

Computational Identification, Characterization and Analysis of Conserved miRNAs and their Targets in *Amborella Trichopoda*

Hajieghrari B^{1,2}, Farrokhi N³, Goliaei B⁴ and Kavousi K¹

¹Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

²Department of Agricultural Biotechnology, College of Agriculture, University of Jahrom, Jahrom, Iran

³Department of Agronomy and Plant Breeding/ Biotechnology Research Group, Faculty of Agriculture, Shahrood University, Shahrood, Iran

⁴Laboratory of Complex Biological Systems and Bioinformatics (CBB), Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

Abstract

MicroRNAs (miRNAs) are single stranded non-coding endogenous small RNAs of about 22 nucleotides, which are directly involved in regulating gene expression at post transcriptional level. miRNAs play key roles in development and response to biotic and abiotic stresses. Homology searches allow identification of new miRNAs due to their relative high conservation in plant species. Here, miRNAs were identified for *Amborella trichopoda*. Known and unique plant miRNAs from miRBase were BLAST-searched against Expressed Sequence Tag (EST) and Genomic Survey Sequence (GSS) in *A. trichopoda*. All candidate sequences with appropriate fold back structures were screened by a series of miRNA filtering criteria. Finally, we identified and analysed conservation of 5 potential conserved miRNAs belonging to 5 miRNA gene families from ESTs as well as 82 newly identified miRNAs belonging to 39 miRNA families from GSSs. Potential target genes of identified miRNAs were identified based on their sequence complementarities to the respective miRNAs using psRNA target against scaffold assignment of *A. trichopoda* genome sequences. Totally, 1219 target sites in *A. trichopoda* genome were identified. From which, 941 (77.19%) were predicted to be the subject of miRNA cleavage and 278 (22.81%) scaffolds were regulated via translational repression of mRNA. From the predicted miRNAs, 18 had no target sequence in *A. trichopoda*.

Keywords: Micro RNA (miRNA); *Amborella trichopoda*; Homology search; Target genes

Introduction

Micro RNAs (miRNAs) are a class of endogenous, single-stranded, non-protein-coding small RNAs that negatively regulate expression of varieties of protein-coding genes at post transcriptional level. This ancient evolutionary mechanism controls the expression by both targeting and cleavage of complementary mRNA, or in some cases by translational repression [1].

In animals miRNA genes are derived from introns, untranslated regions of transcripts or primary transcripts containing tandem precursors [2]. The animal miRNAs have multiple complementarily recognition sites, an imperfect complementarity that is located at 3'-untranslated region (3'UTR) of their targets. In plants, miRNA genes usually exist as independent transcriptional units and are transcribed by RNA polymerase II into long primary transcripts (Pri-miRNA) [3] with a sole and specific target site. The transcripts from miRNA genes are capped by adding a 5'-7-methylguanosine cap [4], spliced and polyadenylated at 3' end [5,6]. In plants, maturation of miRNA from pre-miRNA is processed in the nucleus by miRNA processing machinery. RNase III like protein DCL1 (Dicer like 1 enzyme), the core component of miRNA processing machinery, shapes the pri-miRNA into the precursor miRNA (pre-miRNA; the hairpin form of pri-miRNA) within the cytoplasm. The precursor miRNA folds back on itself to form a hairpin secondary structure. Mature miRNA sequences are located on one arm of this hairpin structure. Several paralogs of Dicer-like proteins are present in plants but only DCL1 incorporates in pre-miRNA processing. Other family members have evolved to protect plants against viruses [7]. The pre-miRNA is subsequently cleaved into a 22 bp double-stranded RNA. One strand is the mature miRNA, while the other comes from the opposite arm of the hairpin known as the miRNA* [8]. Similar to other types of RNA molecules processed by RNase III family of enzymes, miRNA duplex bears two protruding nucleotides at 3' end [9]. Subsequently, the miRNA/miRNA* duplex

is methylated by methyltransferase activity [9]. The processed 2'-O-methylated miRNA/miRNA* may be exported from nucleus to cytoplasm. This transport can be carried out either by HASTY (the Exportin 5 homolog in plants) -dependant or HASTY-independent nucleo cytoplasmic pathways resided on nucleus membrane [10-12]. Finally, the single-stranded mature miRNA is assembled into the ARGONAUTE 1 (Ago1) associated RNA-induced silencing complex (RISC) [13]. This complex is capable of binding to the complementary sequence of an mRNA molecule; the binding can be partial or complete. This results in either mRNA cleavage or translation arrest by miRNA, while the miRNA* is degraded. A given miRNA may have hundreds of different mRNA targets and it may cleave and/or repress the production of hundreds of proteins. An mRNA target can also be regulated by multiple miRNAs [14].

In animals, miRNA sequences have multiple complementarily recognition sites, located at 3'-untranslated region (3'UTR) of their targets. In plants, target mRNAs most contain one continuous complementary recognition site resulting in cleaving the target mRNA

***Corresponding author:** Goliaei B, Departments of Biophysics and Bioinformatics laboratories, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran, Tel: +98-216111-3356; E-mail: goliaei@ibb.ut.ac.ir

Farrokhi N, Department of Agronomy and Plant Breeding/ Biotechnology Research Group, Faculty of Agriculture, Shahrood University, Shahrood, Iran, Tel: 0098-21-44580; E-mail: nfarrokhi@nigeb.ac.ir

Received December 02, 2014; **Accepted** February 03, 2015; **Published** February 10, 2015

Citation: Hajieghrari B, Farrokhi N, Goliaei B, Kavousi K (2015) Computational Identification, Characterization and Analysis of Conserved miRNAs and their Targets in *Amborella Trichopoda*. J Data Mining Genomics Proteomics 6: 168. doi:10.4172/2153-0602.1000168

Copyright: © 2015 Hajieghrari B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and causing an immediate degradation. Additionally and less often, in plants, target site can be found in non-coding region (3'UTR or 5'UTR) of the transcript [15].

In recent years extensive research efforts have focused on identification of potential miRNAs in plant species. Both experimental cloning (construction and sequencing of small RNA libraries) and computational approaches have been used to identify plant miRNAs. Cloning or deep sequencing is limited to finding miRNAs by highly constrained tissue- and time-specific expression patterns. Since many miRNAs are highly conserved amongst species, homology-based search methods may be a useful task towards discovery of new counterparts with the capability of folding into hairpin secondary structures [16]. Additional criteria are necessary to set for distinguishing miRNA from other types of small RNA, reducing the number of false positives. Minimal folding free energy (MFE) and minimal folding free energy index (MFEI) are amongst such criteria; miRNAs have significantly higher negative MFE and MFEI than other RNA types [17,18].

A common approach for the identification of orthologs of miRNAs in other plant species would be through homology search in EST and GSS databases [19], especially for the species whose genomes are unknown or poorly understood. It has been generally accepted that mature miRNAs are conserved in plants from species to species in contrast to animals that miRNA precursors are usually conserved [5,20]. This feature of evolutionary conservation of miRNA allows for comparative analysis by using the available bioinformatics tools to search for putative miRNAs. Although large numbers of miRNAs have been identified in plants via this approach [21-25], less-conserved miRNA usually remained unidentified. Nevertheless, predicted miRNAs are needed to be experimentally characterized.

Amborella is a monotypic genus of rare understory shrubs or small trees endemic to the main island, Grande Terre, of New Caledonia.

The genus is placed alone in the family *Amborellaceae* and contains a single species, *Amborella trichopoda*. *A. trichopoda* has pivotal phylogenetic position since it has common ancestor with all other extinct angiosperms. To date, totally 124 *A. trichopoda* miRNA have been identified [26] and deposited at the miRNA database current version (release 21; June 2014). Due to the importance of *A. trichopoda* as the basal lineage in angiosperms clade in the evolutionary analysis, we applied an EST and GSS based-homology search to identify its potential miRNAs.

Materials and Methods

Data sets and software

All 7057 mature miRNA sequences of 73 plant species were downloaded from the miRNA registry miRBase (<http://www.mirbase.org>; release 21, June 2014) [27-29]. The repeated miRNA sequences were removed from the data set and only the unique ones were used as the reference set. *A. trichopoda* ESTs and GSSs were obtained from the relevant databases, which were available at nucleotide database of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>). BLAST-2.2.22 [30] was downloaded from NCBI website and set up locally. For prediction of the secondary structure of pre-miRNA and the free energy, Zuker RNA folding algorithm MFOLD 3.5 (<http://www.mfold.rna.albany.edu>) was used [31]. The miRNA target genes were predicted by the plant small RNA analysis server psRNATarget (<http://www.bioinfo3.noble.org/psRNATarget>) [32] and the plant Ensemble database (<http://plants.ensembl.org>).

Procedure and screening criteria for miRNA prediction

The workflow for prediction of the potential miRNAs is shown in Figure 1. After removing redundant sequences, all mature plant miRNA sequences downloaded from miRNA registry miRBase were

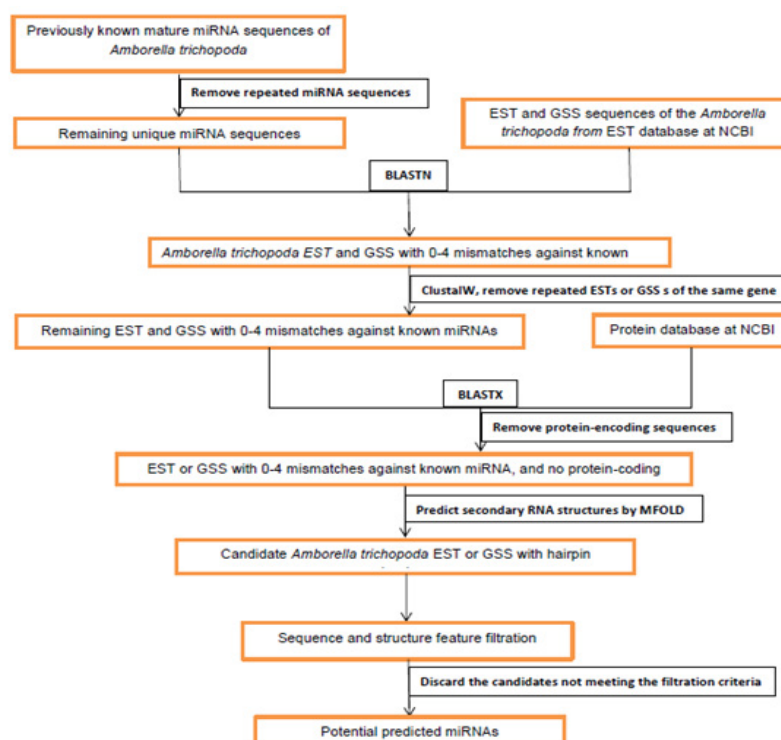


Figure 1: Procedure for the prediction of *A. trichopoda* potential miRNAs via EST and GSS homologs search performed on previously reported plant miRNAs.

used as queries in BLAST homology search against downloaded ESTs and GSSs. The default settings of BLAST parameters were used, except for maximum target sequence and expected threshold that were set to 1000 and 10, respectively. All ESTs and GSSs with no more than 4 mismatches were selected. The protein-coding sequences were removed from extracted sequences by searching against NCBI non-redundant (nr) protein databases using BLASTX (<http://www.BLAST.ncbi.nlm.nih.gov/BLAST.cgi>). The precursor sequence of 400 nt was extracted from each sequence by selecting 200 nt upstream and 200 nt downstream of each BLAST hit. If the length of the query was shorter than 400 nt, the entire sequence was used as a putative miRNA precursor. The secondary structures of putative pri-miRNAs were predicted using MFOLD 3.5 program. All parameters were set to default values. The A+U/T and C+G content and minimal free folding energy index (MFEI) were calculated according to [17] and based on the following equation: $MFEI = [(MFE / \text{length of the RNA sequence}) * 100] / (G+C)\%$. The sequences were considered as potential miRNA candidates if they met the following criteria: 1) Mature miRNA should be 18-22 nt in length. 2) The predicted pre-miRNA sequence folded into perfect or nearly perfect stem-loop hairpin secondary structure. 3) The potential mature miRNA sequence located on one arm of the hairpin structure. 4) No loops or breaks were allowed in the miRNA/miRNA* duplex. 5) Predicted mature miRNA sequence had less than 4 nt impairs with the miRNA* sequence. 6) No loops or breaks in miRNA* and miRNA sequences. 7) A+U content should be 30-75%. 8) The predicted pre-miRNA secondary structure had a high negative MFE (lower than -20kcal/mol) [22] and high MFEI values usually over 0.8.

Phylogenetic analysis of the miRNAs

Most plant miRNA and pre-miRNA are strongly conserved with high sequence identity even between distantly related species [20,33], belonging to the same families with low rate of evolution. Therefore, multiple sequence alignment of consensus structure of precursor sequences of the newly identified miRNAs with all previously verified members of each predicted miRNA family in the plant species (obtained from the miRNA database: <http://www.mirbase.org/>) (release 21, June 2014) [27-29] was constructed by the web based software Loc ARNA (<http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>) [34]. Thus, to obtain further insights on the evolutionary relationships of the newly identified miRNAs and their counterparts in other plant species, phylogenetic trees of the aligned sequences were constructed using average percentage identity. Moreover, to obtain further insights on the evolutionary relationships of the newly predicted conserved miRNAs and their counterparts in other plant species, phylogenetic trees of predicted pre-miRNA sequences were constructed based on average distance using percent identity in CLUSTALW (available online in EMBL/EBI; <http://www.ebi.ac.uk/Tools/msa/clustalw2>) and the trees were generated in Jalview 2.8.2.

Prediction of potential target genes in *Amborella trichopoda*

Predicted miRNAs were used as query against scaffold assignment of *Amborella trichopoda* draft genome sequence gene index using psRNA Target [32], an updated version of web-based miRU [6]. psRNA Target tool provides reverse complementary matching between miRNAs and their target transcripts and finds target site accessibility by calculating unpaired energy (UPE) necessary for opening the secondary structure around the miRNA target site [32] with the following criteria: 1) No gaps and no more than 4 mismatches are allowed between the mature miRNA and its potential target(s). 2) No mismatch was allowed between position 10th and 11th. 3) No more than

one mismatch was allowed at nucleotide positions 2nd -12th and up to three mismatches between positions 12th -15th. 4) No more than two consecutive mismatches were allowed [25]. Although homology based computational methods can cross species to identify conserved miRNA molecules, they fail to pinpoint rather unknown sequences. Having said this, due to advantages those methods can offer they have gained more popularity in recent years. Low cost and being capable of determining low abundant miRNAs as well as recognition of their temporal and spatial expression patterns are amongst some of their benefits.

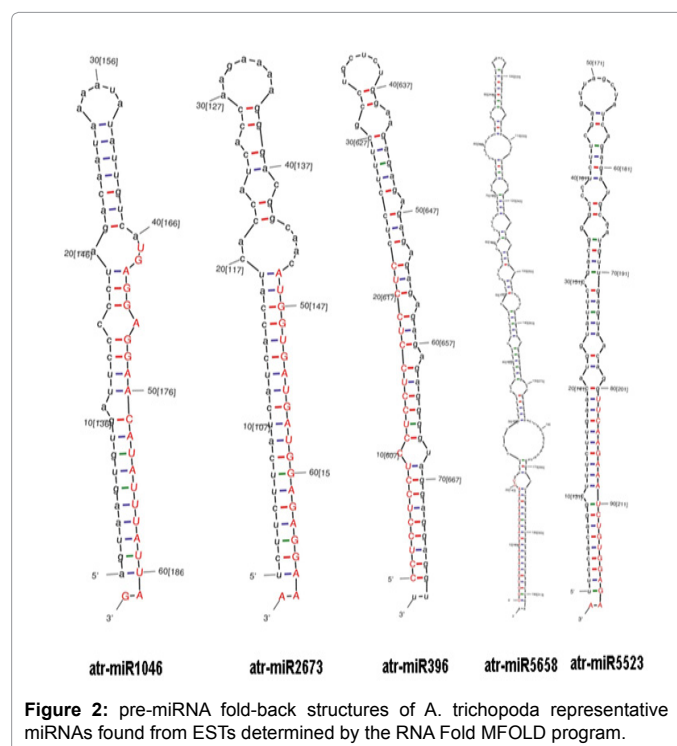
Results and Discussion

Identification of *A. trichopoda* potential miRNAs

Homology-based analyses were conducted by comparing *A. trichopoda* miRNAs presented as EST (26382) and GSS (72160) deposited in GenBank. miRNAs were identified according to the procedure depicted in Figure 1. Redundant sequences were removed from the data set. The remaining sequences were subjected to secondary structure prediction by MFOLD and inspected manually against filtering criteria as indicated to check for any discrepancies. The MFEI value, gold standard in differentiating miRNAs from other small RNAs, was established [35]. All found *A. trichopoda* sequences were in miRNA gene families. The miRNA gene families may give rise to mature miRNAs with one or more (up to 4) different nucleotides. Newly identified miRNA from *A. trichopoda* EST and GSS were named by the procedure of miRNA nomenclature proposed by miRBase [28].

Identification of miRNAs via EST search

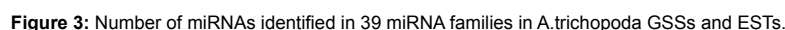
Potential miRNA genes (5) were detected in *A. trichopoda* ESTs (Table 1). These miRNAs fell in 5 different families. The length of identified precursor and mature miRNAs ranged from 61-192 nt (with an average 99.4 nt) and 19-21 nt, respectively. Despite the differences in precursor sequence length noted here, found miRNAs were predicted



Here it needs to be emphasized that GSSs are genomic in origin and therefore the predicted miRNAs require to be considered tentatively. However, the homology and structural data indicate that most derived pre-miRNA-like sequences from GSSs are genuine pre-miRNAs. These putative miRNA sequences may have the potential to be transcribed at various developmental stages of cell. Having said this, samples need to be prepared at different temporal and spatial stages or in response to biotic and abiotic stresses to determine if they are true miRNAs. Nevertheless, such global analyses of miRNA transcripts are cost-effective and time consuming and bioinformatics predictions may shed some light in targeted identification of miRNAs.

GSSs predicted miRNAs were diverse in structure (Supplementary file 2) and size, even if they were from the same family. Moreover, the distribution of GSSs identified miRNAs in each family was different (Figure 3 (Included as supplementary data)). In miR5057 family, although predicted *atr-miR5057a-1* and *atr-miR5057a-2* as well as *atr-miR5057d-1* and *atr-miR5057d-2* had the same precursor length and fold back secondary structure, their genomic positions were quite different. This can be explained either by miRNA duplications or via clone overlap. In contrast, other members of the predicted *atr-miR5057* family had different precursor miRNA sequence length with different fold back secondary structures that predicted to be in different genomic positions. Meanwhile, *atr-miR5057f* and *atr-miR5057g* were originated from both sense and antisense strands of

Using miRNA homology-based GSS analysis, following sets of strict criteria, total of 89 conserved miRNAs were detected in *A.trichopoda* that were classified into 36 families (Table 2, *Supplementary table S1*).



GGGAUUUCGCCUACCCUUCUCCACGUCCAAAGUUAUCGGACGGUACGAUAUCCGGUAGCAUUUCAUUUUUCAUUUUUUGUCUAUAAA(miR5057)(miR5568)GGAAUAGCUUAAAUUUUGGACGUGUGUUCUACAAAAUAGGUGAUUCACAAAAAUACAAAAACAUUUUUACUGAAA
AAAUUJGACCCAAUAUACCCUCAAAGUUUUGGAGGGUUAUUUGGUCAUUUUUGCUUACGAAUAUUUUUUUUUJGUGCGCCUCACUAG
AUGUACACCGCCUCCAAUUUCUAUUUCUGUAGUJ(stop)UUAAGAAAAAAAUAUAAAAUAGAACACUACCGGACAUUCGAACUGUGGGAU
GAACUUUGGACAGUGAUGAAGGUUAAGGUGAAUUCACUGUUAUUUUCGCCUUAUAGGCGUUGGGAAGACUUCUCCUUAUAUAUAUAUGA
GUGAUUUGAGUCAAUUUUGUGUAGUCAUAUAAGGAACAAUUAUGCAUAUAUUACUUGACCUUUCAUUUUCUAAUAAAGGCCUUAGGCC
CACCCAAAAGAAAGGAUAGGGUCUUAUCUGGUUAGACCUUUGGUUGGCCAACAAACUUCACAAAGAACGGUUCCAAUAUUUAUUCUUAUUG
GAGAGAACCAGGAGCAACUGUUUCAGAGGAUUAUGAGAGACAGAAGCUGUCUGAUCAUUGCACCCCAAUUUCAGGAGAGACCUUUCUGAU
GAAGAGUUAUCUUUUUCUAAACUCUCUUAUUCUUCUCCUUUUUGUUUCUAGCUCUUCUAGGUGAGGAGUGUUGUUGAGUCCUCCU
UGCCUCAAAGUACUUUCUCGGUCUACUUCUUAACUUGUUUCACUCUUAACCCCAACUUCUCCUUUAAAUCCGACCUCCGAGUGGCCG

J Data Mining Genomics Proteomics, an open access journal
ISSN: 2153-0602

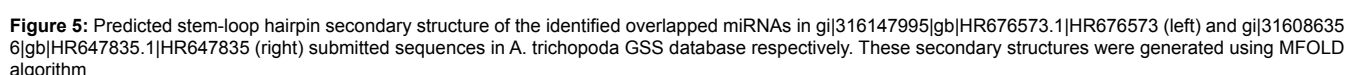


Figure 6: GSS (g|316147995|g|HR676573.1|HR676573) containing the overlap miRNA encoded within the same location. Red colored sequences represent the precursor sequences. Yellow highlighted sequence represent atr-miR169d-1 mature sequences and red colored underline sequences represent atr-miR5169e mature sequence.

the miRNA genomic loci with different precursor miRNA sequence. Both sense and antisense miRNAs seems to be transcribed from the same genomic locus; however the pri-miRNAs of both strands are separately transcribed from their own DNA template strands. Thus, they may be involved in different functions in plants [36]. Here, three pairs of other miRNAs in sense/antisense strands belonging to miR1533 (*atr-miR1533f* and *atr-miR1533g*), miR845 (*atr-miR845a* and *atr-miR845b*: with one ortholog in *Physcomitrella patens*) and miR2928 (*atr-miR2928a* and *atr-miR2928b*: with one ortholog in *Oryza sativa*) families were identified that appeared to have only one pair of sense/antisense miRNA in their families.

Newly identified *atr-miR5057g* and *atr-miR5568* are produced from the same precursor in the same direction on the same genomic locus (gi|316086356|gb|HR647835.1|HR647835; Figures 4,5). These so-called overlapped miRNAs are transcribed from the same pri-miRNA, but they may regulate different gene(s) similar to clustered miRNAs. Another overlapped miRNAs were also identified in miRNA196 family, which included 2 miRNAs belonging to the newly identified *atr-miR169d-1* and *atr-miR169e* with respect to a single nucleotide difference of their precursor and mature miRNA sequences on the same genomic locus (gi|316147995|gb|HR676573.1|HR676573; Figures 5,6). In contrast with animals that miRNA clusters have been identified widely, little reports are available for clustered miRNAs in plants [37-38]. Occurrence of miRNA clusters has been reported previously in the miR169 family [36,39]. However in this study we did not found miRNA clusters in miR169 family. Since the miR169 family has been observed in distantly related plant species, it can be said that it is amongst the most conserved miRNA families. We also identified a member of conserved miR171 family (*atr-miR171d*) that had three previously identified paralogs in *A. trichopoda*, a member of miR165 family (*atr-miR165*), a member of miR394 family (*atr-miR394*) and two members of miR395 family (*atr-miR395b* and *atr-miR395c*). These family members had several orthologs based on search at PMRD (Plant microRNA database) [40].

Moreover, newly conserved miRNAs (10) from miR1533 family were identified. In this family, 4 paralogs of *atr-miR1533e* were identified that had the same precursor sequence and secondary hairpin structure. For this family, only one ortholog (*gma-miR1533*) has been reported in *Glycine max* [41].

Another identified miRNA in *A. trichopoda* was miR529 family with 6 paralogs and several orthologs using PMRD. Meanwhile, 4 paralogs in *A. trichopoda* for miR5658 family were identified with an ortholog in *A. thaliana* reported in PMRD. *Ad infinitum*, 5 newly conserved miRNA genes were identified belonging to miR407 family with 3 orthologs in *A. thaliana* [42], *Zea mays* [43] and *Gossypium hirsutum* [44].

A miRNA gene may rise from multiple locations due to identical sequences of pre-miRNAs. For instance, miRNAs were reported for miR417 in three copies in *A. trichopoda* with two orthologs in *A. thaliana* and *Oryza sativa*. Furthermore, for miR5663 family 2 miRNAs were identified (*atr-miR5663a*, *atr-miR5663b*), which *atr-miR5663a* had 3 copies with one ortholog in *A. thaliana* [45]. Other newly identified miRNAs in *A. trichopoda* were *atr-miR1027* and *atr-miR1024*, each with two previously identified paralogs in *Physcomitrella patens*. Moreover, *atr-miR1446-1* and *atr-miR1446-2*, two identical copies of a sole miRNA appearing in different genome positions, have 5 orthologs in *Populoustrichocarpa* [46]. *atr-miR1514* had 2 orthologs in *Glycine max* and one in *Phaseolus vulgaris*. *atr-miR2082*, *atr-miR2616*, *atr-miR2634*, *atr-miR2919*, *atr-miR2931* and *atr-miR3445* each had one ortholog in *P. patens*, *Medicago truncatula*, *O. sativa*, and *A. thaliana*,

respectively. *atr-miR3512* and *atr-miR3520* were also identified, each with two orthologs in *Arachis hypogaea*. Last, *atr-miR5021a* and *atr-miR5021b* from miR5021 family had one ortholog in *A. thaliana* based on miRNAs registered in PMRD database for these species. *atr-miR5071* (with one ortholog in *O. sativa* registered in PMRD), *atr-miR5147* (with one ortholog in *O. sativa*), *atr-miR5183* (with one ortholog in *Brachypodium distachyon* registered in PMRD), *atr-miR5270* (with two orthologs in *Medicago truncatula* registered in PMRD), *atr-miR5568* (with one ortholog in *Sorghum bicolor* registered in PMRD), *atr-miR5649* (with two orthologs in *A. thaliana* registered in PMRD) and *atr-miR900* (with one ortholog in *P. patens* registered in PMRD) were identified by homology search in GSSs of *A. trichopoda*. The G+C contents of the predicted pre-miRNA sequences were 25.09% to 50.66% with an average of 33.84%. Mature miRNA sequences 16 out of 82 were started with uracil. Meanwhile, for 82 identified miRNAs, 51.22% were found to be located on the 5' arms of the stem loop hairpin while the rest (48.78%) were on the 3' arm (Table 2).

Conservation analysis in predicted miRNA families

Some miRNA families exist broadly in plant species and several miRNA families have multiple members within the same plant species. Different size of miRNA precursors usually results in a slightly different secondary stem-loop hairpin structure; this structure is often conserved within the same family. In this study, secondary structure constructive and structural alignment of the miR1533, miR5057, miR5663, miR529 and miR407 genes were built. Aligned precursors in each family showed some sequence similarity within their paralogs (Figure 7), but continuous conserved regions were not the same in the families. For miR5057 and miR407 families, paralogous sequences demonstrated to have stringent conservation patterns, as shown in peak of conservation (Figure 7).

Plant miRNA genes with high sequence similarity within their precursors are more likely derived from the same gene families. In these paralogs, based on duplication-mutation scenario, the gene expansion seems to have recent origin and this evolutionary pattern is probably an ongoing event. Diverged members of the same miRNA families may have been evolved at different rates within the same plant species and therefore they differ widely within and between species. Different regions in miRNA genes also seem to be under different evolutionary pressures, with higher level of conservation for the vital parts for processing and function as one expects. A growing body of evidence suggests that the mature miRNA sequence (preserved for both maintenance of the double strands and miRNA-target interactions) and its complementary sequence in the opposite arm of the fold back structure of pre-miRNA are the conserved regions [20]. While, other parts of miRNA precursors differ greatly. Here, multiple sequence alignment with a consensus structure of the families showed that the similarities not only lie on the mature miRNA/miRNA* regions, but also throughout the genes (Figure 7). Loss of similarity may be due to miRNA locus age [47]. Accordingly, it would be fair to state that the loci more likely have been generated by duplication of pre-existing miRNA genes in the same family. Moreover, they may share a common ancestry as proposed previously for several plant miRNA families [48]. Interestingly, little gene duplication was noted for some miRNA families; for instance in miR5057 and miR529 that they have 2 copies, and miR1533 has 4 copies. Phylogenetic analysis of the precursor miRNA paralogs in these families also provided additional evidence about the origin of the duplicated loci (Figure 8). In miR5057 family, *atr-miR5057a* had the same origin with other paralog supporting a probable recent duplication event. This evidence also was slightly seen in the miR529 and miR5663 families. Thus, the origin of some miRNA

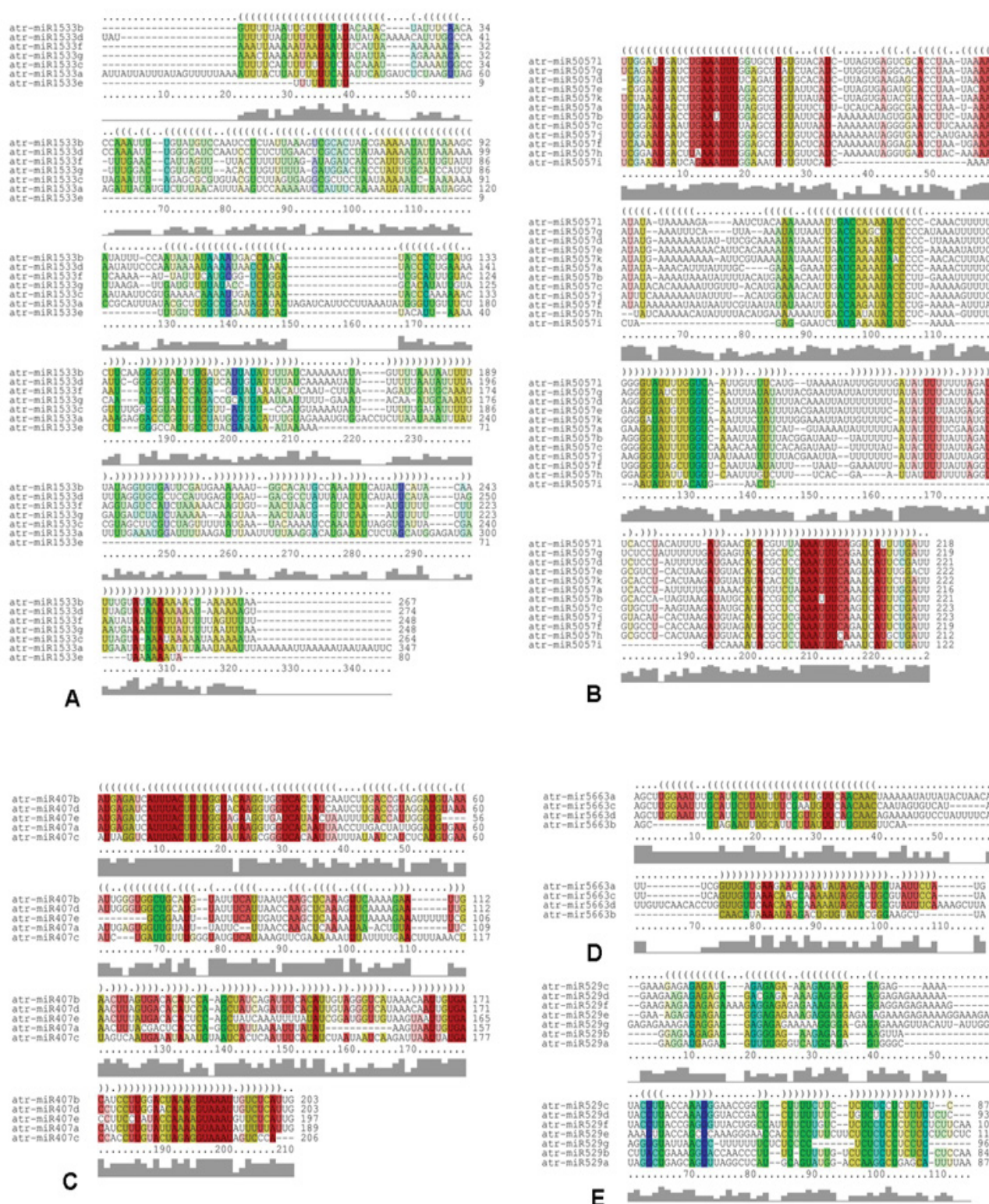


Figure 7: structural multiple sequence alignment of the precursor sequences of the newly and previously identified miRNAs in *Amborella trichopoda* in some conserved families using web-based computer software LocARNA. A: miR1533 family, B: miR5663 family, C: miR529 family, D: miR407 family and E: miR5057 family.

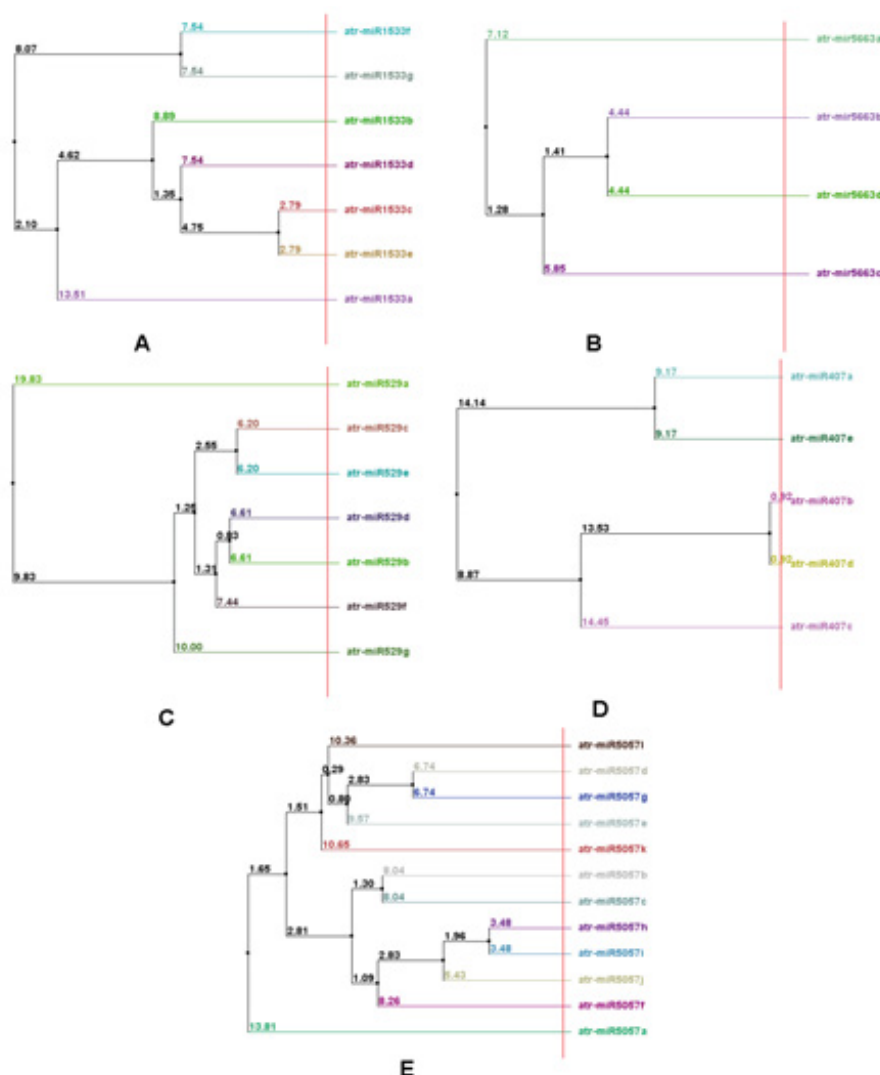


Figure 8: Phylogenetic tree obtained by aligning the precursor sequences of the newly and previously identified conserved miRNAs in *A. trichopoda* in some conserved families. The tree was constructed based on average distance using percent identity. A: miR1533 family, B: miR5663family, C: miR529 family, D: miR407 family and E: miR5057 families.

genes could be explained by duplication-mutation events whereby miRNA may evolve by duplication of a pre-existing miRNA.

Prediction of targets for identified miRNAs

Identification of targets for identified miRNAs is an important step for understanding the role of miRNAs and their various cellular functions via gene regulatory networks. As explained by various studies, plant miRNAs bind to the protein-coding regions of their mRNA targets with perfect or near perfect sequence complementarities, regulating gene expression by either cleavage of mRNA into two pieces or repression of translation [49]. This concept allows searching for target messengers via homology search approach. Here, miRNAs of *A.trichopoda* were subjected as query in psRNA Target against scaffold assignment of *A.trichopoda* genome sequence downloaded from Plant Ensemble. Totally, 1219 scaffold positions were identified in *A. trichopoda* genome. Amongst the relevant miRNA genes, 941 (77.194%) were predicted to be subjected to mRNA cleavage. The rest (278 scaffolds) seem to be regulated via translational repression such

as *atr-miR3512* and *atr-miR1044* (Supplementary Table S2). Based on the results, many miRNAs such as miR2082, miR2634 and miR529 families regulate several different positions on the *A. trichopoda* genome. Activities of the large number of identified scaffolds were also seen to be targeted by multiple miRNAs, similar to other recent studies [23,25]. Therefore regulatory function of miRNAs in biological and metabolic processes should be focused more and more in network concepts instead of individual connections between miRNA and its targets.

On the other hand, for some predicted miRNAs no target sequence was determined. No target miRNAs in *A. trichopoda* scaffolds were *atr-miR1533a*, *atr-miR1533e*, *atr-miR1533f*, *atr-miR1533g*, *atr-miR2919*, *atr-miR407c*, *atr-miR407d*, *atr-miR407e*, *atr-miR417a*, *atr-miR5057d*, *atr-miR5057k*, *atr-miR5057l*, *atr-miR147*, *atr-miR5663c*, *atr-miR5663d*, *atr-miR845a*, *atr-miR900a* and *atr-miR2673*. Lack of targets for predicted miRNAs was also reported in previous studies [23,25]. Targets-less miRNAs may target some positions in the genome

of invasive pathogens such as viroid and/or viruses that inject their genomes into the host plant [50,51]. Moreover and most likely, through evolutionary forces, genomes may have lost their corresponding target sites. Hundreds of thousands of hairpin structure in a genome exist that many of them appear to be derived from repeats, transposable elements and transposable inverted repeats as well as other evolving mechanisms that have the potential to form extremely stable miRNA like hairpin structure. It seems that newly emerged target-less miRNA is functionally inconsequential. Additionally, they are evolving much faster than that of so-called targeted miRNAs to generate seed sequences by chance. However, one needs to keep in mind that the genomic sequences that were taught are junk at one time, are not junk anymore. Accordingly, these non-specific newly emerged miRNAs may have unknown functions such as binding to some target sites (protein/nucleic acid) via an intermediate molecule(s) or get lost. The latter may be either an indication of the loss of target genes during the course of evolution or may be due to the activity of jumping genes. From earlier studies in maize by Barbara Mc-Clintock [52], it is evident that some transposons jump in a gene and once they leave, some parts of the gene that they have resided in get stuck to the transposable element and transfers to somewhere else in the genome. These movements may lead to the loss of target sequences for previously evolved miRNAs. Similar cases have mainly reported in animals [53]. In our study, similar cases were noted (TE Table 3). From the target-less miRNAs only *atr-miR1533a* and *atr-miR2673* genes contained simple repeat elements in their sequences and classified in TE-like miRNAs. However, from this class of miRNAs we identified miRNAs such as *atr-miR396f* that targeted some scaffold positions on *A. trichopoda* draft genome (Supplementary Table S2).

Conclusion

During the past few years, with the advent in availability of sequence resources in databases as well as computational based miRNA identification tools, considerable efforts have been made to predict new miRNAs by sequence and structure homologies search in ESTs and GSSs as well as identifying miRNA targets. In the present study, we performed ESTs and GSSs based-homology search to identify *A. trichopoda* potential miRNAs and their targets. Our data confirmed 5 miRNA genes identified from ESTs as well as 82 potential miRNA genes from GSSs belonging to 5 and 39 miRNA gene families that are homologs to previously deposited miRNAs in public miRNA databases. Also in this study we searched for the identified miRNAs target sites in *A. trichopoda* genome. We have found many miRNAs with no target site in *A. trichopoda* genome. Identifying miRNA and their target transcripts would be useful for other research concerned with the function and regulatory mechanisms of *A. trichopoda* miRNAs and will improve our knowledge about the miRNA mediated mechanisms regulating plant growth and development. Understanding miRNAs, their structures and target sites may deliver greater promise towards designing and engineering new miRNA molecules. Consequently, these newly developed miRNAs may be used in our task towards improving plant resistance/tolerance to biotic and abiotic stresses, eventually improving yield and well-being of crops.

References

- Zhang B Pan X, Cobb GP, Anderson TA (2006) Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* 289: 3-16.
- Berezikov E1 (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* 12: 846-860.
- Yang T Xue L, An L (2007) Functional diversity of miRNA plants. *Plant Science* 172: 42432.
- Xie Z Allen E, Fahlgren N, Calamar A, Givan SA, et al. (2005) Expression of

Arabidopsis MIRNA genes. *Plant Physiol* 138: 2145-2154.

- Bartel DP1 (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.
- Zhang Y1 (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res* 33: W701-704.
- Pashkovskiy PP Ryazansky SS (2013) Biogenesis, evolution, and functions of plant microRNAs. *Biochemistry (Mosc)* 78: 627-637.
- Dezulian T Remmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. *Bioinformatics* 22: 359-360.
- Chen X1 (2005) MicroRNA biogenesis and function in plants. *FEBS Lett* 579: 5923-5931.
- Bollman KM Aukerman MJ, Park MY, Hunter C, Berardini TZ, et al. (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* 130: 1493-1504.
- Park MY Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A* 102: 3691-3696.
- Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell* 136: 669-687.
- Hammond SM Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404: 293-296.
- Krek A Grün D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. *Nat Genet* 37: 495-500.
- Oliveira JS, Mendes ND, Carocha V, Craca C, Paiva JA, et al. (2013) A computational approach for micro RNA identification in plants: combining genome-based predictions with RNA-seq data. *J Data Mining Genomics proteomics* 4: 130.
- Jones-Rhoades MW Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14: 787-799.
- Zhang BH Pan XP, Cox SB, Cobb GP, Anderson TA (2006) Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci* 63: 246-254.
- Ng Kwang Loong S Mishra SK (2007) Unique folding of precursor microRNAs: quantitative evidence and implications for de novo identification. *RNA* 13: 170-187.
- Yin ZJ Shen FF (2010) Identification and characterization of conserved microRNAs and their target genes in wheat (*Triticum aestivum*). *Genet Mol Res* 9: 1186-1196.
- Zhang B Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. *Plant J* 46: 243-259.
- Li Y Li W, Jin YX (2005) Computational identification of novel family members of microRNA genes in Arabidopsis thaliana and Oryza sativa. *Acta Biochim Biophys Sin (Shanghai)* 37: 75-87.
- Wang X Zhang J, Li F, Gu J, He T, et al. (2005) MicroRNA identification based on sequence and structure alignment. *Bioinformatics* 21: 3610-3614.
- Das A, Mondal TK (2010) Computational identification of conserved micro RNA and their targets in Tea (*Camellia sinensis*). *American Journal of Plant Science* 1: 77-86.
- Galli V Guzman F, de Oliveira LF, Loss-Morais G, Körbes AP, et al. (2014) Identifying microRNAs and transcript targets in *Jatropha* seeds. *PLoS One* 9: e83727.
- Han J Xie H, Kong ML, Sun QP, Li RZ, et al. (2014) Computational identification of miRNAs and their targets in *Phaseolus vulgaris*. *Genet Mol Res* 13: 310-322.
- Amborella Genome Project (2013) The Amborella genome and the evolution of flowering plants. *Science* 342: 1241089.
- Griffiths-Jones S (2004) The microRNA Registry. *Nucleic Acids Res* 32: D109-111.
- Griffiths-Jones S Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140-144.

29. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36: D154-158.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
31. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.
32. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39: W155-159.
33. Han J, Kong ML, Xie H, Sun QP, Nan ZJ, et al. (2013) Identification of miRNAs and their targets in wheat (*Triticum aestivum* L.) by EST analysis. *Genet Mol Res* 12: 3793-3805.
34. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* 18: 900-914.
35. Lu Y, Yang X (2010) Computational Identification of Novel MicroRNAs and Their Targets in *Vigna unguiculata*. *Comp Funct Genomics*.
36. Patanun O, Lertpanyasampatha M, Sojikul P, Viboonjun U, Narangajavana J (2013) Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz). *Mol Biotechnol* 53: 257-269.
37. Zhang B, Pan X, Stellwag EJ (2008) Identification of soybean microRNAs and their targets. *Planta* 229: 161-182.
38. Xie F, Frazier TP, Zhang B (2011) Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*). *Gene* 473: 8-22.
39. Dhandapani V, Ramchiary N, Paul P, Kim J, Choi SH, et al. (2011) Identification of potential microRNAs and their targets in *Brassica rapa* L. *Mol Cells* 32: 21-37.
40. Zhang Z, Yu J, Li D, Zhang Z, Liu F, et al. (2010) PMRD: plant microRNA database. *Nucleic Acids Res* 38: D806-813.
41. Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, et al. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* 9: 160.
42. Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* 16: 2001-2019.
43. Zhang B, Pan X, Anderson TA (2006) Identification of 188 conserved maize microRNAs and their targets. *FEBS Lett* 580: 3753-3762.
44. Qiu CX, Xie FL, Zhu YY, Guo K, Huang SQ, et al. (2007) Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene* 395: 49-61.
45. Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, et al. (2012) High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in *Arabidopsis*. *Genome Res* 22: 163-176.
46. Lu S, Sun YH, Chiang VL (2008) Stress-responsive microRNAs in *Populus*. *Plant J* 55: 131-151.
47. Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, et al. (2010) MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22: 1074-1089.
48. Nozawa M, Miura S, Nei M (2012) Origins and evolution of microRNA genes in plant species. *Genome Biol Evol* 4: 230-239.
49. Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. *Trends Genet* 22: 268-280.
50. Pérez-Quintero AL, Neme R, Zapata A, López C (2010) Plant microRNAs and their role in defense against viruses: a bioinformatics approach. *BMC Plant Biol* 10: 138.
51. Maghuly F, Ramkat RC, Laimer M (2014) Virus versus host plant microRNAs: who determines the outcome of the interaction? *PLoS One* 9: e98263.
52. Ravindran S (2012) Barbara McClintock and the discovery of jumping genes. *Proc Natl Acad Sci U S A* 109: 20198-20199.
53. Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23: 578-587.