HortFlora Research Spectrum
Vol. 5, Issue 4; 306-309 (December 2016)
www.hortflorajournal.com
ISSN: 2250-2823

In silico IDENTIFICATION AND CHARACTERIZATION OF POTENTIAL miRNAs FROM *Capsicum annuum*

Naveen Duhan[*1], Navraj Kaur Sarao[1], Prashant Mohanpuria[1] and Mohinder Kaur Sidhu[2]

[1]*School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana*
[2]*Department of Vegetable Science, Punjab Agricultural University, Ludhiana*
**Corresponding Author' E-mail: naveen_duhan@pau.edu*

ABSTRACT : Recently, MicroRNAs (miRNAs) have been shown to be important regulator of genes in many organisms and have been implicated in a growing number of diseases. MiRNA are ~22 nt sequences. *C. annuum* is a well known plant in the world. A total of 1,18,578 Expressed Sequence Tags (EST) of *C. annuum* were mined from database of EST's (dbEST) and processed through Seqclean, 490 sequences were trashed and rest of 1,18,088 sequences were masked through RepeatMasker. Contigs were obtained by processing masked sequences through TGICL. A total no. of 25 putative microRNA's with significant similarity with the plant miRNA of closely related species of *C. annuum*. Majority of the predicted miRNAs were of 24, 23 and 22 nucleotides in length. The potential target of these miRNAs were miRNAs encoding enzymes regulating essential plant metabolic pathways including the putative transcription factor, oxygenases, disease resistance proteins, wound-responsive family protein, early E3 ubiquitin ligase, Rho binding family proteins and mostly are related to the responses to the biotic stresses and stress signaling in plants

Keywords : *MicroRNA, C. annuum, Seqclean, Repeat Masker, TGICL, Miranda.*

Capsicum annuum is a domesticated species of the plant genus *Capsicum* belongs to family solanaceae. It is the most common and extensively cultivated of the five domesticated capsicums encompasses a wide variety of shapes and sizes of peppers, both mild and hot, ranging from bell peppers to chilli peppers. The miRNAs are a class of small regulatory RNAs, which negatively regulate gene expression at the post transcriptional levels by binding target mRNAs for mRNA cleavage or inhibition of mRNA translation (Zhang *et al.*, 10). Many investigations have shown that miRNAs play an important role in a variety of biological and metabolic processes in plants and animals (Carrington and Ambros, 4; Ambros and Chen, 3; Zhang *et al.*, 12). In plants, miRNAs function to control tissue (leaf, root, stem, and flower) differentiation and development, phase switching from vegetative growth to reproductive growth, signal transduction, and the response to biotic and abiotic stress (*e.g.,* salinity, drought, and pathogens) (Chen, 5; Zhang *et al.* 10). Since, the first miRNAs were discovered in plants in 2002 (Park *et al.* 8; Reinhart *et al.* 9), several hundred miRNAs have been identified in plants by computational and experimental approaches (Zhang *et al.* 11). However, in Solanaceae, limited research has been attempted to elucidate the role of miRNAs in growth and development, moreover, putative target genes regulated by miRNAs have not yet been identified. The extensive evolutionary conservation of miRNA provides a powerful approach to their identification using comparative genomics. Using this strategy, we recently developed an expressed sequence tag (EST) and a genome survey sequence (GSS) approach to identify miRNAs (Pan *et al.,* 7). There are several significant advantages of using EST analysis for identifying miRNAs: (1) EST analysis can be employed to identify conserved miRNAs not only in model species, whose genomes have been published, but also in species for which only EST sequences have been determined; (2) EST analysis provides direct evidence for miRNA expression that cannot be inferred from genomic sequence surveys since EST are derived from transcribed sequences (mRNA) (Adams *et al.*, 1; Matukumalli *et al.,* 6; Altschul *et al.,* 2). Although several computational programs have been developed for predicting miRNAs, all these programs are based on genome sequences and require that these programs run individually on a computer; there is no any clue on their expression of miRNAs predicted by these programs (Zhang *et al.,* 11). Thus, the difficulty related to genome-based miRNA prediction is remedied by EST based analyses. Based on these three advantages, EST analysis will significantly

Article's History:
Received : 06-11-2016 ***Accepted : 24-11-2016***

enhance our ability to identify miRNAs and to investigate miRNA structure, function and evolution. Currently, 1,15,878 *Capsicum annuum* ESTs have been deposited in National Center for Biotechnology Information (NCBI). The number of miRNAs obtained by EST analysis depends on three factors: number of previously known miRNAs, conservation of miRNA sequence and structure, and the number of ESTs in the database. EST analysis makes it possible to rapidly study miRNAs and their functions in species for which the genome sequences are not well known, which would be impossible using traditional computational approaches.

MATERIALS AND METHODS

For identification of miRNA's in *Capsicum annuum* in-house pipeline was developed. A set of programs were combined systematically for the identification of miRNA's. EST sequences downloaded from dbEST (Expressed sequence database present at NCBI). Several filtering steps (Fig. 1) were applied in this EST based computational approach for reliable prediction of miRNA and their targets.

RESULTS AND DISCUSSION

A total of 25 miRNAs were identified using EST based computational approaches in Capsicum Annuum. These miRNA sequences were numbered accordingly (Table 1). Different characteristic values of these identified miRNA along with their potential targets were also recorded.

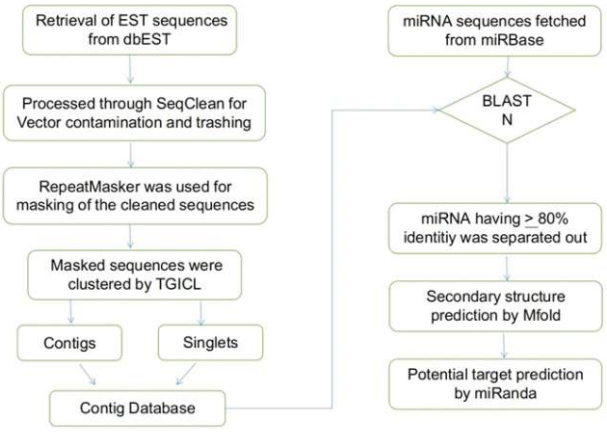

Fig. 1 : Computational approach for microRNA (miRNA) prediction in *Capsicum annuum.*

In silico identification of miRNA from *Capsicum annuum*

In the present study previously designed pipeline was improved. Thus, EST based computational approach was used with several filtering steps that identified the potential miRNAs in a sensitive and specific manner. In this approach after passing the 118578 raw EST sequences of *Capsicum annuum* through SeqClean, 118088 ESTs in output, 5,302 sequences in trimmed and 490 sequences in trashed with trash code were obtained. Seqclean output contained log file, .cln file and .clean file. Log file contained the summary of the process, .cln file contained trimming and trashing details and clean file contained only valid and trimmed sequences. The vector contaminating sequences were removed by SeqClean. Now this seqcl_sequences_fasta.clean file was the next input for RepeatMasker. It contained 118088 sequences and three main file were obtained *viz.,* .tbl, .masked and .out file, in which .tbl file showed the whole process and .masked file showed masked sequences. Through RepeatMasker 1.57% bases were masked. The masked sequences were processed through TGICL. In present approach, the most popular software tool, CAP3 for EST clustering was merged with TGICL. Through TGICL, 12070 contigs and 10515 singletons were obtained. Though, the unaligned assembled ESTs may represent low quality singletons and thus singletons can also be the source of some miRNAs. Thus, singletons were merged into contigs and finally the database of contigs was generated. The final contigs database was blasted with the mature miRNA sequences of Solanum family which are publicly available at http://www.mirbase.org/. Total 25 miRNAs of Capsicum annuum were identified based on their at least 80% identity (Table 1).

Characterization of Identified miRNA

In present study, our main focus was to identify and analyze some miRNAs related to responses to biotic stresses from *Capsicum annuum*. We computationally characterized all 25 predicted miRNAs. Different values such as size, A+U (%), MFEs, scores and hybridization energies of the predicted miRNAs were recorded in addition to their potential targets (Table 1). Out of 25 predicted miRNAs, approximate size of miRNA was 23-24 nucleotide. Other characteristic features of these mature plant miRNA such as sequence starts with 'U' and more than 50% A+U content, lower minimal folding free energy (MFE) and hybridization energy Zhang *et al.*, 11) were also recorded. Out of 25 miRNAs, majority of miRNAs

Table 1 : List of predicted miRNAs of *Capsicum annuum*, their sizes, minimum folding energy, hybridization energy and their predicted targets.

Identified miRNA	miRNA Sequence (5'-3')	Size (nt), A+U (%)	Minimum folding energy (MFE)	Hybridization energy (Kcal/mol)	Predicted targets
chilli-miR-1	UGCCUGGCUCCCUGCAUGCCA	21, 33.3	-4.0	-29.91	E3 ubiquitin ligase
chilli-miR-2	UCCACAGGCUUUCUUGAACGG	21, 47	-0.8	-20.77	Homeodomain-like superfamily protein
chilli-miR-3	UUGACAGAAGAGAGAGAGCAC	21, 52.8	3.5	-22.87	purine permease 4
chilli-miR-4	AGGCAGUGGCUUGGUUAAGGG	21, 42.8	-2.1	-21.70	O-methyltransferase 1
chilli-miR-5	UGUGGGUGGGGUGGAAAGAUU	21, 47	2.5	-22.29	RNA-binding (RRM/RBD/RNP motifs) family protein
chilli-miR-6	UGCCUGGCUCCCUGCAUGCCA	21, 33.3	-4.0	-29.90	E3 ubiquitin ligase
chilli-miR-7	CUCGGAGAGGGAAGAACGCGGUG	22, 36.36	-1.3	-25.58	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
chilli-miR-8	CCUAGGAAAUGAGAAAACUCGGC	23, 56.52	-1.7	-24.90	pentatricopeptide repeat 336
chilli-miR-9	UGAAAGACAAGGGUAGUGAGAUG	23, 56.52	1.4	-21.67	xyloglucan endotransglucosylase/hydrolase 8
chilli-miR-10	CAUCAGAAUUCAUGGAGGCUAG	23, 52.17	-0.3	-22.05	PAK-box/P21-Rho-binding family protein
chilli-miR-11	UCUUUGGUUGUUUAGCUAUAUGA	23, 65.2	-1.1	-21.58	NO TARGET
chilli-miR-12	UAAAGGCUGUUAUGCUUAUUUUUG	24, 70.83	-3.2	-20.81	Coproporphyrinogen III oxidase
chilli-miR-13	AAGUUUGGACUUAAAUUUGGUAAC	23, 73.91	-0.3	-27.47	dicer-like 1
chilli-miR-14	UGAGUAUUACAUCAGGUACUGGU	23, 60.87	-2.7	-24.95	NO TARGET
chilli-miR-15	UCGGGAAGGCAGCUGCGGCGGACU	23, 30.48	-1.7	-20.29	Coproporphyrinogen III oxidase
chilli-miR-16	UGAUUUGAAGUUUAUGAAUGUUGUA	24, 79.16	1.7	-29.49	dicer-like 1
chilli-miR-17	UGAAGAUGCACCAGAUGUUGAGAGGCC	27, 48.14	0.7	-26.49	NO TARGET
chilli-miR-18	UCGAAUUUUUACCCUUUUUUGGUCU	25, 68	-0.9	-24.01	purine permease 4
chilli-miR-19	CGACAACAACAACAAGAAGAAGAG	24, 58.33	0.0	-29.14	PPDK regulatory protein
chilli-miR-20	UAUAGAGAGCAGGAAGAUUAAUGU	24, 62.5	1	-30.99	purine permease 4
chilli-miR-21	UGGGGGAGCCAUGAGAUAAGAGCA	24, 45.83	-2.3	-24.87	Homeodomain-like superfamily protein
chilli-miR-22	UGAUUCGAACGUCACCAUCCAACA	23, 56.52	-0.6	-26.72	Coproporphyrinogen III oxidase
chilli-miR-23	GGAGGAGGUAGAGGGUGGUGGAAUU	24, 45.83	2.0	-20.40	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
chilli-miR-24	GAUGAAGAAGUGGAAGGAAGAAGA	24, 58.3	2.5	-26.29	E3 ubiquitin ligase
chilli-miR-25	UCACAACCUCCUUGAGUGAGUUGA	24, 54.16	-1.9	-20.14	Protein kinase superfamily protein

were having more than 50% A+U content, 16 miRNAs sequences were started with 'U' which shows their possibility of existence in the chilli plant (Table 1).

Based on the predicted targets of these miRNAs and other above mentioned characteristic features, several potential miRNAs were identified, which might be related to any disease response (Table 1). Further validation of these miRNA targets can be done experimentally.

Acknowledgment

Authors are thankful to Dr Parveen Chunneja to provide bioinformatics facility in School of Agricultural Biotechnology to carry out this work.

REFERENCES

1. Adams, M.D., Kelley J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao H., Merril, C.R., Wu A., Olde B., Moreno, R.F., Kerlavage, A.R., Mccombie, W.R. and Venter J.C. (1991). Complemen- tary DNA sequencing: expressed sequence tags and human genome project. *Science,* **252** : 1651–1656
2. Altschul, S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a newgeneration of protein database search programs. *Nucleic Acids Res.,* **25** : 3389–3402
3. Ambros, V. and Chen, X.M. (2007). The regulation of genes and genomes bysmall RNAs. Development **134** : 1635–1641
4. Carrington J.C. and Ambros V. 2003 Role of microRNAs in plant and animal development. *Science*, **301** : 336–338
5. Chen, C.F., Ridzon, D.A., Broomer A.J., Zhou, Z.H., Lee, D.H., Nguyen, J.T., Barbisin M., Xu N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J. and Gegler, K.J. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*, **33** : 179
6. Matukumalli, L.K., Grefenstette, J.J., Sonste-gard, T.S. and vanTassell C.P. (2004). EST-PAGE: managing and analyzing EST data. *Bioinformatics,* **20** : 286–288
7. Pan, X.P., Zhang, B.H., SanFrancisco, M. and Cobb, G.P. (2007). Characterizing viral micro RNAs and its application on identifying newmicroRNAs in viruses. *J. Cell Physiol*., **211** : 10–18
8. Park, W., Li J.J., Song R.T., Messing, J. and Chen X.M. (2002) CARPELFACTORY, a Dicer homolog, and HEN1, a novel protein, act inmicroRNA metabolism in *Arabidopsis thaliana. Curr Biol.,* **12** : 1484–1495
9. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. (2002). MicroRNAs in plants. *Genes Dev.,* **16** : 1616–1626
10. Zhang, B.H., Pan X.P., Cobb, G.P. and Anderson, T.A. (2006a). Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.,* : **289 :** 3–16
11. Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P. and Anderson, T.A. (2006b). Computational identification of microRNAs and their targets. *Comput Biol Chem* **30** : 395–407
12. Zhang B.H., Wang Q.L. and Pan X.P. (2007). MicroRNAs and theirregulatory roles in animals and plants. *J. Cell Physiol.,* **210** : 279–289.

❑

Citation : Duhan N., Sarao N.K., Mohanpuria P. and Sidhu M.K. (2016). *In silico* identification and characterization of potential miRNAs from *Capsicum annuum. HortFlora Res. Spectrum,* **5**(4) : 306-309.