abstract

Micro RNAs (miRNAs) are single stranded, small and non-coding endogenous RNA molecules, which control the gene expression at the post-transcriptional level either by suppression or degradation. Because of its highly conserved nature, *in silico* methods can be employed to predict novel miRNAs in plant species. By using previously known plant miRNAs available at miRBase, we predicted 16 miRNAs, which belongs to 11 miRNA families, and also targets for seven potential miRNAs in coconut leaf transcriptome. A majority of these seem to encode transcription factors. To the best of our knowledge, this is the first report of *in silico* prediction and characterization of miRNA from coconut. These findings form an useful resource for future research into miRNA prediction and function prediction in coconut and for studies on their experimental validation and functional analyses.

Key words: miRNAs, RNA, gene expression, *in silico*, miRBase, coconut, leaf transcriptome

Introduction

Plants encode process and accumulate different types of small RNAs, which range in sizes from 21-24 nucleotides (nt). Of these, microRNAs (miRNAs) are the most abundantly expressed and well-characterized (Li *et al*., 2010). These miRNAs are non-coding RNA molecules of ~ 22 nt, which are usually generated from stem-loop hairpin structures of ~80 nt called miRNA precursors (pre-miRNAs) (Lee *et al*., 2002). These precursors are initially transcribed as longer RNAs (Lee *et al*., 2002) and are processed initially by Drosha, a RNAase III enzyme (Lee *et al*., 2003), and later by another RNAase III enzyme, Dicer, which cuts these ~80 nt precursors to release ~ 22 nt mature miRNA.

In general, miRNA, which possess near-perfect complementary to their target mRNAs, play an important role in target gene expression and regulation, either by degradation or by inhibition of translation of mRNA target in plants (Jones-Rhoades and Bartel, 2004; Voinnet *et al*., 2009). Some of the biological processes in which miRNAs have been implicated to possess key roles include regulation of leaf, stem and root development, signal transduction, developmental timing, floral differentiation and development and defense response against biotic and abiotic stresses. miRNA genes have been reported to constitute 1-2 % of known eukaryotic genomes (Barbato *et al*., 2009). Plant miRNAs have been reported to be evolutionary conserved. Various approaches have been put forth and utilized for identification and characterization of miRNAs. Because of their low abundance, cloning of miRNAs has been found to be cumbersome. Recently, many computational programmes, both web-based or stand alone, have been developed for successful identification/prediction of miRNAs and their targets (Huang *et al*., 2007; Coronnello and Benos, 2013) .

Coconut (*Cocos nucifera* L.) is one of the major perennial plantation crops cultivated in more than 93 countries across the world. It is one of the economic palms used in domestic, commercial and industrial level. In the current study, we have used coconut leaf transcriptome data, generated in an Illumina HiSeq 2000 platform, for the prediction of miRNAs and their targets using bioinformatics approaches.

Materials and methods

Dataset: Coconut leaf transcriptome data of the cultivar Chowghat Green Dwarf (SRX436961), generated in our laboratory using RNA-Seq in an Illumina HiSeq 2000 platform, was utilized for the miRNA prediction. A total of 130942 coconut transcriptome unigenes were used for *in silico* prediction of miRNA. Mature reference plant miRNAs (5940) were retrieved from miRBASE Release 19 (Kozomara and Griffiths-Jones, 2011). The miRNA redundancy was removed using BLASTCLUST (ftp://ftp.ncbi.nih.gov/blast/executables/blast+/) program.

Homology search: A stand-alone database of coconut ESTs (from leaf transcriptome data) was created using FORMATDB (ftp://ftp.ncbi.nih.gov/blast/executables/blast+/) program and homology searches with non-redundant miRNA reference dataset using stand-alone BLASTN (Altschul *et al.*, 1990) program to identify coconut miRNA candidates. Hits with at least 18 nt and mismatch or gap <3 were selected. Sequences 100 bp upstream and downstream of the targeted miRNA candidates were selected. BLASTX (Altschul *et al*., 1990) was performed using the selected sequences to remove sequences, which codes for protein.

Prediction of miRNA: Those sequences which did not give any hits with BLASTX were used for RNA secondary structure prediction using MFOLD program (Zuker, 2003). The following criteria were considered for screening the candidate miRNA homologs: (i) Free energy change (dG), the structure should be less than or equal to -18 kcal/ mole, (ii) The un-pairing bulge size should not be more than 7 bp, (iii) Mature miRNA should be on the stem region of the hair pin structure (Singh and Nagaraju, 2008). Randomization test of predicted premiRNA was carried out using Randfold software (Bonnet *et al*., 2004). Default

randomization parameter (= 999) and simple mononucleotide shuffling were used for the *P*-value calculation for the predicted pre-miRNAs.

Computational prediction of potential miRNA targets: Target prediction of the identified miRNA was done using psRNATarget tool (Dai and Zhao, 2011) , by selecting *Arabidopsis thaliana* as reference.

Results

The work flow followed for *in silico* prediction of miRNA molecule from coconut leaf transcriptome is shown in Figure 1. A total of 5940 plant mature miRNA was retrieved from miRBASE and the redundancy was reduced using BLASTCLUST program (identity 100% and query coverage 95%) and a reference miRNA dataset with 2944 sequences was obtained. The reference dataset was searched against 130942 coconut unigenes assembled from coconut leaf transcriptome. Significant hits with at least 18 nt identity and mismatch or gap <3 were screened. A total of 107 miRNA-like sequences, along with 100 bp upstream and downstream regions, were selected as a rough precursor sequence for further analysis. Based on the BLASTX result against nr protein database, we manually removed protein coding sequences and obtained a total of 40 pre-miRNA-like sequences.

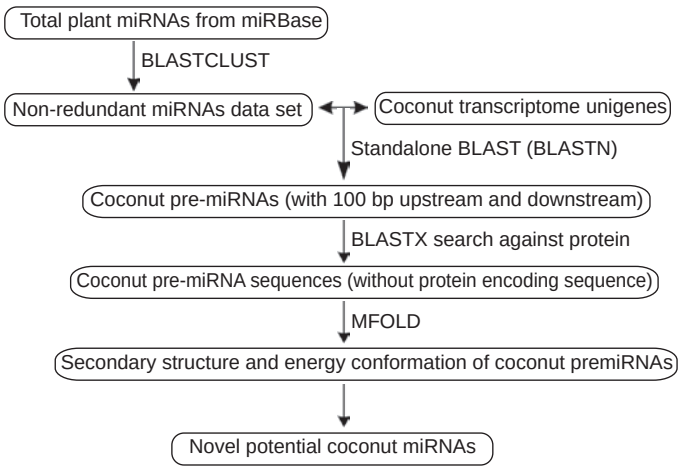

Fig.1. miRNA prediction workflow in coconut.

Based on the secondary structure analysis, 16 pre-miRNA sequences satisfying all the three criteria *viz*., free energy change (dG) lesser than or equal to -18 kcal/ mole, un-pairing bulge size of not be more than 7 bp and present on the stem region of the hair pin structure, were selected for further analysis. The structures of these 16 mature miRNA in coconut, with pre-miRNA predicted, are shown in Fig. 2a, 2b and 2c. The details of sequences of the predicted miRNAs from coconut are furnished in Table 1. Randfold analysis revealed that 14 out of 16 predicted premiRNA had *P*-value < 0.01 (Table 2).

The coconut miRNAs were classified into 11 miRNA families, with miR156 family comprising of three miRNA structures. miR171, miR172 and miR535 families comprise of two miRNA

Table 2. Predicted coconut miRNA

Coconut miRNA	MFE	*p*- value
coc_miR156a	-80.40	0.001
coc_miR157b	-85.70	0.001
coc_miR172a	-61.50	0.001
coc_miR172c	-69.70	0.001
coc_miR168a	-88.50	0.001
coc_miR156g	-78.40	0.002
coc_miR169g	-88.43	0.001
coc_miR171n	-69.10	0.017
coc_miR164	-70.79	0.001
coc_miR530a	-60.66	0.001
coc_miR535	-67.00	0.001
coc_miR167a	-72.26	0.003
coc_miR535-3p	-81.70	0.029
coc_miR528	-74.00	0.001
coc_miR171i	-67.00	0.009
coc_miR156k	-65.30	0.001

Table 1. Predicted coconut miRNAs

Coconut EST id	Coconut miRNA	coc_miRNA Sequence	Length	A+U %	dG (kcal/ mole)	miRdb id
unigene100115263	coc_miR156a	CAGAAGAUAGAGAGCACA	18	55.56	-74.99	MIMAT0004988
unigene100092892	coc_miR157b	GCUCUCUAUGCUUCUGUCAUCA	22	54.55	-84.70	MIMAT0017416
unigene100023756	coc_miR172a	UGUGAAUCUUGAUGAUGCCAC	21	57.14	-53.23	MIMAT0017521
unigene100086551	coc_miR172c	UGUGAAUCUUGAUGAUGCU	19	63.16	-65.90	MIMAT0017525
unigene100124553	coc_miR168a	CCCGCCUUGCAUCAACUGAAU	21	47.62	-86.85	MIMAT0017482
unigene100102683	coc_miR156	GACAGAAGAGAGUGAGCAC	19	47.37	-77.15	MIMAT0001012
unigene100104454	coc_miR169g	AGCCAAGGAUGACUUGCC	18	44.44	-88.40	MIMAT0005618
unigene100113395	coc_miR171n	AGAUAUUGACGCGGCUCAA	19	52.63	-64.09	MIMAT0023202
unigene100067901	coc_miR164	UGCACGUGCUCCCCUUCUCCA	21	38.10	-66.49	MIMAT0022757
unigene100106319	coc_miR530a	AGGUGCAGGUGCAAAUGCA	19	47.37	-58.50	MIMAT0020446
unigene100117720	coc_miR535	GCGUGCUCUCUCUCGUUGUCA	21	42.86	-62.25	MIMAT0012594
unigene100115901	coc_miR167	UGAAACUGCCAGAUGAUCU	19	57.89	-70.60	MIMAT0025922
unigene100013224	coc_miR535-3p	GUGCUUUCUCCCGUUGUCACU	21	47.62	-77.75	MIMAT0022927
unigene100104098	coc_miR528	CCUCUGCAUGCCCCUUCCA	19	36.84	-74.00	MIMAT0002884
unigene100113395	coc_miR171i	GAUUGAGGCACGCCAAUACCU	21	47.62	-60.33	MIMAT0022886
unigene100102683	coc_miR156k	GCUCACUUCUCUUUCUGUCAGC	22	50.00	-77.15	MIMAT0015208

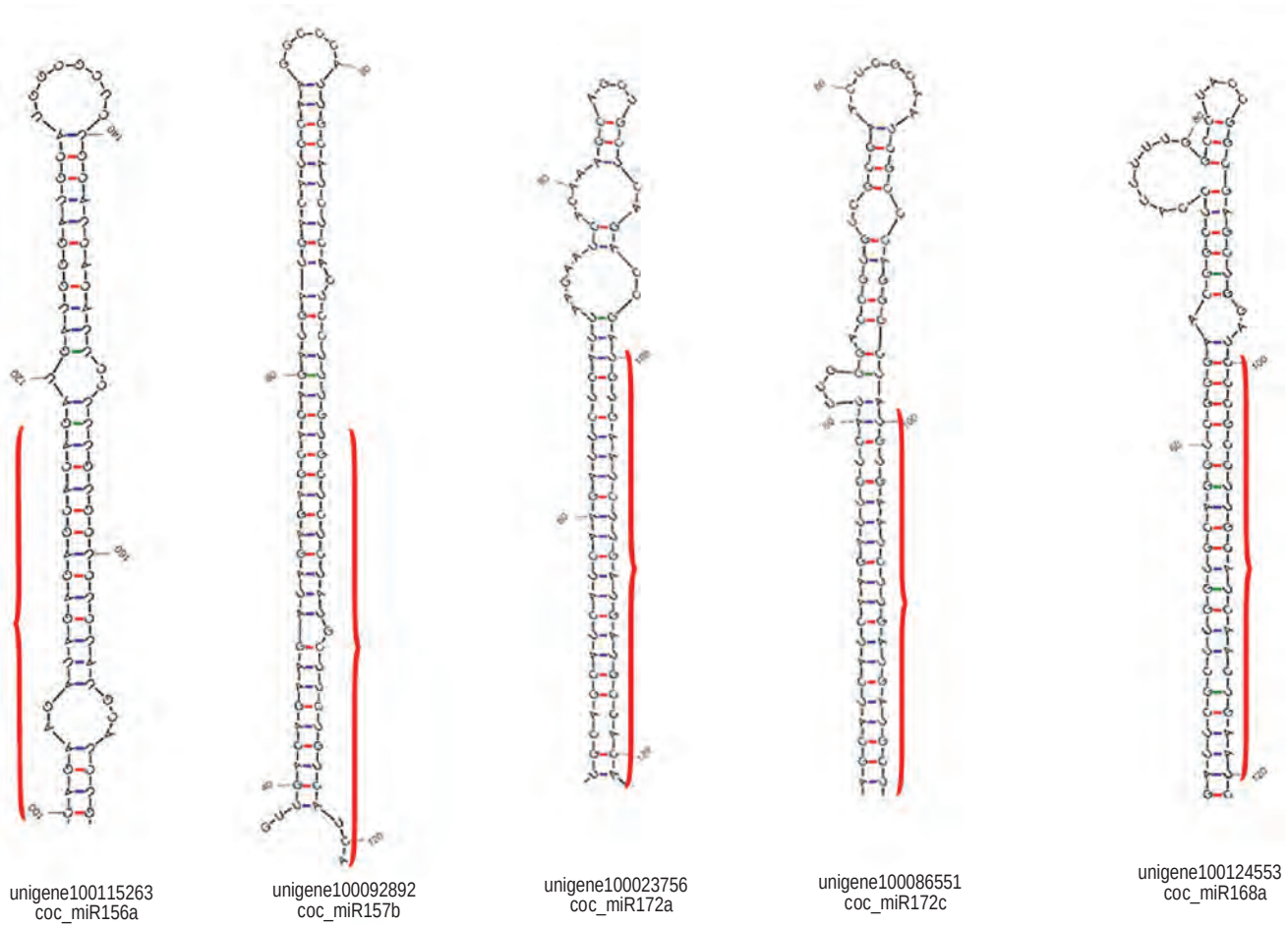

Fig. 2a. Predicted secondary structures of coc_miR156a, coc_miR157b, coc_miR172a, coc_miR172c and coc_miR168a

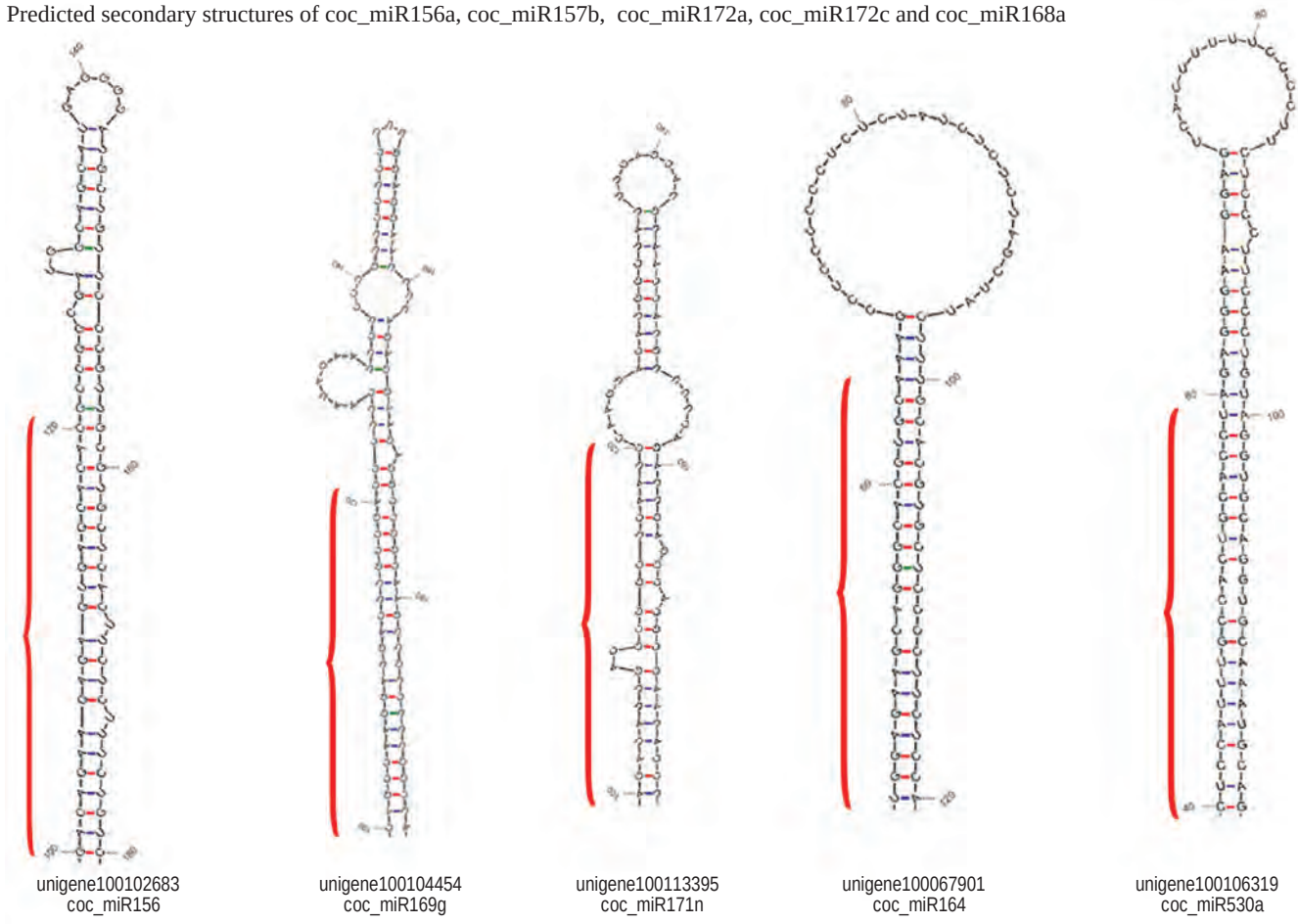

Fig. 2b. Predicted secondary structures of coc_miR156, coc_miR169g, coc_miR171n, coc_miR164 and coc_miR530a

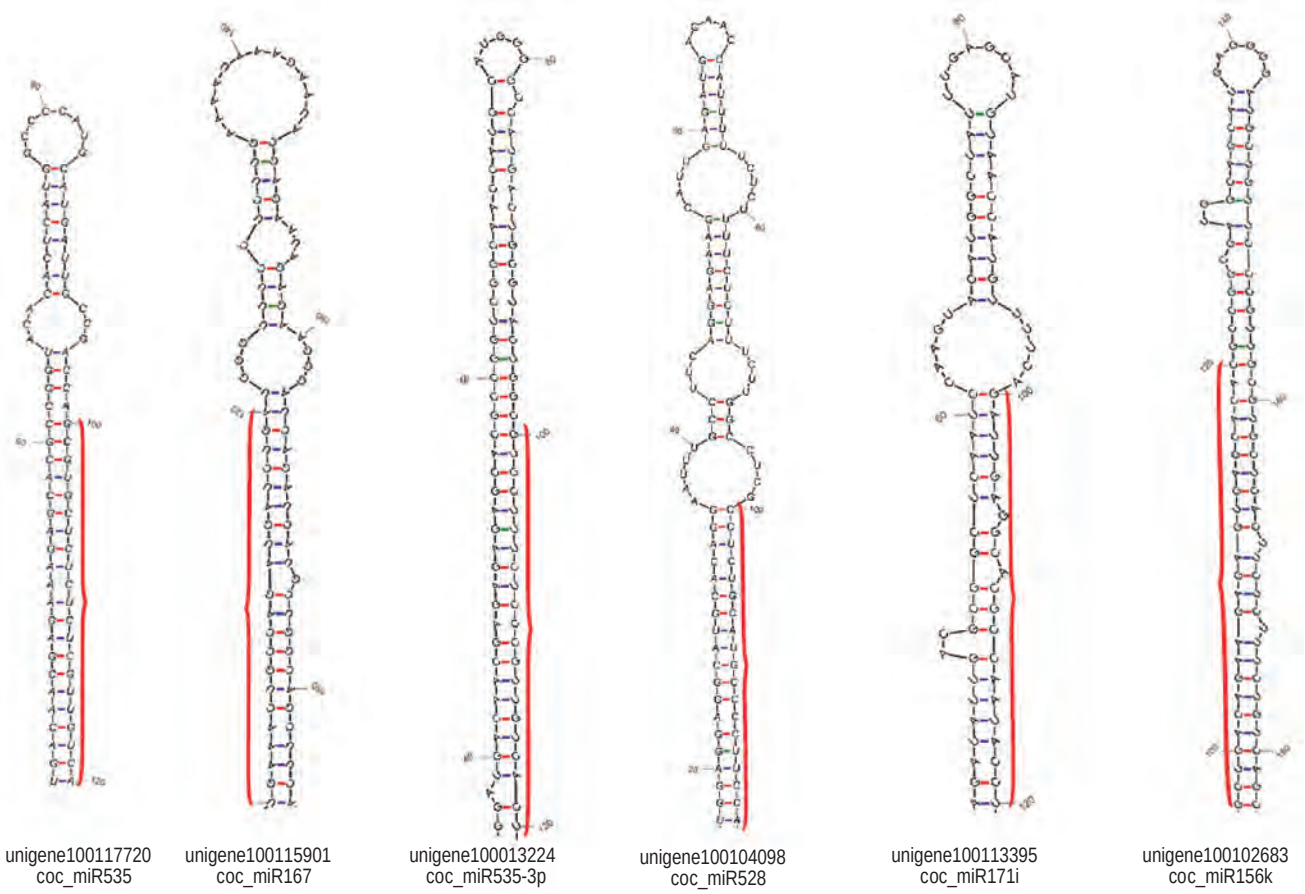

Fig. 2c. Predicted secondary structures of coc_miR535, coc_miR167, coc_miR535-3p, coc_miR528, coc_miR171i and coc_miR156k

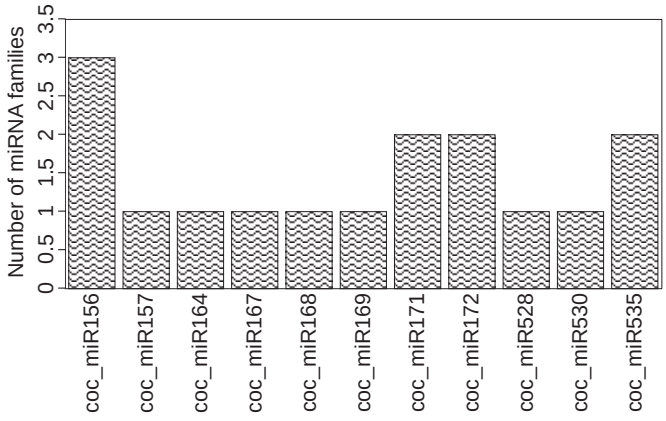

Fig. 3. Distribution of coconut miRNA in different miRNA families

structures each. The distribution of coconut miRNA into corresponding families is shown in Fig. 3. The putative target prediction of these miRNAs was carried out using psRNATarget tool and nine coconut miRNA were found to possess potential targets. The predicted target information is provided in Table 3.

Discussion

miRNAs have emerged as an important regulatory component expressed by most eukaryotic genomes, including plants. Their high evolutionary conservation has aided the prediction of newer miRNAs homolog's using *in silico* techniques. In this study, we have utilized an *in silico* approach to predict miRNAs and their targets in coconut leaf transcriptome. By using previously known plant miRNAs available at miRBase, we have identified 16 novel miRNAs, coming under 11 miRNA families. It is equally important to identify miRNA targets to assess the biological function of miRNA in plants. We have predicted targets for seven potential miRNAs in coconut. Some of the miRNA were found to possess multiple targets too. A majority of the predicted miRNA targets were coding genes for transcription factors.

Coc_miR157b was found to target many multiple proteins, two of them being phospholipases and galactolipases. Phospholipid catabolism has been reported to be essential to cell functions, which include membrane reorganization and degradation, production of secondary messengers and metabolic channeling of unusual fatty acids (Chapman, 1998). Phospholipid-derived molecules have been implicated to act as secondary signal messengers in plant signaling (Ryu, 2004). Yet another predicted target of coc_miR157b was RNA-recognition motif (RRM)-containing protein. Interaction of protein factors with specific RNA sequences is mainly responsible for post-transcriptional regulation of gene expression. A number of plant proteins that contain the principal RNA-binding domain, the RNA-recognition motif (RRM), have been identified (Sun *et al*., 2013) and they are mainly involved in different aspects of RNA metabolism, especially for plastid RNA editing in *Arabidopsis* and maize (Alba and Pages, 1998). Coc_miR157b was also found to target plasma membrane H^+-ATPases, which are the primary pumps responsible for the establishment of cellular membrane potential in plants. These enzymes, in addition to regulating basic aspects of plant cell function, are involved in signaling events in response to diverse environmental stimuli. Their activity is dynamically regulated, mostly during plant immune responses.

Table 3. Potential target of the identified coconut miRNAs

Coconut miRNA Id	Target Gene	Inhibition	Target protein
Coc_miR 157b	AT1G61850.2	Cleavage	Phospholipases; galactolipases
	AT5G25060.1	Cleavage	RNA recognition motif (RRM)-containing protein
	AT5G62670.1	Cleavage	H(+)-ATPase
	AT4G05400.1	Cleavage	Copper ion binding
	AT4G05400.2	Cleavage	Copper ion binding
Coc_miR 172a	AT3G09030.1	Cleavage	BTB/POZ domain-containing protein
	AT1G59171.2	Cleavage	Inositol-pentakisphosphate 2-kinase family protein
	AT1G59171.1	Cleavage	Inositol-pentakisphosphate 2-kinase family protein
Coc_miR 164	AT4G17710.1	Translation	Homeodomain GLABROUS 4
Coc_miR 535	AT3G20140.1	Cleavage	Cytochrome P450, family 705, subfamily A, polypeptide 23
	AT3G63200.1	Cleavage	PATATIN-like protein 9
Coc_miR 535-3p	AT1G04180.1	Cleavage	YUCCA 9
	AT5G43890.1	Cleavage	Flavin-binding mono-oxygenase family protein
	AT2G01040.1	Cleavage	Transposable element gene
	AT1G04610.1	Cleavage	YUCCA 3
	AT2G38470.1	Cleavage	WRKY DNA-binding protein 33
Coc_miR 171i	AT3G47170.1	Translation	HXXXD-type acyl-transferase family protein
	AT4G00150.1	Translation	GRAS family transcription factor
	AT3G60630.1	Translation	GRAS family transcription factor
	AT2G45160.1	Translation	GRAS family transcription factor
Coc_miR 156k	AT5G14460.1	Cleavage	Pseudouridine synthase family protein
	AT3G29175.1	Translation	transposable element gene

Many pathogens target H^+-ATPases during infection (Elmore and Coaker, 2011).

Coc_miR535-3p was found to target YUCCA genes which are mainly involved in auxin biosynthesis. Mutants of YUCCA genes in *Arabidopsis* have shown extreme flaws in vascular formation, floral patterning and various other developmental processes (Cheng *et al*., 2006). Yet another target of Coc_miR535-3p was patatin-like proteins. Recent results have indicated that patatin-related enzymes are involved in different cellular functions, some of which include abiotic stresses, plant responses to auxin, elicitors or pathogens and lipid mobilization during seed germination (Scherer *et al*., 2010).

The BTB/POZ domain, a target of Coc_miR 172a, is a widely distributed structural motif, which is evolutionarily conserved. They are found in a wide range of proteins with functions such as formation of voltage-gated channels, cytoskeletal organization and transcriptional regulation (Collins *et al*., 2001). The GRAS families of proteins, a target of Coc_miR 171i, have been proposed to play key regulatory roles as transcription factors in a wide range of plant processes ranging from shoot and root development, phytochrome A signal transduction and gibberellic acid signalling (Hirsch and Oldroyd, 2009). The target of Coc_miR 535-3p, WRKY proteins belong to a class of transcription factors involved in various plant processes chiefly to cope with various biotic and abiotic stresses (Pandey and Somssich, 2009). The target of Coc_miR 156k, pseudouridine synthases, are the enzymes responsible for posttranscriptional modification of cellular RNAs. They catalyze an isomerization reaction of specific uridine residues within an RNA chain (Mueller and Ferré-D'Amaré, 2009).

To conclude, we report the prediction of many novel miRNA's and their targets in coconut leaf transcriptome. The results of this study, the first of its kind in coconut, will contribute to an understanding of function of miRNAs in coconut. Future studies are required for their experimental validation and functional analyses.

Acknowledgement

We thank Director, CPCRI, Kasaragod for guidance and facilities. This work was supported by a grant (Sub-DIC) from DBT (Department of Biotechnology), Government of India.

References

Albà,M.M., and M. Pagès, 1998. Plant proteins containing the RNA-recognition motif. *Trends Plant Sci*., 3: 15-21.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.

Barbato, C., I. Arisi, M.E. Frizzo, R. Brandi, L.D. Sacco and A. Masotti, 2009. Computational challenges in miRNA target predictions: to be or not to be a true target? *J. Biomed. Biotechnol.*, doi:10.1155/2009/803069.

Bonnet, E., J. Wuyts, P. Rouzé and Y. Van de Peer, 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*, 20: 2911-2917.

Chapman, K.D. 1998. Phospholipase activity during plant growth and development and in response to environmental stress. *Trends Plant Sci*., 3: 419-426.

Cheng, Y., X. Dai and Y. Zhao, 2006. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev*., 20: 1790-1799.

Collins, T., J.R. Stone and A.J. Williams, 2001. All in the family: The BTB/POZ, KRAB, and SCAN domains. *Mol. Cell Biol.*, 21: 3609-3615.

Coronnello, C. and P.V. Benos, 2013. ComiR: Combinatorial microRNA target prediction tool. *Nucl. Acids Res.*, 41: W159-64.

Dai, X. and P.X. Zhao, 2011. psRNATarget: A plant small RNA target analysis server. *Nucleic Acids Res*., doi: 10.1093/nar/gkr319.

Elmore, J.M. and G. Coaker, 2011. The role of the plasma membrane H^+-ATPase in plant-microbe interactions. *Mol. Plant*, 4: 416-427.

Hirsch, S. and G.E.D. Oldroyd, 2009. GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.*, 4: 698-700.

Huang, T.H., B. Fan, M.F. Rothschild, Z.L. Hu, K. Li and S.H. Zhao, 2007. MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics*, 8:341

Jones-Rhoades, M.W. and D.P. Bartel, 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, 14: 787-799.

Kozomara, A. and S. Griffiths-Jones, 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39: D152-D157.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim , J. Lee, P. Provost, O. Radmark, S. Kim and V. N. Kim, 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425: 415-419.

Lee Y., K. Jeon, J.T. Lee, S. Kim and V.N. Kim, 2002. microRNA maturation: Stepwise processing and sub-cellular localization. *The EMBO J.*, 21: 4663-4670.

Li, Y.F., Y. Zheng, C. Addo-Quaye, L. Zhang, A. Saini, G. Jagadeeswaran, M.J. Axtell, W. Zhang and R. Sunkar, 2010. Transcriptome-wide identification of microRNA targets in rice. *Plant J.*, 62: 742-759.

Mueller, E.G. and A.R. Ferré-D'Amaré, 2000. Pseudouridine Formation, the Most Common Transglycosylation in RNA. Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience;. Available from: http://www.ncbi.nlm.nih.gov/ books/ NBK6205/2000.

Pandey, S.P. and I.E. Somssich, 2009. The role of WRKY transcription factors in plant immunity. *Plant Physiol.*, 150: 1648-1655.

Ryu, S.B. 2004. Phospholipid-derived signaling mediated by phospholipase A in plants. *Trends Plant Sci.*, 9: 229-235.

Scherer, G.F.E., S.B. Ryu, X. Wang, A.R. Matos and T. Heit, 2010. Patatin-related phospholipase A: nomenclature, subfamilies and functions in plants. *Trends Plant Sci*., 15: 693-700.

Singh J. and J. Nagaraju, 2008. *In silico* prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). *Insect Mol. Biol.*, 17: 427-436.

Sun, T., A. Germain, L. Giloteaux, K. Hammani, A. Barkan, M.R. Hanson and S. Bentolila, 2013. An RNA recognition motif-containing protein is required for plastid RNA editing in *Arabidopsis* and maize. *Proc. Natl. Acad. Sci. USA*, doi:10.1073/pnas.1220162110.

Voinnet, O., D.T. Gibbings, C. Ciaudo and M. Erhardt, 2009. Multi-vesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. *Nature Cell Biol.*, 11: 1143-1149.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31: 3406-3415.

Received: October, 2014; Revised: November, 2014;
Accepted: December, 2014