Indian J. Genet. Plant Breed., (2022); 82(4): 469-473
doi: 10.31742/ISGPB.82.4.10

ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Computational identification of microRNAs and their target genes in sesame (*Sesamum indicum* L.)

P. Supriya*, Animesh Kumar[1], Sunil Archak[2] and K.V. Bhat[2]

Abstract

MicroRNAs (miRNAs) are non-coding small RNAs, which play a crucial function in specific biological and metabolic processes of plants and animals. MicroRNAs are very small sequences of 18 to 26 nucleotides length, which regulate gene expression through either inhibition of mRNA translation or mRNA cleavage by binding to target mRNAs. EST-based homology search, a powerful and comparative computational approach, is applied to detect conserved miRNAs in sesame. A homology search of previously identified miRNAs of Viridiplantae was performed against 16548 transcripts of sesame so as to detect novel miRNAs of sesame crop. In the present study, 4 potential miRNAs were detected through various stringent filtering criteria. Using psRNATarget server, 143 potential target genes were predicted which corresponds to the predicted miRNAs in sesame. The target genes tend primarily to encode DNA-binding proteins, transcription factors, stress-responsive genes and genes that regulate signal transduction. Results from the current study will hasten the way for understanding some mechanisms of post-transcriptional gene silencing and other functions of miRNAs in sesame.

Keywords: Gene ontology**,** MicroRNA, psRNAtarget server, transcripts

Introduction

Sesame (*Sesamum indicum* L.) is a diploid (2n=26) oil seed crop of the Pedaliaceae family and order Tubiflorae. Earlier studies on sesame were concentrated mainly on conventional breeding, quantitative genetics and genetic diversity among sesame germplasm collections (Laurentin and Karlovsky 2006). Nevertheless, the advancement of next-generation sequencing techniques resulted in a significant leap in sesame crop improvement. However, epigenetic mechanisms *viz*., DNA methylation, RNAi and histone modification have not been so far addressed in sesame. In particular prediction of miRNA and their target gene identification in sesame has been addressed to limited extent.

MicroRNAs are non-coding regulatory RNAs that usually derived from long sequences of self-complementing precursors. The length of small miRNAs ranges between 18–26 nucleotides and are usually conserved in evolution. MicroRNAs regulate gene expression based on the amount of complementarity between mRNA and miRNA by inhibiting translation or cleavage by binding to target mRNAs (Zhang et al. 2006). Several studies revealed that miRNAs regulate several biological and metabolic processes of plants and animals (Zhang et al. 2007). In plants, miRNAs tend to bind with mRNAs and regulate tissue differentiation, growth, and various abiotic and biotic stresses (Chen 2005). Since miRNAs are greatly conserved evolutionarily from species to species, a comparative genomics analysis across different taxa is possible. This function of widespread evolutionary conservation of miRNAs by comparative genomics offers a powerful method to their detection.

In the present study, we carried out EST based homology search of already reported miRNAs from Viridiplantae to identify potential miRNAs and their target genes in sesame. Furthermore, functions of the miRNA target genes were also

ICAR-National Academy of Agricultural Research Management, Hyderabad, India.

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India.

[2]ICAR- National Bureau of Plant Genetic Resources, New Delhi 110 012, India.

***Corresponding Author:** P. Supriya, ICAR-National Academy of Agricultural Research Management, Hyderabad, India, E-Mail: puramsupriya@gmail.com

How to cite this article: Supriya P., Kumar A., Archak S. and Bhat K.V. 2022. Computational identification of microRNAs and their target genes in sesame (*Sesamum indicum* L.). Indian J. Genet. Plant Breed., **82**(4): 469-473.

Source of support: Nil

Conflict of interest: None.

Received: May 2022 **Revised:** Oct. 2022 **Accepted:** Nov. 2022

investigated in order to understand the role of miRNAs in growth, development and egulation of abiotic and biotic stresses.

Materials and methods

Sesame transcripts, reference miRNAs

For the present study, the data has been received from ICAR-NBPGR in the form of transcripts. A total of 16548 transcripts of Swetha variety of sesame were screened against reported plant miRNAs available in miRBase (Release 22). From miRBase, 8410 plant miRNAs were retrieved to quest probable miRNAs in sesame unigenes.

Identification of potential miRNAs

Using BLAST-2.6.0 + (Altschul et al. 1990), the retrieved plant miRNAs were searched against sesame transcripts sequence database for homology. BLASTn algorithm was used for achieving the pairwise alignment of unigenes against known miRNAs with a word size of seven between query and database and a threshold of E-value at 0.01. Potential miRNA candidates were selected after homology search by considering the following criteria (Zhang et al. 2005):
(1) The predicted miRNA should have a length of at least 18nt, without any gaps within the sequence and (2) The mismatches allowed amongst predicted miRNAs and known reported miRNAs is taken as less than or equal to three.

After applying the stringent filtering criteria, the selected candidate miRNAs were further compared to the non-redundant (nr) protein database using BLASTx to discard any sequences that code for proteins. After removal of protein coding sequences, only sequences not coding for any proteins were compared against the Rfam database (Burge et al. 2012). The sequences with significant match were removed to minimize the possibility of putative RNA sequences, and the sequences with no match were subjected for further analysis.

Prediction of secondary structure of putative miRNAs

Secondary structure prediction was achieved through MFOLD software (Zuker 2003). The parameters used for secondary structure prediction of MFOLD are: The RNA sequence must be linear; folding temperature fixed at 37 °C; ionic conditions of 1 M NaCl with no divalent ions; maximum bulge/interior loop size of 30%; sub-optimality number should be 5 and parameters remaining were set by default.

In the current study, we used the following criteria, defined by Zhang et al. (2005) to select potential miRNAs after secondary structure prediction as described below.

(1) The pre-miRNA should consist a minimum length of 60 nt, (2) The pre-miRNA must be folded into a proper stem loop hairpin secondary structure, (3) The mature miRNA must be positioned in any arm of the hairpin structure, (4) The A+U content should be between 30–70% and (5) The secondary structure must have higher minimal folding free energy index (MFEI) and negative minimal folding energy (MFE).

Prediction of miRNA target genes in sesame

The predicted sesame miRNAs were used as query to identify potential miRNA targets in sesame available in the public domain using psRNATarget server (Dai and Zhao 2011) with following parameters.

(1) Maximum expectation value should be 3, (2) The central mismatch range leading translational inhibition should be 10-11 nt, (3) maximum number of mismatches allowed at the complimentary site must be ≤ 4 without any gaps and (4) Length for complementarity scoring = 20 bp.

For the potential miRNA target genes, annotation was performed through Blast2GO software and subsequently, GO terms were assigned. GO terms were used to identify corresponding roles of target genes among categories such as biological process, cellular component and molecular function. The detailed workflow for detection of potential miRNAs and their target genes in sesame is shown in Fig. 1.

Functional annotation of target genes

The putative role of target genes was identified through BLASTx analysis. The best hits were retrieved to confirm the role of target genes as well as miRNA regulated metabolic pathways. Gene Ontology (GO) terms were assigned to

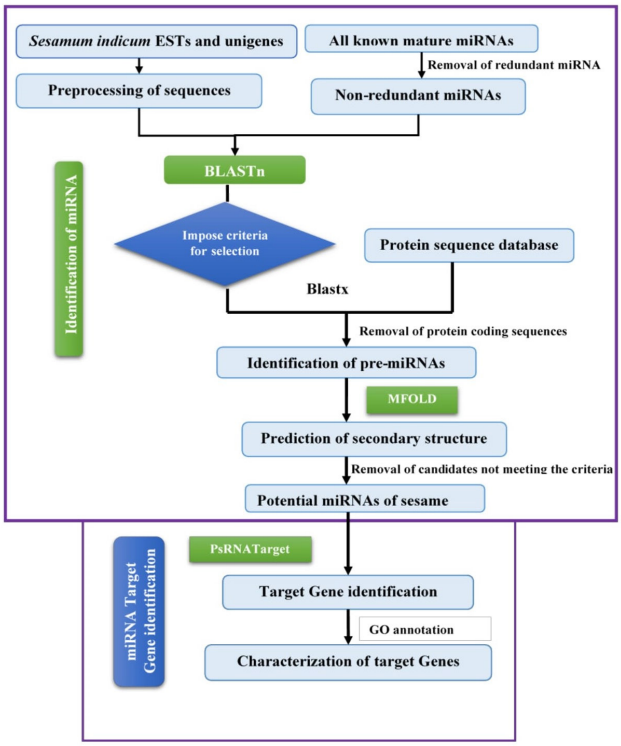

Fig. 1. Workflow for identification of potential miRNAs and their target genes in sesame

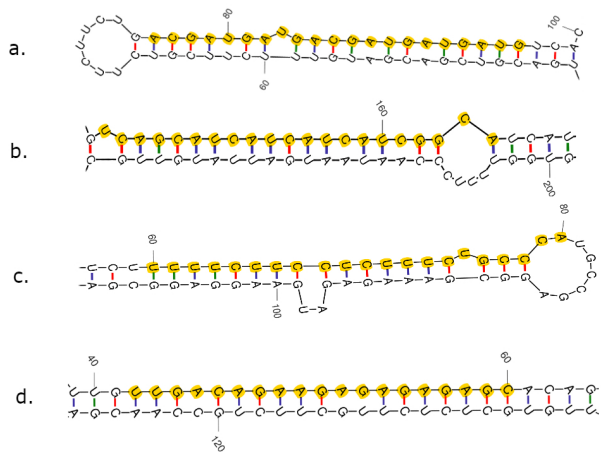

Fig. 2. Stem-loop structures of newly detected miRNAs in sesame. The mature miRNA is highlighted in yellow. (a) Sin-miR1, (b) Sin-miR2, (c) Sin-miR3 and (d) Sin-miR4

miRNA target genes through Blast2GO (Conesa et al. 2005) followed by functional classification by WEGO software (Ye et al. 2006).

Results and discussion

Detection of potential miRNAs in sesame

Analysis based on homology search has enabled the identification of conserved miRNAs in several plants such as maize (Zhang et al. 2006b), *Brassica napus* (Xie et al. 2007), tomato (Yin et al. 2008), soybean (Zhang et al. 2008), tobacco (Frazier et al. 2010), wheat (Han et al. 2009), potato (Zhang et al. 2009), Asiatic cotton (Wang et al. 2012), garlic (Panda et al. 2014) and coffee (Akter et al. 2014). In the current study, we could identify 89 unique sequences which have homology with already reported plant miRNA sequences. These sequences were screened to discard protein-coding sequences using BLASTx. Out of 89 probable miRNA candidates, 68 were coding for proteins and the remaining 21 sequences were subjected to further analysis. Secondary structure prediction was done through MFOLD for these 21 precursor sequences. The predicted structures from MFOLD were assessed to determine the appropriate hairpin stem loop structure by different filtration criteria. Four potential miRNAs were detected from sesame unigenes in the present study, shown in Fig. 2.

The length of predicted miRNAs was 21 nucleotides. The A+U content of miRNA precursors was ranging from 45% to 60%. Nucleotide content analysis of mature miRNAs revealed a prevalence of U (28.8%) followed by C (24.55%), A (23.89%) and G (22.3%). The different parameters viz., length of precursor miRNA (LP), length of mature miRNA (LM), adjusted minimal folding free energy (AMFE), minimal folding free energy index (MFEI) were calculated (Table 1).

Target gene prediction of sesame miRNAs

Target gene prediction is essential for understanding the functions of miRNAs in different biological, cellular, molecular and gene regulation mechanisms. Plant miRNAs tend to show perfect or near-perfect complementarity with the target genes followed by translation inhibition or cleavage (Schwab et al. 2005). In plants, miRNAs usually guide the cleavage of complementary mRNAs for instance microRNA-slicer activity guided by RNA-induced silencing complex (RISC) in Arabidopsis (Qi et al. 2005). In plants, many miRNA target genes which have disrupted miRNA activity as a result of mutation are likely to be expressed at higher rates (Boutet et al. 2003; Vazquez et al. 2004). Plant miRNAs mainly target transcription factors which normally affect plant growth and development indicating that miRNAs are key regulators. Several subsequent experiments have proved that miRNAs target genes involved in different physiological processes namely signal transduction, plant defense, stress response, root growth and development, besides small RNA biogenesis (Zhou et al. 2008; Nodine and Bartel 2010).

Target prediction of sesame miRNAs revealed that the newly detected miRNAs are targeting more than one gene and it is in agreement with the studies reported earlier (Jones-Rhoades and Bartel 2004; Zhang et al. 2006). The predicted sesame miRNAs are targeting genes that regulate transcription, proteolysis, ATP binding, signal transduction, growth and development. Apart from these, sesame miRNAs are also targeting the genes associated with response to stress, fatty acid biosynthesis and brassinosteroid mediated signaling pathway. Accordingly, target gene prediction in sesame ESTs shown that the newly detected miRNA families target genes related to a varied range of roles complementing previous studies such as the dehydration stress regulated by miRNAs in barley (Kantar et al. 2010), switchgrass miRNAs controlling genes related to signal transduction, stress response and plant growth (Xie et al. 2010).

Table 1. Characterization of detected sesame miRNAs

Sesame miRNAs	Mature miRNAs	Source	Homologous miRNA	LP	LM	MFE (−ΔG)	AMFE	MFEI
sin_mir1	ACGAUGAUGACGAUGAUGAUG	Unigene2152	ath-miR5658	171	21	41.99	24.55556	0.494
sin_mir2	UCAGCAUCAUCAUCAUCGGCA	Unigene4958	ath-miR414	175	21	55.08	17.48571	0.420458
sin_mir3	UUUUCUUCCUCUUUCUGCCCA	Unigene13545	aly-miR838-3p	155	21	46.3	29.87097	0.643056
sin_mir4	UUGACAGAAGAGAGAGAGCAC	BU667823.1	ahy-miR156c	137	21	57.23	41.77372	0.880462

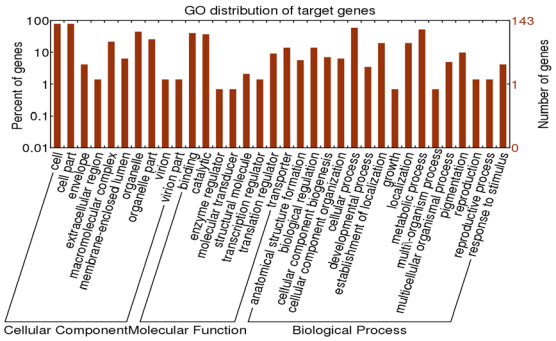

Fig. 3. Distribution of Gene Ontology terms for identified miRNA target genes in sesame

Furthermore, miRNA target genes in the present study are mainly transcription factors that typically affect plant growth and development and are in accordance with the reported studies (Yang et al. 2007; Yin et al. 2008). Sin-miR1 targets ion transmembrane transport, DNA binding, protein binding, lipopolysaccharide biosynthetic process, gibberellic acid mediated signaling and regulation of transcription whereas Sin-miR2 is involved in the regulation of protein serine/threonine phosphatase activity, transcription and translation. Propionate catabolic process, 2-methylcitrate cycle and ATP-dependent helicase activity genes were targeted by Sin-miR3. Sin-miR4 targets the genes involved in fatty acid catabolic process, acetate catabolic process and RNA bindingetc. In general, miRNA functions as a negative regulator of gene expression either by cleavage of target mRNA or repressing its translation, however, in plants, the cleavage of mRNA is likely to be a primary method of gene regulation (Sunkar et al. 2005). Results of the present study showed that a greater number of sesame miRNA target genes are predicted to be repressed by cleavage while very few of them are repressed by mechanism of translation inhibition.

GO term analysis of miRNA targets in sesame

Gene ontology (GO) term analysis of miRNA target genes was made to understand putative role of miRNA. In the present study, 143 miRNA target genes were assigned 37 GO terms in three ontologies namely, biological process (16 GO terms), molecular function (8 GO terms) and cellular component (13 GO terms) (Fig. 3).

The results of KEGG mapping revealed that miRNA target genes are involved in 28 metabolic pathways such as Lipopolysaccharide biosynthesis, Fatty acid degradation and Histidine metabolism etc.

EST based homology search was made in the present study to detect potential miRNAs in sesame. The newly detected four potential miRNAs belong to three distinct miRNA families viz., miR5658, miR838 and miR156. The detected miRNAs primarily target transcription factors, genes that are potentially involved in signal transduction, stress tolerance, lipopolysaccharide biosynthesis and fatty acid biosynthesis that are likely to provide miRNA mediated gene regulation in sesame. Overall our study results will lead to a deeper understanding of miRNA biogenesis and the mechanism of gene regulation.

Authors' contribution

The conceptualization of research (PS, SA, KVB); Designing of the experiments (PS, AK, KVB); Contribution of experimental materials (PS, SA, KVB); Execution of field/lab experiments and data collection (PS, SA, KVB); Analysis of data and interpretation (PS, AK, KVB); Preparation of the manuscript (PS, AK, SA, KVB).

References

Akter A., Islam M.M., Mondal S.I., Mahmud Z., Jewel N.A., Ferdous S., Amin M.R. and Rahman M.M. 2014. Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea arabica*). Saudi J. Biol. Sci., **21**(1): 3-12.

Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. 1990. Basic local alignment search tool. J. Mol. Biol., **215**(3): 403-410.

Boutet S., Vazquez F., Liu J., Beclin C., Fagard M., Gratias A. and Vaucheret H. 2003. Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr. Biol., **13**(10): 843-848.

Burge S. W., Daub J., Eberhardt R., Tate J., Barquist L., Nawrocki E. P. and Bateman A. 2012. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res., **41**(D1): D226-D232.

Chen X.M. 2005. MicroRNA biogenesis and function in plants. FEBS Lett., **579**: 5923–5931.

Conesa A., Gotz S., Garcia-Gomez J. M., Terol J., Talon M. and Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. J. Bioinform., **21**(18): 3674-3676.

Dai X. and Zhao P.X. 2011.psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res., **39**: W155–W159.

Frazier T.P., Xie F., Freistaedter A., Burklew C.E. and Zhang B. 2010. Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). Planta, **232**: 1289–1308.

Han Y., Luan F., Zhu H., Shao Y., Chen A., Lu C., Luo Y. and Zhu B. 2009. Computational identification of microRNAs and their targets in wheat (*Triticum aestivum* L.). Sci.China Life Sci., 52: 1091–1100.

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

Kantar M., Unver T. and Budak H. 2010. Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. Funct. Integr. Genomics, **10**(4): 493-507.

Laurentin H.E., Karlovsky P. 2006. Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet., **7**(1): 10.

Nodine M. D. and Bartel D. P. 2010. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev., **24**(23): 2678-2692.

Panda D., Dehury B., Sahu J., Barooah M., Sen P. and Modi M.K. 2014. Computational identification and characterization of conserved miRNAs and their target genes in garlic (*Allium sativum* L.) expressed sequence tags. Gene, **537**(2): 333-342.

Qi Y., Denli A. M. and Hannon G. J. 2005. Biochemical specialization within Arabidopsis RNA silencing pathways. Mol. Cell, **19**(3): 421-428.

Schwab R., Palatnik J.F., Riester M., Schommer C., Schmid M. and Weigel D. 2005. Specific effects of microRNAs on the plant transcriptome. Dev. Cell, **8**: 517–527.

Sunkar R., Girke T., Jain P. K. and Zhu J. K. 2005. Cloning and characterization of microRNAs from rice. The Plant Cell, **17**(5): 1397-1411.

Vazquez F., Gasciolli V., Crete P. and Vaucheret H. 2004. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not post-transcriptional transgene silencing. Curr. Biol., **14**(4): 346-351.

Wang M., Wang Q. and Wang B. 2012. Identification and characterization of MicroRNAs in Asiatic cotton (*Gossypium arboreum* L.). PLoS ONE, **7**(4): e33696.

Xie F., Frazier T. P. and Zhang B. 2010. Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*). Planta, **232**(2): 417-434.

Xie F., Huang S.Q., Guo K., Xiang A.L., Zhu Y.Y., Nie L. and Yang Z.M. 2007. Computational identification of novel microRNAs and targets in *Brassica napus*. FEBS Lett., **581**: 1464–1474.

Yang T., Xue L., An L. 2007. Functional diversity of miRNA in plants. Plant Sci., **172**(3): 423-432.

Ye J., Fang L., Zheng H. K., Zhang Y. and Chen J. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res., **34**: W293–W297.

Yin Z., Li C., Han X. and Shen F. 2008. Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). Gene, **414**: 60–66.

Zhang B., Pan X., Cannon C. H., Cobb G. P. and Anderson T. A. 2006. Conservation and divergence of plant microRNA genes. The Plant J., **46**(2): 243-259.

Zhang B.H., Pan X. and Stellwag E.J. 2008. Identification of soybean microRNAs and their targets. Planta, **229**: 161–182.

Zhang B.H., Pan X.P., Cobb G.P. and Anderson T.A. 2006b. Identification of 188 conserved maize microRNAs and their targets. FEBS Lett., **580**: 3753–3762.

Zhang B.H., Pan X.P., Wang Q.L., Cobb G.P. and Anderson T.A. 2005. Identification and characterization of new plant microRNAs using EST analysis. Cell Research., **15**: 336–360.

Zhang B.H., Wang Q. and Pan X. 2007a. MicroRNAs and their regulatory roles in animals and plants. J Cell Physiol., **210**: 279– 289.

Zhang W., Luo Y., Gong X., Zeng W. and Li S. 2009. Computational identification of 48 potato microRNAs and their targets. Comput. Biol. Chem., **33**: 84–93.

Zhou X., Wang G., Sutoh K., Zhu,J. K. and Zhang W. 2008. Identification of cold-inducible microRNAs in plants by transcriptome analysis. Biochimica et Biophysica Acta (BBA) - Gene Regul. Mech., **1779**(11): 780-788.

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