European Journal of Medicinal Plants
15(2): 1-14, 2016, Article no.EJMP.25221
ISSN: 2231-0894, NLM ID: 101583475

SCIENCEDOMAIN *international*
www.sciencedomain.org

In Silico MicroRNA Identification from *Stevia rebaudiana* Transcriptome Assembly

Aditya Mehta[1,2*], Hemant Gupta[3], Rakesh Rawal[4], Archana Mankad[2], Tanushree Tiwari[1], Maulik Patel[1] and Arpita Ghosh[1]

[1]Xcelris Labs Ltd., Ahmedabad, Gujarat, India.
[2]Department of Bioinformatics, Gujarat University, Ahmedabad, India.
[3]Eurofins Genomics India Pvt. Ltd, India.
[4]Medicinal Chemistry, Gujarat Cancer and Research Institute, Ahmedabad, Gujarat, India.

Authors' contributions

This work was carried out in collaboration between all authors. Authors AM and HG designed and the managed analysis of the study. Authors AM and MP performed the statistical analysis. Authors AM, TT and AG wrote the first draft of the manuscript. Authors ARM and RR supervised the work. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/EJMP/2016/25221
Editor(s):
(1) Shanfa Lu, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, China.
(2) Paola Angelini, Department of Applied Biology, University of Perugia, Italy.
(3) Marcello Iriti, Prof. of Plant Biology and Pathology, Dept. of Agricultural and Environmental Sciences, Milan State Univ., Italy.
Reviewers:
(1) Manoj Kumar Yadav, SVP University of Agriculture & Technology, India.
(2) Klára Kosová, Crop Research Institute, Prague, Czech Republic.
(3) Anonymous, Al-Farabi Kazakh National University, Kazakhstan.
Complete Peer review History: http://sciencedomain.org/review-history/14803

Original Research Article

Received 23rd February 2016
Accepted 17th May 2016
Published 27th May 2016

ABSTRACT

MicroRNAs are a class of endogenous, approximately 22 nucleotides in length noncoding RNA, which is evolutionary conserved and mediates post-transcriptional gene regulation. MicroRNA play a crucial role in development of plant, cellular processes, biological processes, cell proliferation and stress response. *Stevia rebaudiana* is an economically important and medicinal plant of the Asteraceae family. A total of 1,418,58 unigenes from *Stevia* transcriptome data were used for homology search against known plant miRNA database miRBase version 21. The functionally annotated unigenes were excluded from the studies. Total 381 non-protein coding unigenes were considered for candidates of miRNA precursor in *Stevia*. One potential miRNA from miR168 family with secondary structure was identified through the sequel of stringent filtering criteria. The target prediction of novel miRNA was carried out for using psRNATarget program based on their

**Corresponding author: E-mail: mehtaaditya99@gmail.com;*

sequence complementarities. A total of 31 potential gene targets were predicted for identified novel miRNA, which playing crucial role in various biological processes like development of plant, DNA repair, splicing, post-translational gene silencing, plant defense response, cell growth and proliferation. The phylogenetic analysis was also carried out to study the conserved nature of miRNA. These findings provide significant insights of miRNA and their potential role in S*tevia* as well as their regulatory mechanisms.

Keywords: miRNA; circos plot; psRNATarget; MFEI; MFE; EST; Stevia rebaudiana; small RNA.

1. INTRODUCTION

MicroRNAs (miRNAs) transcribes from RNA polymerase II promoters which located in non protein coding genes and non coding region of genomes [1,2]. MicroRNAs generated from spliced introns which are known as mitrons [3]. The biogenesis of miRNA is a complicated process which involves a series of events such as transcription, dicing, modification, nuclear export and RISC (RNAi-induced silencing complex loading [4]. The miRNA gene transcription is guided by RNA polymerase II, contains cap structures and polyadenylated pre-miRNAs with characteristic hairpin structure, processed by DCL protein. The pre-miRNA is further processed by dicer like-1 (DCL-1) into miRNA/miRNA* duplexes [5]. After dissociation from duplexes, miRNA is then loaded into Argonaute (AGO) protein to concoct the miRNA-induced silencing complex (miRISC). The miRISC directs its activity on target mRNA which depends upon perfect or imperfect complement between miRNA and their responsive elements and fate in miRNA cleavage or translational repression of the target mRNA [1,4]. miRNAs play key role as negative regulators in gene expression. MicroRNA binds to (UTRs) untranslated region or open reading frame (ORF) to mediate endogenous gene silencing. The majority of their targets are plant transcription factors which are involved in transcriptional control [3,6].

Stevia rebaudiana is the important medicinal herb of Asteraceae family. The steviol glycosides content in its leaves has medicinal properties. The two major active constituents of steviol glycosides are stevioside and rebaudioside-A which is a low caloric sweetening agent. Owing to secondary metabolites like alkaloids, phenol, flavanoids makes *Stevia* extract widely used as neutraceuticals, treatment of diabetes mellitus, obesity, hyper tension, beverages, and medicines [7-9].

The miRNAs are highly evolutionarily conserved among the species in plant and animal kingdoms. Many computational tools have been developed to identify potential miRNA in diverse organisms. There are many computational strategies for identifying miRNAs: (1) expresses sequence tag (EST) analysis, (2) genome survey sequence (GSS) approach to identify miRNA and (3) Computational identification of miRNA via homology search. In many species of plant, miRNAs are successfully identified by using computational approaches such as *Arabidopsis thaliana, Coffea arabica, Cyanococcus* (*Blueberry), Helianthus petiolaris, Jatropha curcas, Zea maize* and *Glycine max (Soybean)*. *Stevia rebaudiana* has enormous economical and medicinal significance but till date, no miRNAs of *Stevia rebaudiana* are submitted in miRBase and NCBI database. The whole genome sequence of *Stevia rebaudiana* is not available; however, few genomic information is available in NCBI databases including 470 proteins, 30,417 nucleotide sequences, 5,646 EST sequences. In miRbase, 18,226 miRNAs have been deposited from which 4,014 belongs to viridiplantae from 52 different species [10,11]. The transcriptome data of *Stevia* was downloaded from Sequence retrieval archive (SRA) of NCBI and assembled into transcripts. These transcripts were further used to predict known and novel miRNA. This approach resulted in one potential miRNA from *Stevia rebaudiana,* the targets for the same were predicted, playing important roles in various biological processes.

2. MATERIALS AND METHODS

2.1 Reference miRNA, Stevia ESTs and cDNAs

The transcriptome data of *Stevia* was downloaded from NCBI SRA accession SRR1576548 (http://www.ncbi.nlm.nih.gov/sra/?term=SRR1576548). A total of 30,038 nucleotide sequences and 5,646 EST sequences were retrieved from Genbank nucleotide database at NCBI.

2.2 Reference Set of miRNA

miRBase v21(http://www.mirbase.org/cgi-bin/browse.pl) was used which contains 28,645 hairpin precursors and 35,828 mature miRNAs from 223 species. A total of 8,496 known plant mature miRNA were extracted from total. The duplicate entries were eliminated by using cd-hit-v4.5.4 with identity value 100. As a consequence, 4,617 non-redundant unique viridiplantae miRNAs were obtained.

2.3 Pipeline of Bioinformatic Analysis for Transcript Assembly

The whole transcriptome data of *Stevia rebaudiana* were obtained from the NCBI SRA (Sequence Read Archive) with the link (http://www.ncbi.nlm.nih.gov/sra/?term=SRR1576548). The raw data were obtained in SRA format which was further converted to Fastq format using SRA Toolkit (version 2.4) (http://www.ncbi.nlm.nih.gov/Traces/sra/). The Trimmomatic [12] software was used for quality filtration of RNA-Seq raw reads and low-quality (QV <20) as well as adapter-contaminated sequences were discarded. High-quality reads were assembled de-novo using Bridger [13] assembler. The assembled transcripts subsequently clustered by using CD-HIT-EST [14] on default parameters for remove redundant transcripts. A total of 1,41,858 unigenes were used to predict against known plant miRNA for the homology.

2.4 Identification of Potential miRNA of *Stevia rebaudiana*

The pipeline for identification of potential miRNA from *Stevia rebaudiana* is described in Fig. 1. The known plant miRNA sequences were used as a query for homology search by BLAST-2.2.27 of NCBI Genbank against the assembled unigene as a reference database. The BLASTn was carried out at E-value 1000 and mismatch less than four. For candidate miRNAs, the length of mature miRNA sequences should not be less than 18nt and maximum of 24nt, secondly based on reported length of pre-miRNA, precursor sequences of 200 nucleotides were extracted (100 nt upstream and 100 nt downstream to the blast alignment) using the program FastafromBed which is an utility of BEDTools [15].

2.5 Secondary Structure Prediction

The non-protein coding unigenes were used for secondary structure prediction. The secondary structure of pre-miRNA sequences of potential miRNA homologs were predicted by the mfold web server (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) using RNA folding form with default parameters. The output of mfold includes ID numbers, respective miRNA homologs, total length of the sequences, the number of each nucleotide (A, G, C and U), the number of arms per structure, location of the matching regions, percentage (%) of (A + U/T) and (G + C) content and minimal folding free energy (MFE, Δ*G* in kcal/mol). The adjusted minimal folding free energy (AMFE) and the minimal folding free energy index (MFEI) were calculated according to the equation AMFE=(MFE/(length of a potential pre-miRNA))*100 and MFEI=((100 *MFE)/Length of RNA/(G+ C))%. The key features to be considered are: 1) The sequence of pre-miRNA precursors must fold into appropriate stem-loop hairpin secondary structure 2) mature miRNA sequence must be located within one arm of the hairpin structure 3) Loop should not break in miRNA sequences 4) No more than 6 substitutions in miRNAs and the opposite miRNA∗ sequence in the other arm 5) A+U content of miRNA should be 30-70% 6) maximum size for bulge in miRNA sequence should be 3 nt 7) predicted secondary structures must have higher negative folding free energy (MFE) and minimal folding free energy index (MFEI), the MEFI ranging between 0.65 to 0.85 [4,5,10,16].

2.6 Potential miRNA Target Prediction and Functional Annotation

Stevia rebaudiana assembled unigene, *Stevia rebaudiana* EST database and nucleotide databases were used as reference for identifying targets of the candidate miRNA by option "function User submitted small RNAs /user submitted transcripts" in the psRNATarget: A Plant Small RNA Target Analysis Server (http://plantgrn.noble.org/psRNATarget. Due to inadequate information about *Stevia rebaudiana* genomic data, the putative targets of novel miRNA were also predicted against Arabidopsis thaliana. The blastx analysis against non-redundant data set (nr database January 2015) of NCBI was carried out followed by gene ontology analysis using Blast2Go pro server [17] to distinguish biological and molecular functional. The plant transcription factor analysis was carried out for predicted target sequences using blastx against plant transcription factor database PlantTFDB.

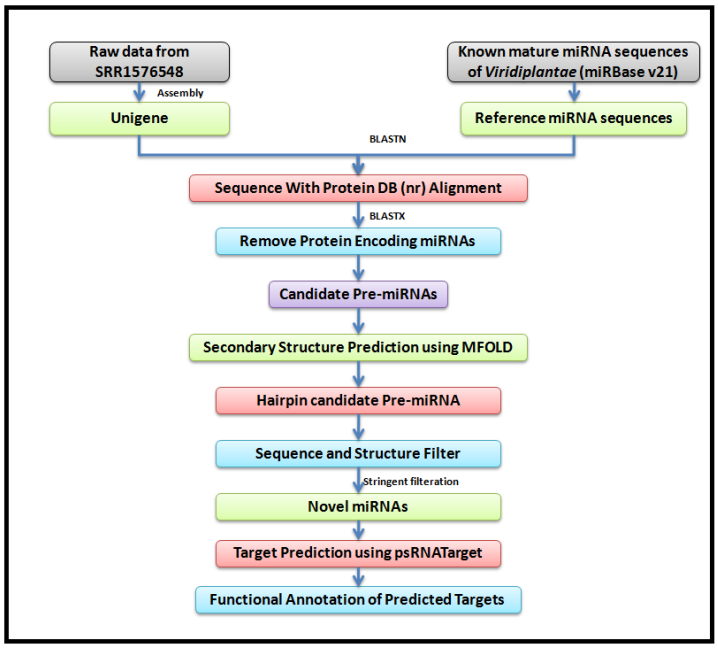


Fig. 1. Workflow for *Stevia* miRNA prediction

2.7 Circos Plot

The circos plot to visualize the novel miRNA and their corresponding target was done using circos (http://circos.ca/) [18]. Novel miRNA and their putative targets were "connected" using individually colored lines to indicate each miRNA's connection to their putative target genes.

2.8 Phylogenetic Analysis

The precursor sequences of pre-miRNA's were previously identified by BLAST. These precursor sequences of a same miR family were further aligned with predicted novel miRNA of *Stevia* using online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The phylogenetic analysis was carried out using MEGA6 tool [19]. Then, the Neighbor-joining tree was constructed to study their evolutionary relationships of miRNAs within the same family [3,4,5].

3. RESULTS

The 95,795,141 raw reads of *Stevia rebaudiana* transcriptome was filtered using Trimmomatic at QV> 20. The 91,044,283 high-quality reads were obtained and assembled into unigene using Bridger-de novo transcriptome assembler for RNA-Seq data which generated 1,41,858 unigenes. A total of 4,616 unique Viridiplantae miRNA sequences were aligned against assembled unigene sequences of *Stevia rebaudiana* leaf by using the standalone BLAST program, 4,616 *Stevia rebaudiana* unigenes showed perfect or near perfect match with reference miRNA. Subsequently, the BLASTX against NCBI non redundant protein database carried out for 4,616 unigenes for removal of protein coding sequences and other noncoding RNAs. Thus, we have obtained 381 non-protein coding unigenes which were considered as candidate miRNA. These miRNA precursor sequences were used for RNA secondary structure prediction using MFOLD server for validation process of predicted miRNA. To prevent false prediction of other RNA as a potential miRNA candidate, the crucial filtration, parameters were used. After filtration only one potential miRNA from *Stevia* was obtained (Fig. 2). This potential miRNA candidate encodes for miR168 family. The secondary structure of predicted miRNA was shown in (Fig. 3A, B) and criteria are listed in Table 1.

The target genes for putative miRNA were predicted using the psRNATarget program with default parameters against assembled Unigene of *Stevia*, *Arabidopsis thaliana*, this resulted in 31 targets. All the predicted targets of putative miRNA are listed in Table 2. The novel miRNA

and their corresponding target was visualized using circos plot (Fig. 4).

The Stv-miR is a member of miR168 family. The miRNAs belonging to family 168 were used for phylogenetic tree analysis. The miRNA sequences were subjected to multiple sequence alignment using Clustal Omega server (http://www.ebi.ac.uk/Tools/msa/clustalo/). The phylogenetic tree was generated using the NJ method using MEGA6 with a bootstrap value of 1000. The phylogenetic tree is divided into two major groups branching out from *Arabidopsis thaliana* miRNA168a (ath-miR168a) (Fig. 5). The group1 has multiple subgroups. In our tree, *Medicago truncatula* miR168 (mtr-miR168a) forming the early branch followed by *Oryza sativa* miRNA 168b (osa-miR168b) and *Aquilegia caerulea* miR168 (aqc-MIR168) clustered into a single group followed by *Solanum lycopersicum* miRNA 168a(sly-miR168a), *Oryza sativa* miRNA 168a(osa-miR168a), *Zea mays* 168a(zma-miR168a), *Brassica rapa* miRNA 168a(bra-miR168a) and *Arabidopsis thaliana* miRNA168b(ath-miR168b) clustered into single group with good bootstrap support. In group2, *Stevia* miRNA (Stv-miR) groups with *Saccharum officinarum* miRNA 168b (sof-miR168b) and *Brassica rapa* miR168b(bra-miR168b) with good identity [20].

A Gene Ontology enrichment analysis was conducted to identify the functional categories of predicted targets. The Blast2GO (www.blast2go.com) program was used to categorize 27 functional groups under three main categories: Biological process (BP), cellular component (CC) and molecular function (MF) (Fig. 7). A total of 21, 14, and 21 targets were annotated for biological process, cellular component and molecular functions, respectively (Fig. 6) [21].

```
>Unigene
CCAGTCAAAGCCCTCGACAACCTTCGCCCACGTCTTTTTACTCTTCCACCCTTCTCACTATCTTTAACA
CCGCCTTATACCAGAGGCATTCTCGTCAATCCATATAACGGAATTCCGTTCTCCCACATCCCACTCCCA
ATCTGCAACAACCAGTTGCCGCTAAATAAATTCGCCAGATTTGCCCCAAACCCTCTATGCCACCTCCAC
GTGTCACAACACCATTGGTCTGTTTACACTCTAAATAACCTTCACTTCTCCGTCCCTTACTGAAAGATA
GGGTTCATCTTCTTCATCCTCTCTCTCTTTATCTGGAAAATTTACGCTTTATATAACATGCAGATCTAA
TGATTTTGTATCTATACGTATTGACTTTGTGTCAGTGATTGACTGATTGTGGATTGAGATGACGGATGA
AGCGATTTGTGTTTCTGTTGTTACGGATGTTTCTGATTCGCTTGGTGCAGGTCGGGAACGGT
ATTACCGGCGTGGATCGTAATTGTAATACTTGAGGATAATGATTCCGGTGAACGACGGTCCCGCCTTGC
ATCAAGTGAATTGGAAACTGCTGTGATTGTTTGGAGTTATCGCAGTGGATGATCAATGATGGAGATCTT
TCATCCGGATTCGAAAAGGTTAATTGATTTGTAATTTGTATACGCTATTGAACTTCGATTTCGTGGTTT
GATTTTTATTGTAGATGTTTATTGATTTTTGAATCTCAATAGGAAATTAGGAATTGGAATGAATGAATG
ATTGATGGCGCTTTACGTTTTGATTTGTGAGTTGAATAATTGATAAATGAAATTTTGCTGTCTATATTA
TGTGTCAATCAGATGCTAACTGTTTTGTGTTAATGAAGAAAATAGA
```

Fig. 2. Unigene sequence containing miRNA. Green color highlighted sequence shows pre-miRNA precursor and Blue color highlighted sequence shows mature miRNA

Table 1. Details of newly predicted miRNA in Stevia

Properties	Value
Length of unigene sequence	867 nt
Precursor coordinates	352 nt -572 nt
Precursor length	221 nt
Mature miRNA length	20 nt
Mature miRNA sequence	UCGCUUGGUGCAGGUCGGGA
miRNA * Sequence	AGUGAACUACGUUCCGCCCU
MFE	-80.30 kcal/mol
AMFE	36.33 kcal/mol
MFEI	0.81
Nucleotide mismatch	1nt
Number of nucleotides in pre-miRNA sequences	A=48, C=33; G=65; T/U=75
(A+T/U)%	55.66%
(G+C)%	44.34%
Family of miRNA	miR 168

MFE- minimal folding free energy; AMFE- adjusted minimal folding free energy; MFEI- minimal folding free energy index

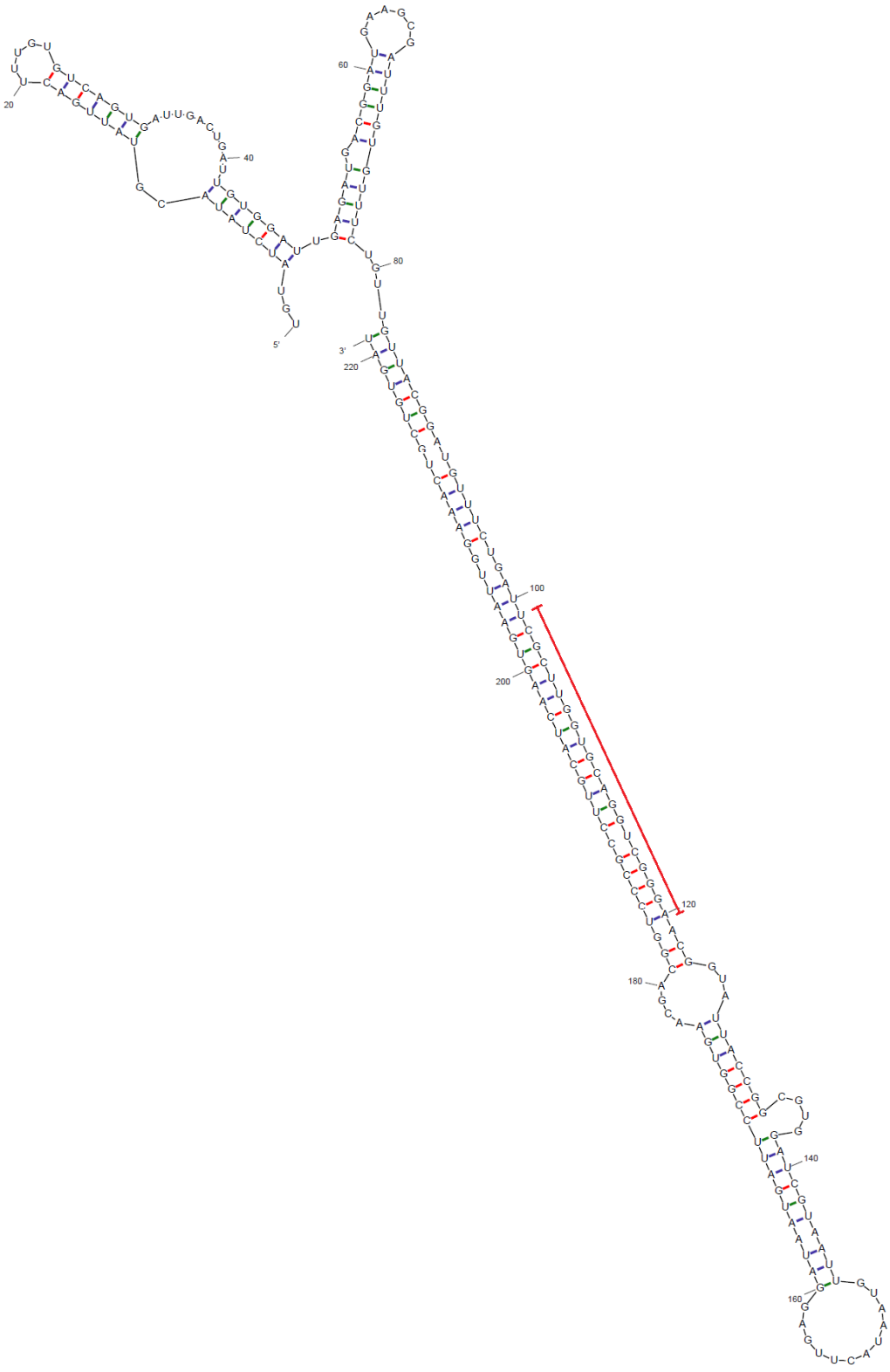


Fig. 3A. The Stem loop hairpin secondary structure of the stevia miRNA. Mature miRNA sequence is highlighted in red colour

Fig. 3B. The predicted fold back secondary structure of stevia miRNA. The actual size of precursors may be slightly longer

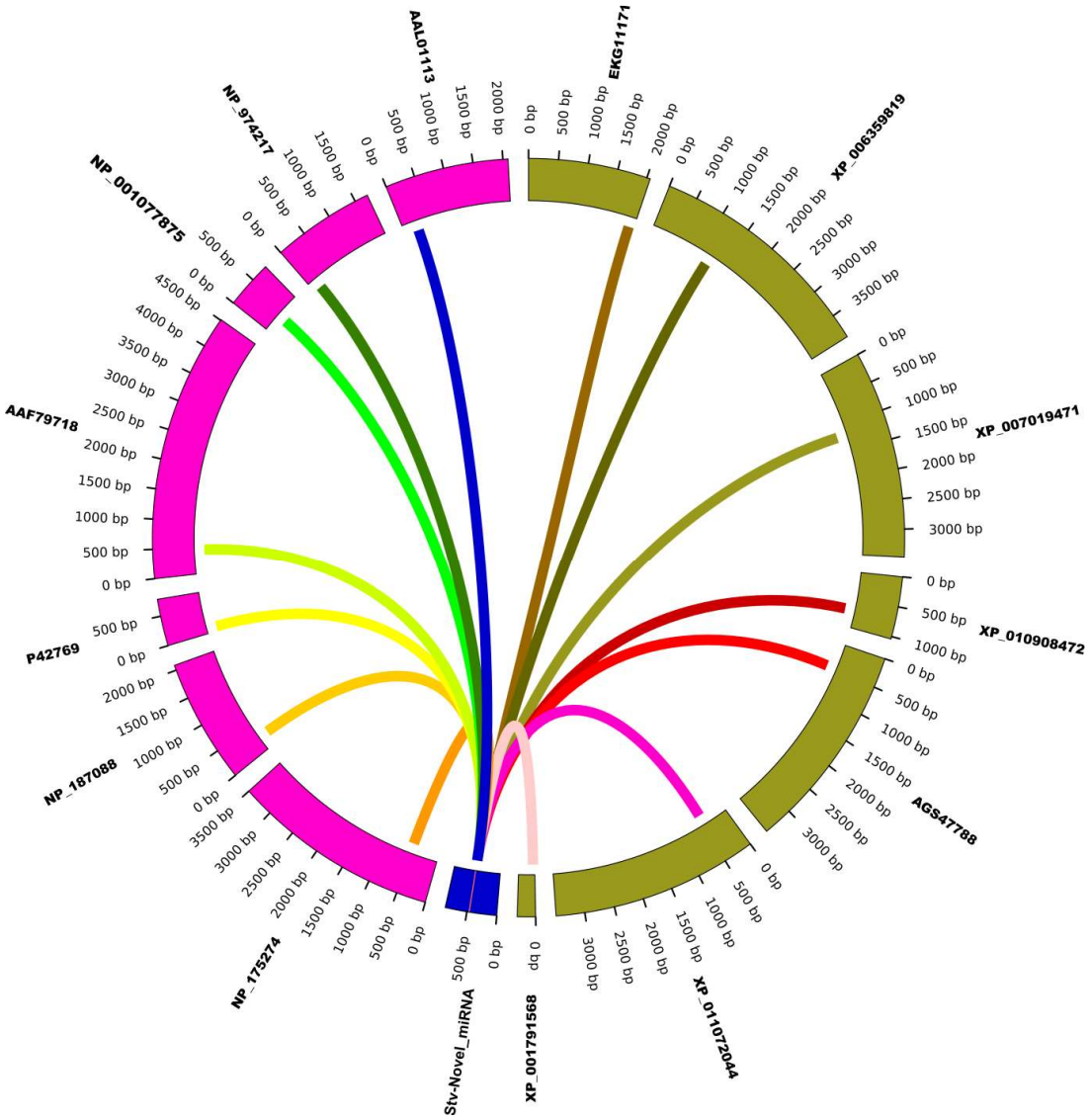


Fig. 4. Circos plot between Stv-Novel_miRNA and their targets

The pink colored blocks in the outer circle represent Arabidopsis thaliana target protein hit accession and light green circle represents stevia unigene target protein hit accession. The targets are labeled according to their Blastx annotation of the target sequence. The scale in the margins of each blocks represents the length of the corresponding sequence in basepair(bp). The blue colored block represents unigene from stevia in which novel microRNA is identified. The position of novel microRNA is represented as dark red line in the blue colored block. The miRNA and its target sequences interconnected by coloured lines

Table 2. List of potential targets for novel predicted miRNA in *Stevia rebaudiana* and their functional annotations

Target ID	E-Value	UPE	Inhibition type	Alignment between miRNA and target	Target annotation	Target accession
unigene_32030	0.5	18.36	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::::::.::::::: Target 2221 UCCCGACCUGUACCAAGCGA 2240	No Hits Found	No Hits Found
unigene_32029	0.5	18.36	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::::::.::::::: Target 1749 UCCCGACCUGUACCAAGCGA 1768	No Hits Found	No Hits Found
unigene_32028	0.5	18.36	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::::::.::::::: Target 507 UCCCGACCUGUACCAAGCGA 526	No Hits Found	No Hits Found
unigene_67745	0.5	13.277	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::::::.::::::: Target 3 UCCCGACCUGUACCAAGCGA 22	No Hits Found	No Hits Found
unigene_56595	2.5	16.605	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :.::::.:::: :::::::: Target 1196 UUCCGAUCUGCUCCAAGCGA 1215	Sugar/inositol transporter	EKG11171
unigene_56594	2.5	16.605	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :.::::.:::: :::::::: Target 1196 UUCCGAUCUGCUCCAAGCGA 1215	Sugar/inositol transporter	EKG11171
unigene_56596	2.5	16.605	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::.::.:::.:::::::: Target 1952 UCCUGAUCUGUACCAAGCGC 1971	Sugar/inositol transporter	EKG11171
unigene_12603	2	13.577	Translation	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::.::.:::.:::::::: Target 1909 UCCUGAUCUGUACCAAGCGC 1928	exocyst complex component 2-like	XP_006359819
unigene_12604	2	13.577	Translation	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::.::.:::.:::::::: Target 1882 UCCUGAUCUGUACCAAGCGC 1901	exocyst complex component sec5 isoform 1	XP_007019471
unigene_38051	2.5	12.882	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::::::. :::::::::: Target 704 ACCCGACUAGCACCAAGCGA 723	2-oxoglutarate-dependent dioxygenase	XP_010908472
unigene_7187	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::.:::::::: : Target 467 UCCCGAGCUGUACCAAGCAA 486	protein argonaute 1	XP_011072044

Target ID	E-Value	UPE	Inhibition type	Alignment between miRNA and target	Target annotation	Target accession
unigene_7186	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::.:::::::: : Target 467 UCCCGAGCUGUACCAAGCAA 486	protein argonaute 1	AGS47788
unigene_7185	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::.:::::::: : Target 467 UCCCGAGCUGUACCAAGCAA 486	protein argonaute 1	XP_011072044
unigene_7183	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::.:::::::: : Target 467 UCCCGAGCUGUACCAAGCAA 486	protein argonaute 1	AGS47788
unigene_7184	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::.:::::::: : Target 467 UCCCGAGCUGUACCAAGCAA 486	protein argonaute 1	XP_011072044
unigene_97585	3	19.933	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::.: :::::::::.:::: Target 23 UCUCCUCCUGCACCAGGCGA 42	hypothetical protein SNOG_00901	XP_001791568
gi|537290375 |gb|GANE01029285.1	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::::::::::.:::::::: Target 2221 UCCCGACCUGUACCAAGCGA 2240	protein argonaute 1	XP_011072044
gi|537290287 |gb|GANE01029326.1	3	19.756	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::::::::::.:::::::: Target 1749 UCCCGACCUGUACCAAGCGA 1768	protein argonaute 1	XP_011072044
AT1G48410.3	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 522 UCCCGAGCUGCAUCAAGCUA 541	protein argonaute 1-like	NP_849784
AT1G48410.1	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 512 UCCCGAGCUGCAUCAAGCUA 531	protein argonaute 1-like	NP_175274
AT1G48410.2	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 512 UCCCGAGCUGCAUCAAGCUA 531	protein argonaute 1-like	NP_849784
AT3G04380.2	3	12.343	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::.:::. :::: ::::::: Target 143 UCUCGAUAUGCAACAAGCGA 162	histone-lysine n-methyltransferase suvr4	NP_974217
AT3G04380.1	3	12.343	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::.:::. :::: ::::::: Target 257 UCUCGAUAUGCAACAAGCGA 276	histone-lysine n-methyltransferase suvr4	NP_187088

Target ID	E-Value	UPE	Inhibition type	Alignment between miRNA and target	Target annotation	Target accession
NP031662	3	17.551	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :.:::::::::.::::: Target 82 ACUCGACCUGCACUAAGCGU 101	glutathione s-transferase	P42769
TC358800	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 512 UCCCGAGCUGCAUCAAGCUA 531	protein argonaute 1-like	NP_175274
TC372434	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 512 UCCCGAGCUGCAUCAAGCUA 531	protein argonaute 1-like	NP_849784
TC383299	3	20.259	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::::: :::::.::::: : Target 459 UCCCGAGCUGCAUCAAGCUA 478	protein argonaute 1-like	AAF79718
EG484074	3	14.195	Translation	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::.:: :: :::.:::::: Target 342 UCCUGAACUCCACUAAGCGA 361	asparagine synthetase domain-containing protein 1-like	NP_001077875
EG484076	3	13.88	Translation	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 :::.:: :: :::.:::::: Target 342 UCCUGAACUCCACUAAGCGA 361	asparagine synthetase domain-containing protein 1-like	NP_001077875
NP1659635	3	12.343	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::.:::. :::: ::::::: Target 143 UCUCGAUAUGCAACAAGCGA 162	histone-lysine n-methyltransferase suvr4	NP_974217
TC362362	3	12.343	Cleavage	miRNA 20 AGGGCUGGACGUGGUUCGCU 1 ::.:::. :::: ::::::: Target 257 UCUCGAUAUGCAACAAGCGA 276	histone-lysine n-methyltransferase suvr4	AAL01113

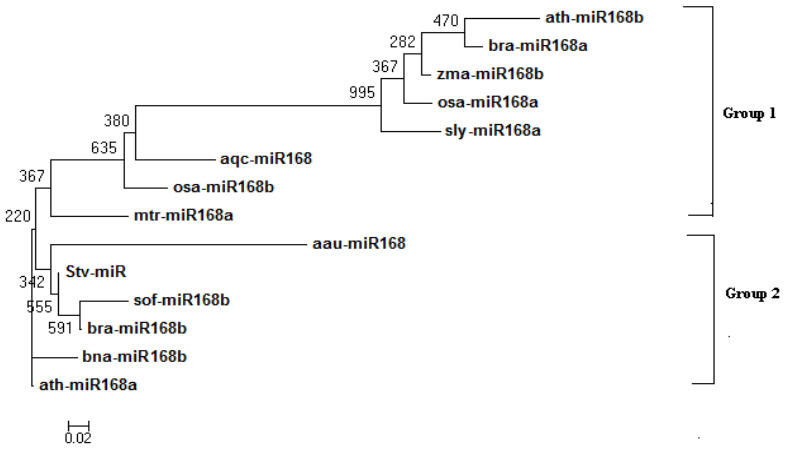
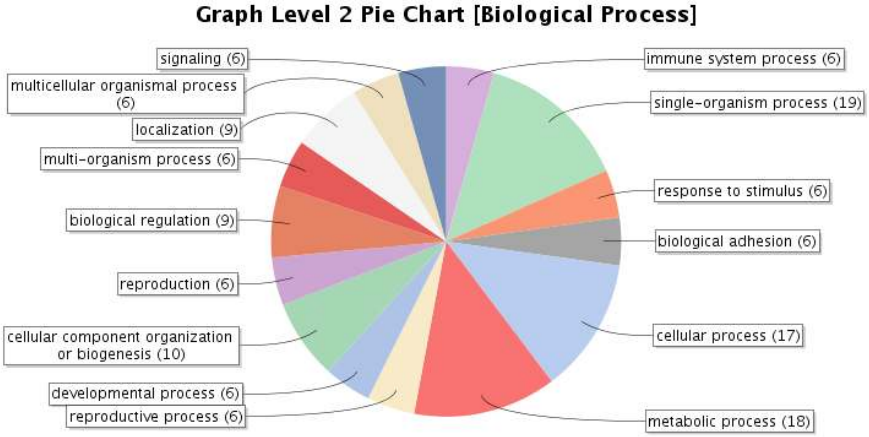


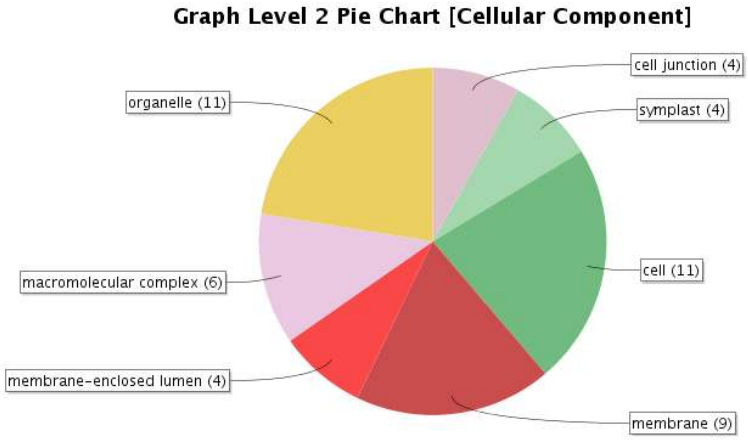
Fig. 5. Phylogenetic tree of newly identified miRNA in Stevia with published miRNAs of miR168 family of other plant species

The tree was constructed using the Neighbor-Joining method in MEGA6 software and evaluated with 1,000 bootstrap replications (bootstrap values are labeled along the branches). The branch length is proportional to the substitution per site

(A)



(B)



(C)

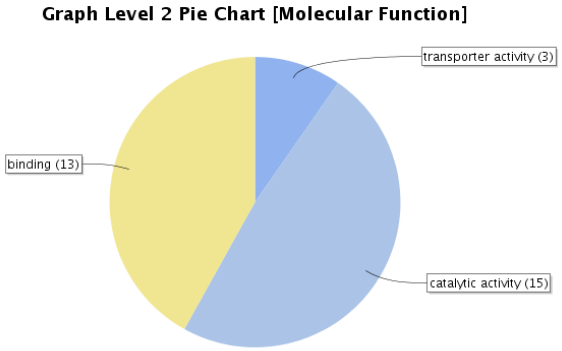


Fig. 6. GO term distribution of (A) Biological process (B) Cellular component (C) Molecular function

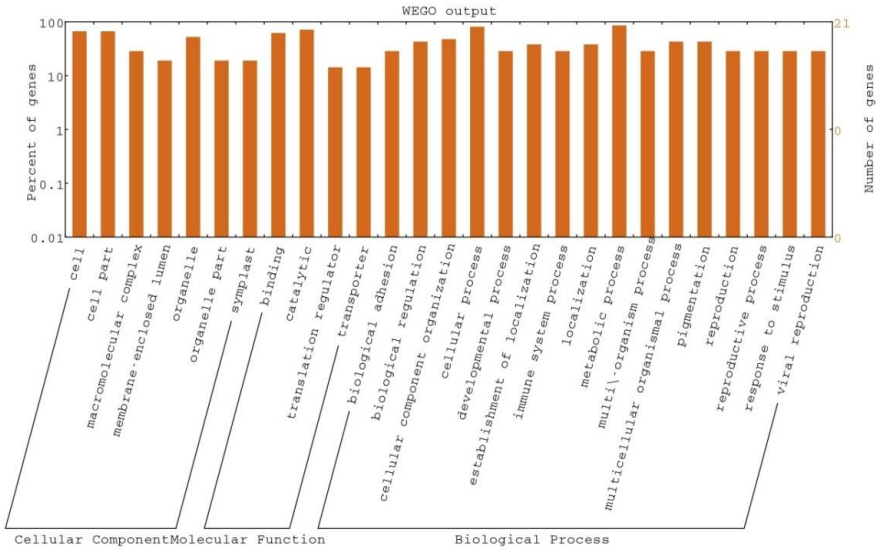


Fig. 7. WEGO plot of Gene Ontology classification

The GO term is categorized into three main GO categories: Cellular Component, Molecular Function and Biological process

4. DISCUSSION

The plant miRNAs are highly conserved among different species including dicots and monocots and they are involved in important metabolic activities like a diverse process of development and growth. In the present study, only one potential miRNA in *Stevia* from transcriptomic data was identified by homology-based approach. The newly identified miRNA is member of miR168 family. The important characteristic of pre-miRNA and the crucial step in miRNA maturation is the formation of stem-loop hairpin secondary structure. To evade the designating miRNAs from other non-coding RNAs, three important criteria were considered such as minimal fold energy (MFE), adjusted minimal fold energy (AMFE), and the minimal fold energy index (MFEI). The MFE determines the secondary structure of nucleic acids. Its lower value indicates the higher stability of the secondary structure of the corresponding sequence. The predicted structure of the hairpin is shown in Fig. 3 [22]. The hairpin structure depicts there are nucleotides engaged in G/U pairings between miRNA and miRNA* in stem region and does not contain large internal loops or bulges. So the newly identified *Stevia* miRNA satisfied the criteria for the annotation of novel miRNA. The phylogenetic tree among different plant species of miR168 family shows the evolutionary relationship of newly identified Stv-

miRNA. The phylogenetic tree illustrated that the Stv-miR branched out from ath-miR168a and aau-miR168. The bra-miR168b and sof-miR168b branched out from Stv-miR forming a subgroup with good bootstrap value from group 2.

The computational approach was used to predict the targets to understand the role played by predicted miRNA in plant development (Table 2). It was seen that majority of the predicted target genes coding for argonaute protein. The Argonaute (AGO) proteins interact with miRNA or short interfering RNA (siRNA) and are facilitating the major downstream events of cytoplasmic post-translational gene silencing processes and they are involved in alternative splicing and DNA repair also [23]. AGO protein plays a role in RNA silencing also which function as primary antiviral immune systems in plants [24]. Another miRNA target was asparagine synthetase domain-containing 1-like protein, which plays role in plant defense responses and primary nitrogen metabolism [25]. Other targets were sugar/Inositol transporter, histone-lysine n-methyltransferase suvr4, exocyst complex component sec5 which are involved in plant cell growth and morphogenesis [26], 2-oxoglutarate-dependent dioxygenase look after epigenetic regulation [27] and glutathione S-transferase have non-catalytic roles as carriers for phytochemicals [28]. These findings provide us a potential to identify miRNA and their probable roles in plant development. This research will add value in further in-depth research with reference to the regulatory mechanism of *Stevia* leaf miRNAs.

5. CONCLUSION

In this research work, a new approach to identify potential new miRNA from transcriptome data of Stevia leaf which is a highly economical plant species with the medicinal property. One potential novel miRNA from assembled unigenes of Stevia leaf was identified. Most of the miRNA targets were involved in plant growth, metabolism and also can provide insight into the potential role in post-translational gene silencing process. This study employs the bioinformatics approach for the identification and analysis of miRNAs of those plant species whose genome information is not yet fully available or known in public databases.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

ACKNOWLEDGEMENT

We acknowledge management of Xcelris Labs Ltd., for support and encouragement to carry out this study.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Lu Y, Yang X. Computational identification of novel MicroRNAs and their targets in *Vigna unguiculata*. Comp Funct Genomics; 2010.
2. Xie FL, Huang SQ, Guo K, Xiang AL, Zhu YY, Nie L, Yang ZM. Computational identification of novel microRNAs and targets in *Brassica napus*. FEBS Lett. 2007;581(7):1464-74.
3. Sarika Sahu, Dev Bukhsh Singh, Kusum Kumari Yadav, Rai DV, Rekha Dixit. Computational identification and functional annotation of miRNAs in medicinal plant *Helianthus petiolaris.* Network Modeling Analysis in Health Informatics and Bioinformatics. 2(4):277-284. DOI:10.1007/s13721-013-0044-8
4. Li X, Hou Y, Zhang L, Zhang W, Quan C, Cui Y, Bian S. Computational identification of conserved microRNAs and their targets from expression sequence tags of blueberry (*Vaccinium corybosum*). Plant Signal Behav; 2014.
5. Vishwakarma NP, Jadeja VJ. Identification of miRNA encoded by *Jatropha curcas* from EST and GSS. Plant Signal Behav; 2013
6. Mandhan V, Kaur J, Singh K. smRNAome profiling to identify conserved and novel microRNAs in *Stevia rebaudiana* Bertoni. BMC Plant Biol; 2012.
7. Pandey M, Chikara SK. *In vitro* regeneration and effect of abiotic stress on physiology and biochemical content of *Stevia rebaudiana* 02 'Bertoni'. J Plant Sci Res; 2014.
8. Chen J, Hou K, Qin P, Liu H, Yi B, Yang W, Wu W. RNA-Seq for gene identification and transcript profiling of three *Stevia*

rebaudiana genotypes. BMC Genomics; 2014.

9. Arpan R. Modi, Sandip Raj. Analysis of differentially expressed genes involved in stevioside biosynthesis in cultures of *Stevia rebaudiana* Bertoni Treated with Steviol as an Immediate Precursor Springer; 2013.
10. Akter A, Islam MM, Mondal SI, Mahmud Z, Jewel NA, Ferdous S, Amin MR, Rahman MM. Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea arabica*). Saudi J Biol Sci; 2014.
11. Chao YT, Su CL, Jean WH, Chen WC, Chang YC, Shih MC. Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. Plant Mol Biol; 2014.
12. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics; 2014.
13. Hang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X. Bridger: A new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol; 2015.
14. Weizhong Li, Adam Godzik Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics; 2006.
15. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics; 2010.
16. Wang CM, Liu P, Sun F, Li L, Liu P, Ye J, Yue GH. Isolation and identification of miRNAs in *Jatropha curcas*. Int J Biol Sci; 2012.
17. Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics; 2005.
18. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Res; 2009.
19. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0; 2013.
20. Barik S, SarkarDas S, Singh A, Gautam V, Kumar P, Majee M, Sarkar AK. Phylogenetic analysis reveals conservation and diversification of micro RNA166genes among diverse plant species. Genomics; 2014.
21. Rastogi S, Meena S, Bhattacharya A, Ghosh S, Shukla RK, Sangwan NS, Lal RK, Gupta MM, Lavania UC, Gupta V, Nagegowda DA, Shasany AK. De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. BMC Genomics; 2014.
22. Hemant Gupta, Tanushree Tiwari, Maulik Patel, Aditya Mehta, Arpita Ghosh, An approach to identify the novel miRNA encoded from H. Annuus EST sequences, Genomics Data; 2015.
23. Meister G. Argonaute proteins: Functional insights and emerging roles. Nat Rev Genet; 2013.
24. Carbonell A, Carrington JC. Antiviral roles of plant ARGONAUTES. Curr Opin Plant Biol; 2015.
25. Hwang IS, An SH, Hwang BK. Pepper asparagine synthetase 1 (CaAS1) is required for plant nitrogen assimilation and defense responses to microbial pathogens. Plant J; 2011.
26. Hála M, Cole R, Synek L, Drdová E, Pecenková T, Nordheim A, Lamkemeyer T, Madlung J, Hochholdinger F, Fowler JE, Zárský V. An exocyst complex functions in plant cell growth in Arabidopsis and tobacco. Plant Cell; 2008.
27. Farrow SC, Facchini PJ. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. Front Plant Sci; 2014.
28. Edwards R, Dixon DP, Walbot V. Plant glutathione S-transferases: Enzymes with multiple functions in sickness and in health. Trends Plant Sci; 2000.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/14803