Rajesh Singh Tomar et al /J. Pharm. Sci. & Res. Vol. 9(5), 2017, 667-671

ISSN:0975-1459
Journal of Pharmaceutical Sciences and Research
www.jpsr.pharmainfo.in

In-Silico identification of miRNAs and their targets using Expressed Sequence Tags (ESTs) in Plants

Rajesh Singh Tomar, Gagan Jyot Kaur, Shuchi Kaushik, Raghvendra Kumar Mishra*

Amity Institute of Biotechnology,
Amity University Madhya Pradesh, Gwalior

Abstract

MicroRNAs (miRNA) are endogenous, non protein coding, small RNA molecules (21-24 nts in length) found in plants, animals, insects and viruses. They play a key role in transcriptional and post-transcriptional gene regulation. miRNAs perform their function by pairing with the complementary sequences of the target mRNA molecules, which results in silencing of gene in two ways, either by translational repression or target degradation. Various high-throughput methods are used for the identification of miRNAs. Methods like direct cloning as well as deep sequencing are widely used apart from the homology-based comparative-genome analysis methods which use expressed sequence tags (ESTs) and genomes survey sequences (GSS), which are easily available in the public databases. Compared with the other high-throughput methods, analysis through ESTs is a powerful and a better tool for identifying conserved miRNAs and studying the evolution of conserved miRNAs in several plant species, mainly for those species whose whole genome sequencing has not been done and not available. Various steps are involved in identifying the miRNAs and their targets using computational approach. The goal of this review is to summarize the process of in-silico identification of miRNAs and their targets using ESTs.

Keywords: miRNAs, high-throughput methods, in-silico identification, ESTs

INTRODUCTION

Since we have known the existence of genes, continuous efforts are been made to determine how the gene regulation of so many genes are controlled. Regulation of gene expression is surrounded by many ambiguities and there are many proteins which were found to control DNA translation and transcription and structural changes in different levels of genome organization. A short time ago, many scientists found that beside proteins, RNA also regulates the gene expression in the cell. miRNA is such type of regulatory RNA, which was first discovered in 1993.

MicroRNAs (miRNA) are endogenous, non protein coding, small RNA molecules (21-24 nts in length) found in plants, animals, insects and viruses. They play a key role in transcriptional and post-transcriptional gene regulation. miRNAs perform their function by pairing with the complementary sequences of the target mRNA molecules, which results in silencing of gene in two ways, either by translational repression or target degradation. It is estimated that human genome code for about 50,000 miRNAs, which may regulate different genes and are present abundantly in many human cells (Glaser et al. 2008).

Various researches have proved that miRNA has more functions along with repression of gene expression and stopping the protein translation. In few cases, miRNA binds to the promoter site of some genes and suppress expression of those genes (Place et al. 2008). Small RNA-induced gene activation, or RNAa is the term used for activation of gene expression using miRNAs.

Eukaryotic organisms have well conserved miRNAs and are considered as an important and evolutionarily part of genetic regulation. They play a vital role in growth and development of a plant (Bartel, 2004). Additionally, miRNAs may be considered as a good candidate for the early detection of various diseases (Pritchard et al. 2012).

Various high-throughput methods are used for the identification of miRNAs. Methods like direct cloning as well as deep sequencing are widely used apart from the homology-based comparative-genome analysis methods which use expressed sequence tags (ESTs) and genomes survey sequences (GSS), which are easily available in the public databases. Therefore, miRNAs identification and their target gene regulation is very important to understand their role in controlling key development processes in plants. Several programs are designed to search plant miRNAs using the whole genome sequences. However, the application of this strategy has been greatly limited because whole genome sequences of many plants are not available. Compared with the other high-throughput methods, analysis through ESTs is a powerful and a better tool for identifying conserved miRNAs and studying the evolution of conserved miRNAs in several plant species, mainly for those species whose whole genome sequencing has not been done and not available (Mishra et al. 2011, 12; Gangadhar et al. 2011, 14).

miRNA history

Victor Ambros et al., discovered miRNAs in 1993 in *C.elegans* by studying the role of *lin-4* gene in its development.

They discovered that *LIN-14* protein regulation was controlled by a short RNA product which encodes the *lin-4* gene. *lin-4* gene has a 61-nucleotide long precursor which matured to form a RNA which has 22-nucleotide length and whose sequences show partial complementarity to *LIN-14* mRNA sequences. This complementarity inhibits the translation of the *lin-4* mRNA into the *LIN-14* protein.

Hence, the first microRNA discovered was *lin-4* small RNA.

In 2000, second miRNA *let-7* was identified, which was responsible for repression of many different genes expression during various developmental stages of C. elegans life. *let-7* was further found to be conserved in large species, which indicates that wider phenomenon exists, related to the miRNAs.

Biogenesis of miRNA

Multiple steps are involved in the synthesis of miRNAs (Bartel, 2004). Same as proteins, the synthesis of miRNA takes place from transcription of DNA into RNA. The transcription product is longer than the final miRNA which is further processed by the Dicer-like 1 enzyme in plants, forming pre-miRNA having stem-loop structure (Tang et al. 2003). This]pre-miRNA moves out from the nucleus and travels into the cytoplasm with the help of exportin 5 protein. This pre-miRNA is processed by the RNaseIII-like enzyme called Dicer. Pre-miRNA is cut by dicer into two complementary structures, finally producing mature miRNAs. Two argonaute proteins, ALG-1 and ALG-2 bind to the miRNAs (Jannot, 2008). This leads to the formation of RNA-induced silencing complex (RISC) which is a multicomponent nuclease in which argonaute proteins are active RNase enzymes. After assembly, a RISC complex is formed. miRNAs bind to their complementary target mRNA and interrupts the process of translation. This fially leads to the reduction in the rate of expression of genes (Bernstein et al. 2001).

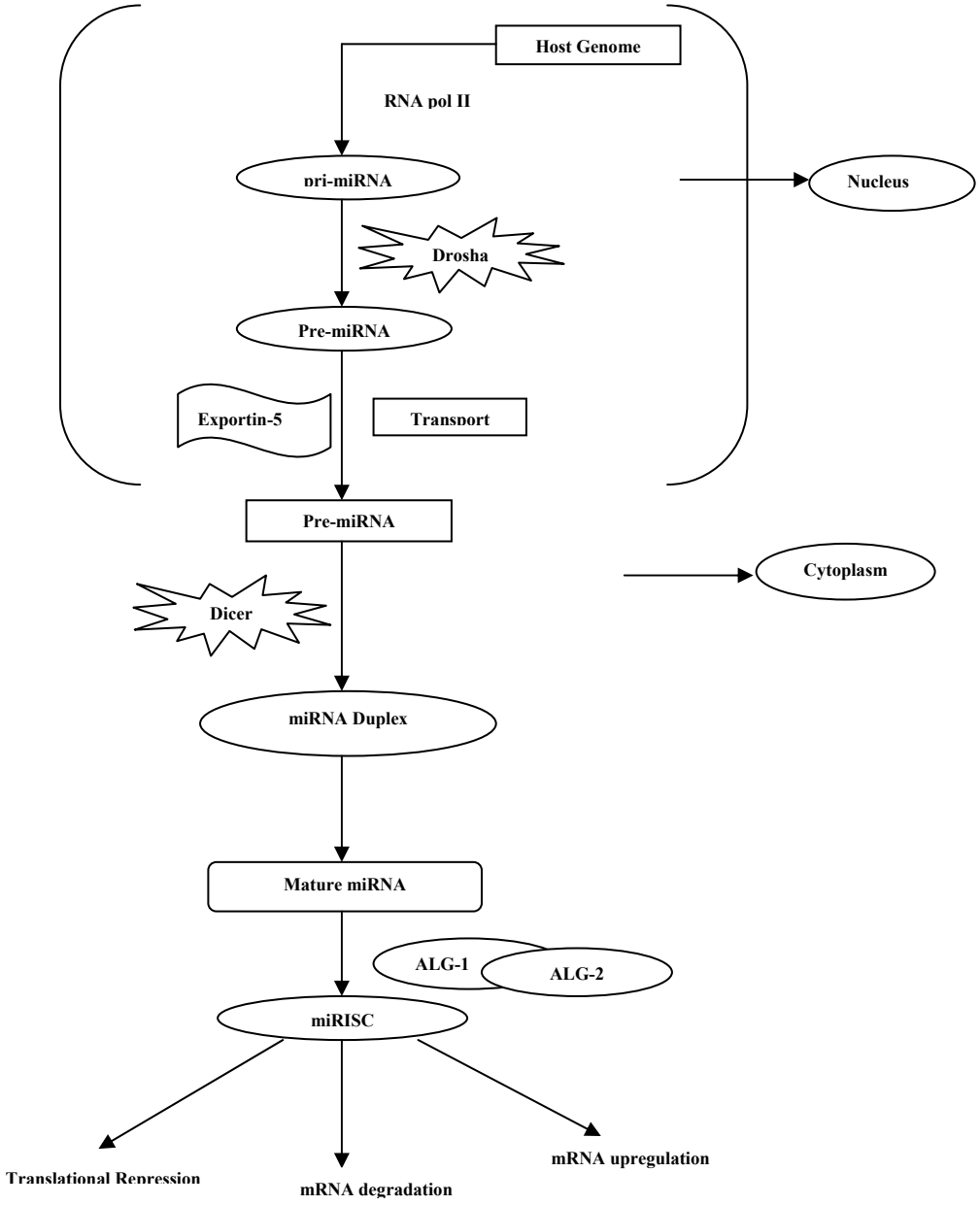

Figure 1: Schematic representation of biogenesis of miRNA

Mechanism of Gene Regulation by miRNA

miRNA functions by binding to the complementary region of mRNAs, leading to changes in the expression of gene. In eukaryotes, the miRNA binds with the 3' UTR of the target mRNA. Previously, it was discoverd that in addition to working through the 3' UTR of the mRNA, miRNA also targets some sites in the amino acid-coding regions and the 5' UTR between the 5' cap and the coding region (Glaser et al. 2008). miRNA binds to mRNA and blocks the translation of protein without destroying the mRNA, other studies revealed that miRNA can also degrade the mRNA. RNA interference is also responsible for the degradation of mRNA. This process involves the use of small interfering RNAs called siRNAs, which are short and double stranded molecules resembling to miRNA. The difference between siRNA and miRNA is that, siRNA is fully complementary to the mRNA target whereas miRNA is not completely complementary to the target mRNA (Elbashir et al. 2001).

In-silico Identification of miRNAs in Plants

miRNAs are crucial for the growth of plants in terms of leaf morphology, leaf polarity, formation of roots, signalling of hormones, change from juvenile phase to vegetative phase and then to flowering phase, time of flowering, identity of floral organs, reproduction etc (Sunkar et al. 2007). miRNAs play an important role in plant stress responses. This was proved by many recent studies which suggested that miRNA-mediated post-transcriptional regulation of target genes helps the plants to cope up with or bear the stress conditions (Matthew et al. 2006; Bartel, 2004).

Dehury et al. (2013) reported the in-silico techniques for the identification of miRNAs in Sweet Potato along with their target genes using their ESTs and most of the genes discovered were responsible for growth and development of plant, signal transduction and many other biological functions like defence and stress response etc.

The overexpression of miR156 in switchgrass resulted in various morphological changes and also the biomass production was improved (Fu et al. 2012).

In-silico methods were used for identifying miRNAs which are used in the biological synthesis of artemisinin, which is an anti-malarial compound produced by the plant *Artemisia annua* (Alvaro et al. 2012).

Ocimum basilicum (Basil) is a very popular herb for its medicinal properties. To identify miRNAs and their targets in *O.basilicum*, in-silico approach which involved the use of ESTs was used. The targets identified were regulating different functions in *O. Basilicum* (Singh and Sharma, 2014). miRNAs along with their targets were identified in *Rauvolfia serpentine* and their potential targets were found to be associated with regulation of different biological processes (Prakash et al. 2016).

Many such *in-silico* approaches have been made for the identification of miRNAs using computational methods.

Requirements for *in-silico* identification of miRNAs an their targets:

Plant miRNA and EST sequences of the plant whose miRNAs are to be identified:

To find out potential miRNAs in a plant, previously identified plant mature miRNA sequences are needed which can be downloaded from NCBI of *Oryza sativa*, *Glycine max*, *Zea mays*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Sorghum bicolor*, and other plants(http://www.mirbase.org/ Release 21.0, June, 2014). To avoid redundancy, the repeated plant miRNA sequences from different plant species should be removed, and the remaining miRNA sequences are used as query sequences for BLAST (Altschul et al. 1970) search against the plant EST database. The ESTs of the plant can be downloaded from the NCBI EST databases (http://www.ncbi.nlm.nih.gov/dbEST/).

Bio-informatic Tools:

BioEdit software (Hall, 1999), can be used to perform local blast of the plant's ESTs against miRNA database. BLASTX and BLASTN (http://www.ncbi.nlm.nih.gov/BLAST/) are done to search for miRNA homologs and to find the protein-encoding sequences, respectively.

The RNA folding algorithm available at Mfold server (Zuker, 2003), can be used for the prediction of the secondary structures of RNAs, the Minimal Folding Energy (MFE) and Minimal Folding Energy Index (MFEI).

The web tool psRNATarget (http://bioinfo3.noble.org/psRNATarget/) can be employed to predict the targets of newly identified miRNAs.

miRNA Identification:

Sequences of mature miRNAs of different plants are used for the prediction of potential miRNAs. Same methodology which was described by Zhang et al. (2005) is used for the identification of new miRNAs using EST analysis. Unique miRNA sequences are then employed as query for a BLASTN search against the EST database for miRNA homologs. All EST sequences which had less than four mismatches against the query miRNA sequences are selected. These sequences are used for a BLASTX search against the protein database at NCBI to remove the protein-encoding sequences. The pre-miRNAs region from 100 bp upstream and 100 bp downstream regions of the identified mature miRNA can be selected for further analysis. Stem-loop structures of the pre-miRNA sequences are generated online using the MFOLD 3.5 software (Zuker, 2003). MFOLD server with default parameters is used to predict the structures of the selected sequences. For the selection of pre-miRNAs following criteria can be used: (Ambros et al. 2003; Zhang et al. 2006) 1) less than four mismatches in the folding of the mature miRNAs; 2) the pre-miRNA sequence should fold properly into a stem-loop structure in which the miRNA sequence should be present in the stem; 3) upto 6 mismatches can be allowed between the mature miRNA sequence and the opposite miRNA sequence; 5) the folded structures of the identified pre-miRNA should have higher MFEI and negative MFE as compared to the other small RNAs.

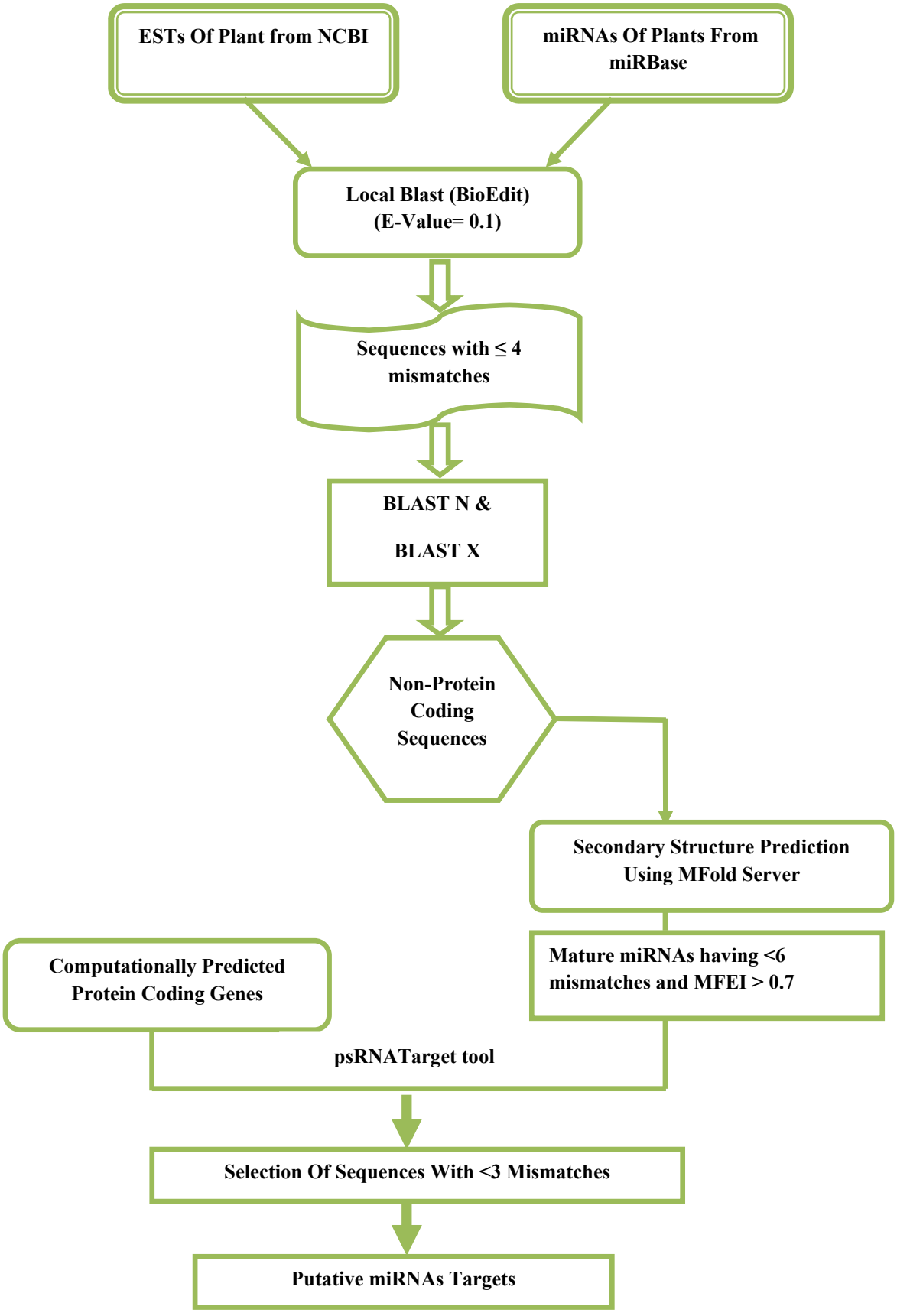

Figure 2. Schematic illustration of the workflow for identification of miRNAs and their targets in *A.barbata* ESTs.

Target Prediction of miRNAs

Targets of the newly identified miRNAs can be searched using the web tool psRNATarget. Available sequence library of various plants can be used for target search, with 3 set as the maximum expectation, not more than 4 mismatches between miRNA sequences and mRNA targets, and there should not be any gap at complementary sites.

REFERENCES

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool (1990). *J. Mol. Biol*. 215:403-410.
2. Álvaro L., Quintero P, Sablok G, Tatarinova TV, Conesa A, Kuo J, Lopez C. Mining of miRNAs and potential targets from gene oriented clusters of transcripts sequences of the anti-malarial plant, Artemisia annua (2012). *Biotechnol. Lett.* 34: 737-745.
3. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T. A uniform system for microRNA annotation (2003). *RNA.* 9: 277-279.
4. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and Function (2004). *Cell.* 116: 281–297.
5. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference (2001). *Nature*. 409, 363– 366.
6. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server (2011). *Nucleic Acids Res.* 39: W155-W159.
7. Dehury B, Panda D, Sahu J, Sahu M, Sarma K, Barooah M, Sen P, Modi MK. In silico identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) Expressed Sequence Tags (ESTs) (2013). *Plant Signal Behav*. 8:e26543.
8. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs (2001). *Genes Dev.* 15: 188-200.
9. Fu C, Sunkar R, Zhou C, Shen H, Zhang JY, Matts J, Wolf J, Mann DGJ, Stewart CN, Tang Y, Wang ZY. Overexpression of miR156 in switchgrass (*Panicum virgatum* L.) results in various morphological alterations and leads to improved biomass production (2012). *Plant Biotech J*. 10:443-452.
10. Gangadhar BH, Mishra RK, Park SW (2011). In silico EST-SSRs Development and Functional Validation in *Cicer arietinum*. Konkuk Journal.
11. Gangadhar BH, Tomar RS, Jyoti A, Shrivastava V, Kaushik S, Mishra RK (2014). Development of in-silico based molecular marker systems in medicinal plants of India: Future prospects. *Journal of European Academic Res.* 2: 581-611.
12. Griffiths-Jones S, Grocock RJ, Van DS, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature (2006). *Nucleic Acids Res.* 34: D140-D144.
13. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT (1999). *Nucl. Acids. Symp. Ser*. 41: 95-98.
14. Han J, Kong ML, Xie H, Sun QP, Nan ZJ, Zhang QZ, Pan JB. Identification of miRNAs and their targets in wheat (*Triticum aestivum* L.) by EST analysis (2013). *Genet. Mol. Res.* 12: 3793-3805.
15. Jannot G. Two molecular features contribute to the argonaute specificity for the microRNA and RNAi pathways in *C. elegans* (2008). *RNA.* 14: 829-835.
16. Jones-Rhoades MJ, Bartel DP. Computational identification of plant microRNAs and their targets, including a stress induced miRNA (2004). *Mol Cell.* 14:787-799.
17. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* (1993). *Cell.* 75:843–54.
18. Matthew W, Jones-Rhoades, David P, Bartel B. MicroRNAs and their regulatory roles in plants (2006). *Annu. Rev. Plant Biol*. 57:19–53.
19. 7) Mishra RK, Gangadhar BH, Kumar S, Park SW (2012). Development of EST derived SSR markers in pea (Pisum sativum) and their potential utility for genetic mapping and transferability to other legumes" Plant Bredding 131, 118—124.
20. 8) Mishra RK, Gangadhar BH, Yu JW, Kim DW, Park SW (2011). Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plant Omics. 4(3):154-162.
21. Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R. MicroRNA-373 induces expression of genes with complementary promoter sequences (2008). *Proc. Natl. Acad. Sci. USA*. 105: 1608-1613.
22. Prakash P, Rajakani R, Gupta V. Transcriptome-wide identification of *Rauvolfia serpentine* microRNAs and prediction of their potential targets (2016). *Comp. Biol. Chem.* 61:62-74.
23. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations (2012). *Nat. Rev. Genet.*13:358-369.
24. Singh N, Sharma A. In-silico identification of miRNAs and their regulating target functions in *Ocimum basilicum* (2014). *Gene.* 552:277-282.
25. Sunkar R, Chinnusamy V, Zhu J, Zhu JK. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation (2007). *Trends in Plant Sci.* 12**:**301-309.
26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0 (2013). *Mol Biol Evol*. 30: 2725-2729.
27. Tang G, Reinhart BJ, Bartel DP, Zamore PD. A biochemical framework for RNA silencing in plants (2003). *Genes Dev.* 17: 49–63.
28. Zhang B, Pan X, Cobb GP, Anderson TA. Plant microRNA: A small regulatory molecule with big impact (2006). Dev. Bio. 289: 3 – 16.
29. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction (2003). *Nucleic Acids Res.* 31: 3406- 3415.