Research & Reviews: Journal of Botanical Sciences

Computational Prediction, Target Identification and Experimental Validation of miRNAs from Expressed Sequence Tags in *Cannabis sativa*. L

Ganesh Selvaraj Duraisamy[1], Ajay Kumar Mishra[1], Jernej Jakse[2] and Jaroslav Matousek[1*]

[1]Biology Centre ASCR v.v.i, Institute of Plant Molecular Biology, Branisovska 31, ceske Budejovice 370 05, Czech Republic

[2]University of Ljubljana, Biotechnical Faculty, Agronomy Department, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia

Research Article

*For Correspondence

Jaroslav Matousek , Biology Centre ASCR v.v.i, Institute of Plant Molecular Biology, Branisovska 31, ceske Budejovice 370 05, Czech Republic ,Tel: (+420) 387 775 525

E-mail: jmat@umbr.cas.cz

ABSTRACT

MicroRNAs (miRNAs) are approximately 20-22 nucleotide non-coding RNAs, which play an important role in posttranscriptional degradation of target mRNA or inhibition of protein synthesis through binding the specific sites of target mRNA. miRNAs have been extensively studied in various plant species, however, there is no miRNA identified in *Cannabis sativa*, a crop which has long been used for hemp fibre, for seed and seed oils, for medicinal purposes, and as a recreational drug. In this study, a computational based approach was used to identify and characterize *C. sativa* miRNAs. A total of 7 miRNAs belonging to 3 miRNA families were identified in cannabis based on homolog search and series of filtering criteria. The identified miRNAs were validated by end-point PCR and quantitative reverse transcription-polymerase chain reaction (qRT-PCR), confirmed the existence of conserved miRNAs in *C. sativa*. Based on near-perfect complementarity between cannabis miRNAs and their target mRNA gene sequence, a total of 23 miRNAs targets were identified involved in processes such as plant development, signal transduction, secondary metabolite production, protein degradation, response to environmental stress and pathogen invasion, and ability to regulate their own biogenesis. The *cis*-regulatory elements relevant to biotic and abiotic stress, plant hormone response flavonoid and cannabinoids biosynthesis were identified in the promoter regions of those miRNA genes. Overall, findings from this study will accelerate the way for further researches of miRNAs and their functions in *C. sativa*, particularly cannabinoid metabolic pathway.

INTRODUCTION

Cannabis (*Cannabis sativa L., Cannabaceae*) is an annual herbaceous plant in the *Cannabis* genus, indigenous to Central and South Asia. Cannabis produces many secondary compounds such as flavonoids, stilbenoids, alkaloids, lignanamides, phenolic amides, but main active psychotropic constituent of cannabis is Δ^9-tetrahydrocannabinol (Δ^9-THC). Cannabinoids, terpenoids, and other compounds are secreted by glandular trichomes that occur most abundantly on the floral calyxes and bracts of female plant. Despite of its medicinal properties, *C. sativa* is among the very oldest of economic plants providing humans with fiber for spinning, weaving cloth, and making paper; seed for human foods and animal feeds; and aromatic resin containing compounds of recreational and medicinal value and also a useful source of foodstuffs and bio-fuels [1,2].

The recent progress in functional genomics research based on large-scale expressed sequence tag (EST) generation, analysis and cloning of genes in cannabis plant has provided a critical significance on elucidating the molecular mechanism of growth, development, differentiation, metabolism, quality, yield, and stress resistance, as well as genetic manipulation via biotechnological approaches in the foreseeable future. However, in cannabis, only a few regulatory players mostly related to regulating gland development and THCA biosynthesis have been studied so far [3,4]. Up to date, limited progress has been made

in molecular and genetic studies, including small RNA mediated gene regulation. MicroRNAs (miRNAs) are a small non-coding RNA molecule containing about 20-22 nucleotides found in plants, animals, and some viruses, which functions in RNA silencing and post-transcriptional regulation of gene expression. Increasing evidence suggests that miRNAs play pivotal roles in multiple biological processes in plants, especially controlling tissue (leaf, root, stem, and flower) differentiation and development self-negative feedback regulation of metabolism, phase transition from vegetative growth to reproductive growth, signal transduction and response to environmental conditions such as biotic and abiotic stresses including nutrient stresses, for example, phosphorus starvation, nitrogen starvation, and micronutrient deficiency or toxicity. Considering the importance of miRNAs, several approaches have been established for identifying miRNAs in various plant species, such as, gene expression analyses, high-throughput method and computational prediction [5-11]. The limitation of using differential gene expression to identify miRNA targets is that they are observed amongst a pool of indirect changes in transcript abundance. This may assist in describing the predominate genes and pathways affected by a miRNA but does not distinguish between direct targets [12].

A comparative genomics study across hugely contradictory taxa has shown that many miRNAs are highly evolutionarily conserved from species to species in the plant and animal kingdom. This feature of extensive evolutionary conservation of these miRNAs among themselves renders a powerful approach to their identification using computational based approach. On the basis of this strategy, researchers developed an expressed sequence tag (EST) and a genome survey sequence (GSS) approach to identify miRNAs from plants or even from various animals [13]. EST analysis has some substantial advantages over the other approaches such as: (1) conserved miRNAs can be identified whose complete genome sequences are unavailable, (2) provides direct evidence for miRNA expression that cannot be inferred from genomic sequence surveys, and (3) miRNA identification can be conducted without highly specialized software. EST-based approaches have been successfully used to identify miRNAs genes in various plants such as maize, wheat, soybean, tobacco, potato and Asiatic cotton [14-19].

Recently, molecular biology of cannabis plants has been one of the most active and kinetic research fields of pharmaceutical industries. Furthermore, miRNAs are being used by industrial companies both as targets and therapeutic agents in order to generate new treatment methods for diseases [20]. However, there has been no report on experimental or computational identification of miRNA in *C. sativa* species. Understanding gene function and their regulatory mechanisms could provide new insight into the mechanisms of cannabis growth and development, the response to environmental abiotic and biotic stresses and particularly on glandular trichomes initiation and development.

Therefore, in this study, we applied computational approach of comparative genome based homologue search to identify miRNAs in cannabis using their current expressed sequence tags (ESTs) available in the NCBI Genebank database. The newly identified miRNAs will enable us to study miRNAs-mediated gene network, transcriptional factors (TFs) involved in regulating secondary metabolite pathway as well as understanding the growth and development and other physiological process involving in glandular trichomes and leaf development in relation to abiotic and biotic stress.

MATERIAL AND METHODS

Sequences and Software

The known plant miRNA sequences from *Arabidopsis, Brassica, Glycine, Saccharum, Sorghum, Vitis, Solanum, Oryza, Triticum, Chlamydomonas*, and other plant species were downloaded from the miRNA database miRBase (http://www.mirbase.org/) (miRBase version 21.0, accessed June 2014). After removing redundant sequences, the unique sequences were used as reference miRNA set for later Blastn seaches for identifying *C. sativa* miRNAs. Cannabis expressed sequence tags (ESTs), cDNAs, and mRNAs were downloaded from the GenBank nucleotide databases at the National Center for Biotechnology Information (NCBI). Currently, a total of 12,907 cannabis ESTs are available in the NCBI EST database (dbEST release 130101, January 1, 2013; http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary .html). Comparative software mpiBLAST-1.6.0 parallel implementation of NCBI BLAST was downloaded and set up locally. To predict the secondary structure of pre-miRNA, Zuker folding algorithm software MFOLD 3.2 (http://mfold.rna.albany.edu/ ?q=mfold/RNA-Folding-Form) was used online to analyze the secondary structure and calculating the minimum free energy [21-23]. The Plant Small RNA Analysis Server (psRNATarget) formerly known as miRU was used for miRNA target analysis [24].

Identification of Potential miRNAs in *C. sativa*

The mature sequences of all currently available plant miRNAs after removing redundant sequence were subjected to a Blastn search against all of the currently available cannabis EST sequences using mpiBLAST-1.6.0 algorithm in Linux based 32 core cluster system. Adjusted blast parameter settings were as follows: the default word-match size between query and database sequences was set at seven; the expected values were set at 1,000 to increase the hit chance for more potential sequences and the sequence number of the Blastn search and the sequence alignments were set to 1,000. If only a partial reference mature miRNA sequence was aligned to a Expressed sequence tag (EST) sequence, the non-aligned parts were manually inspected and compared to determine the number of matching nucleotides. These sequences were considered as potential miRNA candidates only if they fulfill the following criteria with slight modification from a previous study (1) at least 18-22 nt length was assumed between the predicted mature miRNAs, and (2) allowed to have 0-3 nt mismatches in sequence with all previously known plant

mature miRNAs. These whole EST sequences (n=12,907) were used for BLASTx analysis using NCBI BLAST+2.2.29 against NCBI nr (non-redundant protein) database on January, 2014 for removing the protein-coding sequences and retaining only the non-protein coding sequences [22,25]. About 25 EST sequences were found that were non-protein-coding sequences.

The secondary structure of candidate pre-miRNA sequences of these potential miRNA homologs was predicted using the Zuker folding algorithm with MFOLD-3.2 [23]. All parameters were set to default values. All outputs obtained from mfold were recorded into spreadsheet, which included the EST 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, DG in kcal/mol). Then, the adjusted minimal folding free energy (AMFE) [AMFE= (MFE/length of a potential pre-miRNA) *100)] and the minimal folding free energy index (MFEI) [MFEI= AMFE/(G+C)%, where (G+C)% represents %GC content over pre-miRNA sequence] were calculated according to a previous report [14]. An EST was considered a miRNA candidate when it fit all of the following criteria: 1) the predicted mature miRNAs had no more than three nucleotide substitutions compared with a known mature miRNAs; 2) the EST sequence could fold into an appropriate stem-loop hairpin secondary structure; 3) the mature miRNA was localized in one arm of the stem-loop structure; 4) there was no loop or break in the miRNA or miRNA* sequences; 5) there were no more than 6 mismatches between the predicted mature miRNA sequence and its opposite miRNA* sequence in the secondary structure; and 6) the predicted secondary structure had high negative MFE and high MFEI value. Candidate miRNAs those met the criteria were further scanned to remove repeated sequences and the remaining list was used for subsequent analysis [26].

Prediction of Potential miRNA Targets

Prediction of candidate miRNA targets was done using psmiRNATarget (psRNAtarget, http://plantgrn. noble.org/psRNATarget/) with following parameters: (1) maximum expectation value 3; (2) multiplicity of target sites 2; (3) range of central mismatch for translational inhibition 9-11 nucleotide; (4) maximum mismatches at the complementary site ≤4 without any gaps. psRNATarget provides reverse complementary matching between miRNAs and its target transcript and finds the target site accessibility by calculating unpaired energy (UPE) necessary for opening the secondary structure around the miRNA target site [24].

Potential Core Promoter Identification and Analysis of cis-Acting elements of *C. sativa* miRNAs

For promoter prediction, draft genome sequence of *C. sativa* was downloaded from the National Center for Biotechnology Information (NCBI accession number AGQN00000000). The upstream 2000 bp or longer sequences of pre-microRNAs (hairpin precursors) of each miRNA gene was used for promoter prediction. Promoters were predicted by the plant promoter identification program TSSP (http://www.softberry.com), which is designed for predicting plant Pol II promoters [27]. The predictions were obtained at the default TSSP settings. Potential promoter regions from transcription start site (TSS) to 800 bp upstream were obtained to predict the potential *cis*-acting elements and motifs. If there were overlapping TSSs in the selected region, only the sequence between the two TSSs was analyzed to exclude redundancy [28]. The PlantCARE database (http://bioinformatics. psb.ugent.be/webtools/plantcare/ html), a database of plant *cis*-acting regulatory elements was used to analyze the *cis*-acting elements of the miRNAs [29].

Experimental validation of *C. sativa* miRNAs

The predicted cannabis miRNAs were validated using stem-loop RT-PCR and End-point PCR according to the previous report [30]. A stem-loop-containing RT primer with its 5′ end complementary to target the miRNA's last six nucleotides at the 3′ end was designed. Seeds of *C. sativa* were collected from village Chbany (Northern Bohemia, Czech Republic), germinated and 9-day-old seedlings were planted into 11 LC pots with soil (substrate 45 L, Czech Republic). Plants were grown and maintained under natural light conditions with supplementary illumination [170 μmol m-2 s-1 PAR] to maintain a 16 h day period collected and maintained in the green house at a temperature of 25 ± 3˚C. Small RNA was extracted from 200 mg leaf and flower sample using PureLink™ miRNA Isolation Kit (Invitrogen) After measuring the concentration and quality of the miRNAs, 200 ng of small RNA was used to reverse transcribe a miRNA into cDNA using a miRNA-specific stem-loop primer and the SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. Reverse transcription reactions (20μl) contained 2 μl of miRNA, 50 nM stem-loop RT primer, 0.25 mM each of dNTPs, 50 units reverse transcriptase, 1x reverse transcriptase buffer, 10 mM DTT, and 4 units RNase inhibitor. The reactions were incubated for 30 minutes at 16°C, followed by pulsed RT of 60 cycles at 30˚C for 30 seconds, 42˚C for 30 seconds, and 50˚C for 1 second. Pulsed RT reactions provide better detection sensitivity compared with non-pulsed reactions [31]. Reactions were terminated by incubating at 85°C for 5 minutes to inactivate the reverse transcriptase. Ubiquitin U6, one of the uniformly expressed small RNAs, was used as the internal control for stem-loop RT-PCR. All the oligos used in this study were listed in supplemental table **Table S1**.

The End-point PCR reaction mixture (V = 20 μl) contained 1 μl of RT product, 0.25μM each miRNA-specific forward primer and universal reverse primer, 0.6 units of Hot Start Ex Taq polymerase (TaKaRa Bio), 1x Taq buffer, 200 μM dNTPs mixture. The amplifications were carried out in thermal cycler (Bio-Rad) with an initial denaturing step at 94˚C for 2 min, followed by 40 cycles of 95˚C for 15 s and 60˚C for 1 min. The product size was confirmed by melting analysis and 4% agarose gel electrophoresis.

Quantitative real time PCR (qRT-PCR) was performed on an iQ5TM Real-Time PCR Systems (Bio-Rad). For each reaction, 20 µL PCR reaction mixtures were prepared and each contained 2 µL of RT product from the reverse-transcription reaction, 0.5µM miRNA-specific forward primer and stem-loop reverse primer, 1x Universal Syber Green PCR Master Mix (Invitrogen). Amplification curves were generated with an initial denaturing step at 95˚C for 10 min, followed by 45 cycles of 95˚C for 15 s and 60˚C for 60 s. Melting curves were generated using the following program: PCR products were denatured at 95°C and cooled to 65˚C at 20˚C per second to determine the specificity of each reaction. The fluorescence signals at a wave length of 530 nm were then collected continuously from 65˚C to 95˚C as the temperature was increased at 0.2˚C per second. All reactions were performed in triplicate and a control without template was included for each miRNA. The threshold cycle (C_T) values were determined automatically by the instrument and the fold changes of each miRNA gene were calculated as relative quantity (RQ) values using the comparative C_T ($2^{-\Delta\Delta Ct}$) method.

RESULT

Identification of potential miRNAs in *C. sativa*

A reference set of 10,898 plant miRNAs were blasted against 12,907 ESTs from NCBI EST database (dbEST) of cannabis. After circumspectly considering the Blastn and Blastx and after removing protein coding sequences, 25 unique EST sequences were identified that showed homology with known reference plant mature miRNA. The selected EST sequences were subjected to secondary structure prediction by using MFOLD 3.1 program. The results obtained by MFOLD were inspected for determining the sequence of a miRNA precursor and appropriate stem-loop structure using the criteria described in material and methods. After analyzing all outputs we identified 7 potential miRNA from cannabis EST sequences belonging to 3 different miRNA families, which share a high degree of sequence identity with known mature miRNA according to criteria mentioned in materials and methods **(Table 1 and Fig.ure1)**. The predicted miRNAs were named in accordance with miRBase and mature sequences are designated 'miR' and the precursor hairpins are labeled as 'MIR'with the prefix 'csa' for *Cannabis sativa* [32].

Characteristics of miRNAs identified in *C. sativa*

The miRNA candidates identified in *C. sativa* have are different nucleotide identity ratios and majority of identified mature cannabis miRNAs had 2 or 3 nucleotide changes as compared with miRNAs in other plant species. All identified miRNAs were obtained from the plus strand. However, there were several miRNAs identified from the minus strand of certain ESTs. The mature miRNA sequences could be located within either the 3′ or 5′ arm of the secondary stem-loop hairpin structures. Among the 7 identified cannabis miRNAs, 3 (42.85%) were found to be located on the 5′ arm of the hairpin secondary structure. The remaining 4 (57.14%) located on the 3′ arm **(Figure 1)**. The length of mature miRNAs varied from 20 to 21 nt. The majority of identified potential cannabis miRNAs were 20nt in length (57.14%) followed 21 nt (42.85%) length respectively **Table 1**, which is similar to those of other plant species and length of identified miRNA precursor sequences was also varied among identified miRNAs. The predicted precursor sequences ranged in length from 200 to 409 nt with an average of 277 ± 76.60 nt **Table 2**. The composition of the four nucleotides (A, G, C, and U) is an important parameter, which is an indicator for species evolution as well as for the stabilization of RNA secondary structure. The percentage composition of each nucleotide was not evenly distributed in the identified cannabis pre-miRNAs **Table 1**. With an unknown reason, the nucleotide uracil (U) is found to be dominant in both mature miRNAs and pre-miRNAs of plant and animal. In accordance with this study, we observed that the U content varied from 21% to 42.54% with an average of 35.74 ± 6.39 % in the identified *C. sativa* pre-miRNAs **Table 2** which is significantly higher than the content of other nucleotides, particularly much higher than nucleotides C (12.85 ± 4.64%) and G (16.09 ± 2.60%). A majority (71.42%) of pre-miRNAs contained more than 30% of the nucleotide U. Nucleotides G and C contributes to the formation and stabilization of the secondary structure of stem-loop hairpins. In the identified cannabis pre-miRNAs, the GC content (28.95 ±6.48%) was much lower than the AU content (71.04 ± 6.48%) **Table 2**. The average A/U and C/G ratio were 1.08 and 0.80 respectively, which suggested that G and U nucleotide were distributed predominantly in the pre-miRNA sequences.

Minimal folding free energy (MFE) is a prominent characteristic for determining the secondary structure of nucleic acids such as DNA and RNA. The lower value of MFE signifies thermodynamically more stable secondary structure of the corresponding DNA or RNA sequence, (Bonnet et al., 2004). The MFE of the 7 identified cannabis pre-miRNAs varied from -42.20 to -105.60 kcal/mol with an average of -67.03 ± 24.15 kcal/mol **Table 2** [33]. The large variance in MFE value is attributed to significance variation of nucleotide in their length. It is also an issue for determining the stability of RNA or DNA using MFE because different RNA or DNA strands contains a different number of nucleotides. To better measure the stability of RNA or DNA strands, the adjusted minimal folding free energy (AMFE) strategy was developed, which is the MFE of a RNA/DNA sequence that is 100 nt in length. The AMFE of the 7 identified cannabis pre-miRNAs ranged from -21.10 to -35.20 kcal/mol with an average of - 23.90 ± 4.80 kcal/mol, which is a smaller range as compared to the MFE range **Table 1.** The minimal folding free energy index (MFEI) is a new criterion for assaying and distinguishing miRNAs from other coding and non-coding RNAs. The MFEI of the identified cannabis pre-miRNAs ranged from 0.80 to 0.92 with an average of 0.83 ± 0.03 **Table 2**.
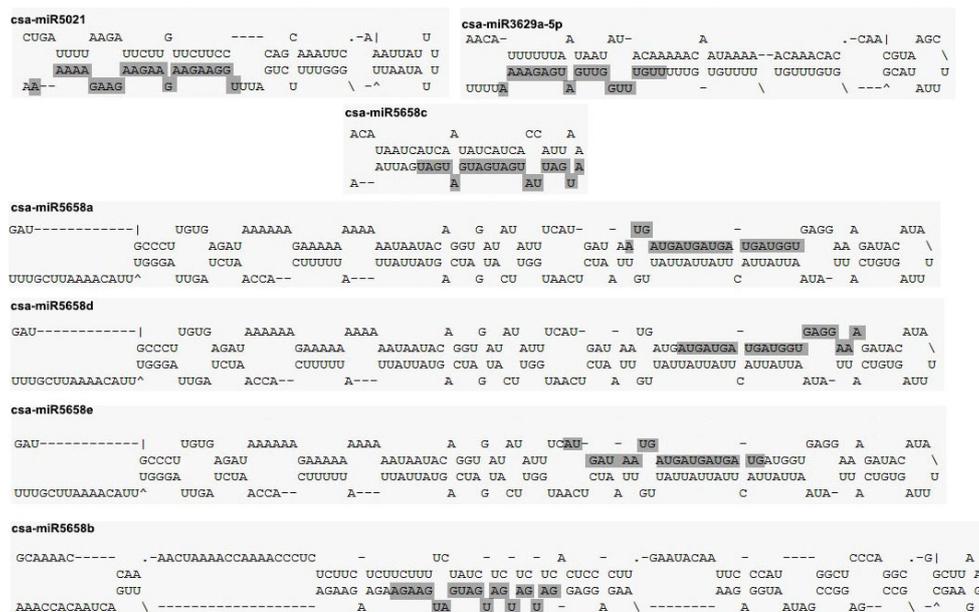
Figure 1. Predicted hairpin secondary structures of the C. sativa miRNAs identified in this study. Mature miRNA sequences are in shadow. The length of the accurate miRNA precursors may be slightly longer than what is presented here.

Table 1. Characteristics of Cannabis sativa miRNAs identified by homolog search and secondary structure

miRNAs	Source miRNAs homologous	Family	EST source	Mature sequence of miRNA (5' to 3')	ML (nt)	PL (nt)	A	U	G	C	A+U (%)	G+C (%)	MFE	AMFE	MFEI
csa-miR5021	ath-miR5021	5021	JK499439	UGGAGAAGAAGAAGAAGAAAA	21	280	83	110	43	44	68.93	31.07	69.82	24.94	0.80
csa-miR3629a-5p	vvi-miR3629a-5p	3629	JK494915	UUGUUUGGUUGAUGAGAAAA	20	409	118	174	63	54	71.39	28.61	93.31	22.81	0.80
csa-miR5658a	ath-miR5658	5658	GR220701	AUGAUGAUGAUGAUGAUGGU	20	200	74	75	33	18	74.50	25.50	42.2	21.10	0.83
csa-miR5658b	ath-miR5658	5658	JK495066	GAUGAUGAUGAUGAUGAAGA	20	300	106	63	63	68	56.33	43.67	105.6	35.20	0.81
csa-miR5658c	ath-miR5658	5658	JK500755	AUGAUUAUGAUGAUGAUGAU	20	350	148	122	40	40	77.14	22.86	73.9	21.11	0.92
csa-miR5658d	ath-miR5658	5658	GR220701	AUGAUGAUGAUGGUGAGGAAA	21	200	74	75	33	18	74.50	25.50	42.2	21.10	0.83
csa-miR5658e	ath-miR5658	5658	GR220701	AUGAUAAUGAUGAUGAUGAUG	21	200	74	75	33	18	74.50	25.50	42.2	21.10	0.83

ML: Mature sequence length; PL: Length of pre-miRNAs; MFE: Minimal folding free energy (-kcal/mol); AMFE: Adjust minimal folding free energy (kcal/mol); MFEI: Minimal folding free energy index.

Table 2. Statistics of the characterized parameters of C. sativa miRNA precursors

Parameter	Minimal	Average	Maximal	Median	Standard Deviation
MFE (-kcal/mol)	42.20	67.03	105.60	69.82	24.15
AMFE (-kcal/mol)	21.10	23.90	35.20	21.11	4.8
MFEI	0.80	0.83	0.92	0.83	0.03
Length (nt)	200	277	409	280	76.60
(G+C)%	22.86	28.95	43.67	25.50	6.48
(A+U)%	56.33	71.04	77.14	74.50	6.48
A%	28.85	35.30	42.29	37.00	4.32
U%	21.00	35.74	42.54	37.50	6.39
C%	9.00	12.85	22.67	11.43	4.64
G%	11.43	16.09	21.00	16.50	2.60
A/U	0.68	1.04	1.68	0.99	0.30
C/G	0.55	0.80	1.08	0.86	0.22

Prediction of potential miRNA targets of *Cannabis sativa*

The knowledge on target function of the identified cannabis miRNA will assist us to gain insight into the important function and regulation of miRNAs in this plant. A total of 23 potential target genes were identified and these targets encode proteins with various biological functions, ranging from transcription factors, stress response, signal transduction, Ubiquitination system, Intracellular trafficking, metabolic enzymes, RNA and protein processing proteins, other proteins for plant growth and development

Table 3. Results from previous study revealed that one miRNA could target one gene or several genes or multiple miRNAs could target only one gene [34]. In our study we identified cannabis miRNAs (csa-miR5658a, csa-miR5658b, csa-miR5658d and csa-miR5658e) that could target several TFs such as Myb and GRAS which play an important role in a variety of biological functions including plant development, hormone signaling and metabolism. csa-miR5021 targets MAPK protein kinase which regulate various biotic stress associated with hormonal responses through transcriptional, translation regulation and protein-protein interaction in many plant species [35]. The other putative target of csa-miR5021 is SLP3 subtilisin-like serine protease 3 which play a role in regulating morphogenesis and development as well as in controlling pathogenesis-related responses and signalling in plants [36]. The common target of csa-miR5658a, csa-miR5658b, csa-miR5658d and csa-miR5658e is BAK1-interacting receptor-like kinase 1 having putative role in plant growth and development such as in lateral roots via auxin regulation [37]. csa-miR5658c targets Serine/threonine protein phosphatase predominantly involved in a wide range of cellular processes, including meiosis and cell division, apoptosis, protein synthesis, metabolism, cytoskeletal reorganization, and the regulation of membrane receptors and channels [38]. Dof zinc finger protein is another important target for csa-miR5658c which plays critical roles as transcriptional regulator of growth and development [39]. Csa-miR5658e target nucleotide-binding site-leucine-rich repeat (NBS-LRR) domain associated with a rapid defense response, termed as hypersensitive response and utility of NBS-LRR R genes for engineering pathogen resistance in crop plants was highlighted by [40].

Table 3. Potential targets of the newly identified miRNAs in *C. sativa*

miRNA	Target Protein	Target Function	Target Genes
csa-miR5021	SLP3 subtilisin-like serine protease 3	Plant Development, defense and signaling	AT2G19170
	Mitogen-activated protein kinase 21	Signal Transduction	AT4G36950
	6F-phosphate phosphohydrolase 2	Intracellular trafficking	AT3G52340
csa-miR3629a-5p	Auxin-repressed protein (ARP1)	Stress response	TC143047
	Pyruvate kinase	Metabolic pathway	TC141417
	DNA binding transcription factors	Transcription factor	AT4G30180
csa-miR5658a	Senescence-associated protein	Transcription factor	TC394397
	Serine/threonine protein kinase	Transcription factor	TC374162
	MYB domain protein 120	Transcription factor	AT5G55020
	BAK1-interacting receptor-like kinase 1	Plant Development	AT5G48380
csa-miR5658b	BAK1-interacting receptor-like kinase 1	Plant Development	AT5G48380
	GRAS family transcription factor	Signal transduction	AT5G59450
	Ubiquitin-like superfamily protein	Cell Physiology	AT5G40630
csa-miR5658c	Protein phosphatase 2A regulatory B	Metabolic pathway	AT3G26020
	Dof zinc finger protein DOF5	Growth and development	TC366889
	Serine/threonine protein phosphatase	Diverse function	TC361599
csa-miR5658d	BAK1-interacting receptor-like kinase 1	Plant Development	AT5G48380
	GRAS family transcription factor	Signal transduction	AT5G59450
	TCP3 Teosinte Branched 1	Transcription factor	AT1G53230
csa-miR5658e	NBS-LRR disease resistance protein	Stress response	AT2G36724
	WD40 repeat-like superfamily protein	Transcription factor	AT3G27640
	MYB transcription factor-like protein	Transcription factor	TC378399
	BAK1-interacting receptor-like kinase 1	Plant Development	AT5G48380

Prediction of potential core promoter and *cis*-regulatory elements of *C. sativa* miRNAs

The upstream sequences up to 2,000 bp from the 5′-end of the entire hairpin structure of miRNA precursor sequence of the *C. sativa* miRNA genes were analyzed by using PlantCARE to reveal known *cis*-regulatory elements that could regulate their expression. As expected, promoter regions of all miRNA genes have been found to contain a TATA-box (a core promoter sequence served for initiation of the transcription process) and a CAAT-box (a common *cis*-element in promoter and enhancer regions) confirming that miRNA genes were transcribed by RNA Polymerase II similar to other protein-coding genes **(Figure 2)** [41]. The *cis*-regulatory element involved in light responsiveness was found to be more prevalent suggesting the important role of light in miRNA gene expression. **Table 4** showed known stress-responsive elements, such as ABRE (*cis*-regulatory element involved in abscisic acid responsiveness) and other stress-relevant elements such as P-box, GARE motifs (*cis*-element involved in gibberellin responsiveness), TC-rich repeats, W-box and WRKY-box (*cis*-element involved in pathogen defense and also reported to involve in other activities) [42]. The *cis*-regulatory elements for phytohormone responses also existed in promoter region of various miRNA genes in *C. sativa* **Table 4**. Interestingly, we found that all identified miRNAs contains putative GARE in its promoter region. This observation hinted that predicted miRNAs in *C. sativa* might regulate itself through the feedback circuit, in order to mediate the balancing between the level of TFs and miRNA expression, which is consistent with previous reports in *Arabidopsis* [43,44]. Moreover, the promoter analysis also showed that WRKY-box, MYB-box and bZIP box are widely distributed as core promoter *cis*-element. The promoter analysis also showed that WRKY-box, MYB-box and bZIP box are widely distributed as *cis*-element in core promoter of csa-miR5658a, csa-miR3629a-5p, csa-miR5658c, csa-miR5658d and csa-miR5658e suggesting their putative role

in flavonoid biosynthesis. The results of this study present a set of putative TFs binding site that could be further investigated for evidence of miRNA regulation in *C. sativa*. These sites could be used to search for a functional connection between TFs and the targets of the miRNAs potentially regulated by them.
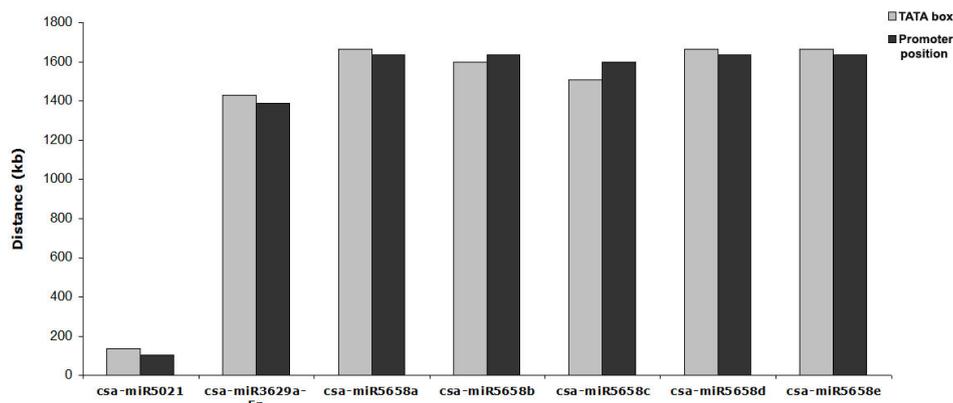

Figure 2. The distribution of the distances between TATA-box and putative promoters: The vertical axis shows the positions of TATA-box and putative promoters with respect to the corresponding microRNA hairpins and the horizontal axis shows the C. sativa miRNA genes.

Table 4. Type of cis-regulatory elements in the upstream regions of miRNAs in *C. sativa*

Cis-regulatory element type	Element function	csa-miRNAs with cis-regulatory element
MYB-box	Involved in the anthocyanin production, flavonoid production, and trichome differentiation	csa-miR5658a, csa-miR3629a-5p, csa-miR5658c, csa-miR5658d, csa-miR5658e
CAAT-box	Enhancing the binding specificity of different transcription factor	csa-miR5021, csa-miR3629a-5p, csa-miR5658e, csa-miR5658a, csa-miR5658b, csa-miR5658c, csa-miR5658d
GATA-box	Involved in light inducible, regulation of the cell cycle and DNA replication and nitrate-dependent control of transcription	csa-miR3629a-5p, csa-miR5658a, csa-miR5658c, csa-miR5658d, csa-miR5658e
TATA-box	Involved in the formation of a transcription factor initiation complex	csa-miR5021, csa-miR3629a-5p, csa-miR5658a, csa-miR5658b, csa-miR5658c, csa-miR5658d, csa-miR5658e
W-box	Involved in regulation of gene expression, pathogen responsive, flavonoid production, anthocyanin production and act as a complex regulator	csa-miR5021, csa-miR5658a, csa-miR5658d, csa-miR5658e
WRKY-box	Involved in biotic and abiotic responses, hormone signaling and secondary metabolism	csa-miR5658a, csa-miR3629a-5p, csa-miR5658c, csa-miR5658d, csa-miR5658e
bZIP box	Involved in pathogen defense, light and stress signaling, seed maturation and flower development	csa-miR5658a, csa-miR3629a-5p, csa-miR5658c, csa-miR5658d, csa-miR5658e
G box, Box I and IV P-box, GARE motifs	Light responsive element Gibberellin-responsive element	csa-miR5021, csa-miR3629a-5p , csa-miR5658a, csa-miR5658b, csa-miR5658c, csa-miR5658d, csa-miR5658e csa-miR5021, csa-miR3629a-5p , csa-miR5658a, csa-miR5658b, csa-miR5658c, csa-miR5658d, csa-miR5658e
TC-rich repeats	Involved in defense and stress responsiveness	csa-miR3629a-5p, csa-miR5658a, csa-miR5658c, csa-miR5658d, csa-miR5658e
TCA	Involved in salicylic acid responsiveness	csa-miR5021, csa-miR5658a, csa-miR5658b, csa-miR5658d, csa-miR5658e
5' -UTR Py-rich stretch	*Cis*-acting element conferring high transcription levels	csa-miR5021, csa-miR5658a, csa-miR5658b, csa-miR5658d, csa-miR5658e
ABRE	Involved in the abscisic acid responsiveness	csa-miR5658a, csa-miR5658b, csa-miR5658d, csa-miR5658e

Validation and Expression Analysis of Selected miRNAs in Cannabis

In this study, identified miRNAs were experimental validated by using a stem-loop RT-PCR by using miRNA-specific primers. The designed primer for specific detection of miRNAs yielded 60-70 bp fragments in the gel **(Figure 3)** and further sequencing confirmed the amplified miRNAs. Real-time PCR expression analysis showed that all of the miRNAs were expressed in leaf, and flower tissues of *C. sativa* **(Figure 4)**. Of the seven miRNAs detected, four miRNAs (csa-miR3629a-5p, csa-miR5021 and csa-miR5658d) were expressed significantly higher in young leaves and flowers **(Figure 4)**.
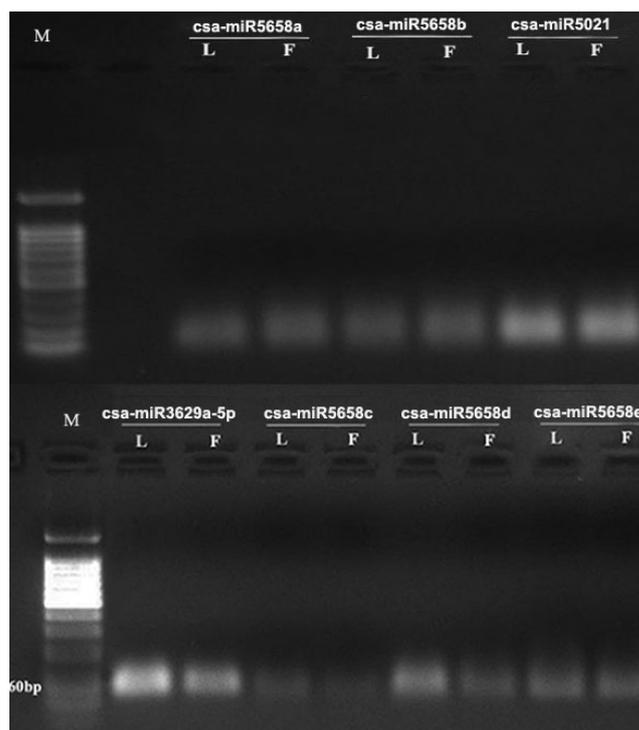
Figure 3. End-point PCR validation of selected *C. sativa* miRNAs isolated from leaf, root and cones. M: 100bp ladder marker.
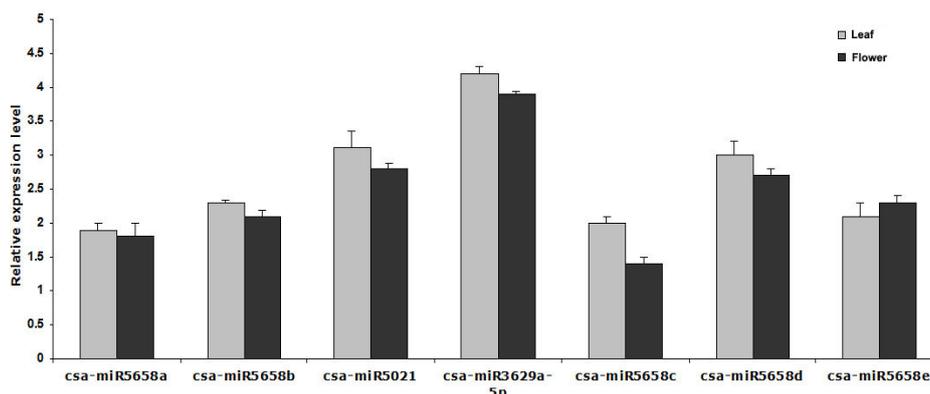

Figure 4. Expression of *C. sativa* miRNAs in different tissues: Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was performed with small RNA isolated from leaf, root and cones. U6 was used as the reference gene.

DISCUSSION

MicroRNAs-related research is one of the hottest research topics in biological or bio-medicinal fields. Many highly conserved miRNAs that exhibit particular expression patterns with specific timing and tissue specificity, have critical functions in growth, development, differentiation, apoptosis, metabolism and biotic and abiotic stress responses, regulating specific target mRNAs. However, no comprehensive research studies have been reported on discovery of novel miRNAs identification and expression analysis in Cannabis plant. In this present study, we identified 7 potential miRNAs along with their potential target genes from a total of 12,907 available *C. sativa* ESTs by homologous search. This indicates that about 0.054% of cannabis ESTs contain potential miRNAs and the ratio is as high as the previously reported about 0.0277% for switch grass [45]. EST based homology search method offer better option to identify miRNAs. This approach has many advantages over the other two approaches for miRNA cloning, namely forward genetics and direct cloning (e.g., library construction and high-throughput sequencing [46].All the characteristic parameters of the miRNAs identified, such as pre-miRNA length, AU content, and MFEI were consistent with previous findings [47]. Unlike animal miRNAs whose precursors are usually less than 100 nt [48] plant pre-miRNA length varies greatly and ranges from 60 to more than 600 nucleotides. We have also observed varying degree of sequence dissimilarities among precursor sequences dataset from diverse plant species, mainly in the sequences flanking the region of mature miRNA used in this study suggesting that different members of the same miRNA family evolved at different rates within plant species.

Considerable number of miRNAs with longer pre-miRNAs ranging from 200 to 409 nt were identified. It has been reported that most miRNA precursors contain more A+U nucleotides than G+C [48]. Similarly, we found that the A+U contents of all *C. sativa* pre-miRNAs were higher than G+C content. In addition, the predicted miRNA hairpin structures showed that there are at least

12-21 nt engaged in Watson-Crick or G/U pairings between the miRNA/miRNA* in the stem region and do not contain large internal loops or bulges. The other major parameters of pre-miRNAs are the length of mature miRNAs and miRNA precursors, the nucleotide composition, and higher MFEI value compared to some other types of RNA molecules, since they form a stable stem-loop structure. In cannabis miRNAs, we found majority of pre-miRNAs predicted have high MFEI values, higher than other types of noncoding RNAs e.g., tRNAs (0.64), rRNAs (0.59) and mRNAs (0.62–0.66). It supports the previous report that, miRNA precursor sequences have significantly higher MFEI value than other non-coding or coding RNAs e.g. tRNA, rRNA [49]. All of these data suggest that the cannabis miRNAs identified in this study are genuine miRNAs. miRNA target gene identification is an important step. In the recent research, several attempts have been made to explore the active role of miRNAs in various cellular functions and gene regulation networks [50]. The conserved nature of miRNAs in different organisms suggests their conserved function and their targets (mostly TFs) which affect plant development and specific genes [51]. Our prediction of target genes for the C. sativa miRNA discovered that more than one gene can be regulated by individual miRNA. This result is accordance with some recent research in other plant species [47,18,19] which also suggest that miRNA research should be focused on regulatory networks rather than individual connections between miRNA and their predicted target genes or regulatory factors. Some miRNAs directly target TFs which directly or indirectly affect growth and development and also specific genes which control metabolism of plants [19]. These conserved miRNAs and their conserved target TFs again highlight their versatile functions, and provide further evidence of the phylogenetic distribution of miRNA families irrespective of species boundaries.

The preferential expression of miRNAs in specific tissues or developmental stages suggests a role in development [52]. We quantified changes in the expression of the seven selected miRNAs in C. sativa leaf and flower tissues using stem-loop real-time RT-PCR-a high throughput quantification methods for authenticating miRNAs. Result showed that the expression profiles of the different miRNAs differed significantly. The difference in expression patterns of miRNAs suggests that the expression pattern can differ among different plants despite the conserved nature of miRNAs [53]. There is growing evidence of role of bZIPs TFs in modulating the accumulation of flavonol glycosides, phenolic acids, and anthocyanin pigmentation in infiltrated *Petunia hybrida* leaves and involvement of several TFs involve in flavonoid biosynthesis in hop through the regulation of chalcone synthase (e.g. HlMYB1, HlMYB2, HlMYB3, HlMYB7, HlbHLH2 and HlWDR1 [54]. Therefore, it is possible that csa-miR414 and csa-miR1886 could be involved in regulation of genes of the cannabinoids biosynthesis pathway in *C. sativa*. Differential expression of various identified miRNAs implies their involvement in various physiological and developmental processes including cannabinoids biosynthesis, which needs to be unraveled in future.

Micro RNA genes are mainly transcribed by RNA Pol II [55] resulting primary transcript is capped at the 5′ end and polyadenylated at the 3′ end similar to mRNAs and the abundance of pri-miRNAs ultimately determines the level of mature miRNAs present in the cell. Identification and analysis of *cis*-regulatory elements of miRNA genes provides important temporal and spatial measurements regarding transcription initiation, and therefore are useful to illustrate the regulatory networks in a broad range of plant species. Many studies indicated the characteristics of miRNAs promoters, including the relative frequencies of CpG, TATA box, TFs binding site recognition, initiator elements and other chromatin signatures, are similar to protein-coding genes [56]. The core promoters and *cis*-acting elements of predicted miRNAs from *C. sativa* revealed the general features of the core promoters and provide insights into the transcriptional regulation and functions of miRNAs.

Cannabinoids are the best known group of natural products and more than 80 different cannabinoids have been found so far. Several therapeutic effects of cannabinoids have been described and the discovery of an endo-cannabinoid system in mammals marked a renewed research interest in these compounds [4]. The dissection of biosynthetic pathway(s) of this compound and its regulation by TFs is important for efficient biotechnological manipulation of the secondary metabolome in Cannabis. The promoters of polyketide synthesis contains TATA box, E-box and W-box *cis*-acting elements, and considered to be binding sites of WRKY, MYB, WDR and bZIP1A TFs [57]. They also provide a basis for further research on the functional role of miRNA and the regulation of cannabinoid biosynthesis in *C. sativa*. Findings of our study will be supportive to understand the gene regulation mechanism and the biogenesis of miRNA in the near future. It also fortifies the present bioinformatics approach for new miRNAs identification from plant species whose genome is not yet fully known. Taken together, the knowledge gained from this research will provide an insight into *C. sativa* miRNAs-mediated genes network which would undoubtedly be hugely beneficial for the further researches of miRNAs function and regulatory mechanisms of cannabinoid biosynthesis pathway

ACKNOWLEDGEMENTS

REFERENCES

1. Clarke RC and Watson DP. Cannabis and Natural Cannabis Medicines. Forensic Sci Med. 2007;35:1-15.

2. Hillig KW. Genetic evidence for speciation in Cannabis (*Cannabaceae*). Genet Res Crop Ev. 2005;52:161-182.

3. FloresSanchez IJ and Verpoorte R. Secondary metabolism in cannabis. Phytochem Rev. 2008; 7: 615-639.

4. Marks MD, et al. Identification of candidate genes affecting ᐞ9-tetrahydrocannabinol biosynthesis in *Cannabis sativa.* J Exp Bot. 2009; 60:3715-3726.

5. Palatnik JF, et al Control of leaf morphogenesis by microRNAs. Nature. 2003;425:257-263.

6. Aukerman MJ and Sakai H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell. 2003;15:2730-2741.

7. Chen X. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science. 2004;303:2022-2025.

8. Gu M, et al. Expression analysis suggests potential roles of microRNAs for phosphate and arbuscular mycorrhizal signaling in *Solanum lycopersicum*. Physiol Plant. 2010; 138:226-237.

9. Zhao M, et al. Involvement of miR169 in the nitrogen-starvation responses in *Arabidopsis*. New Phytol. 2011;190:906-915.

10. Valdes Lopez O, et al. MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity. New Phytol. 2010;187:805-818.

11. Lakhotia N, et al. Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (*Solanum tuberosum* L.) by high-throughput sequencing. BMC Plant Biol. 2014;14:6.

12. Rapaport F, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. 2013;14:R95.

13. Zhang BO, et al. Identification and characterization of new plant microRNAs using EST analysis. Cell Res. 2005;15:336-360.

14. Zhang BH, et al. Conservation and divergence of plant microRNA genes. Plant J 2006;46:243-259.

15. Han YS, etal. Computational identification of microRNAs and their targets in wheat (*Triticum aestivum* L.). Sci China Ser C-Life. 2009;52:1091-110.

16. Chen R, et al. Identification of MicroRNAs in Wild Soybean (Glycine soja) J. Integr. Plant Biol. 2009;51:1071-1079.

17. Frazier TP, et al. Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). Planta. 2010;232:1289-1308.

18. Xie F, et al. Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*). Gene 2011;473:8-22.

19. Wang M, etal. Identification and characterization of MicroRNAs in Asiatic cotton (*Gossypium arboreum* L.). PLoS ONE 7:e33696.

20. Jackson A and Linsley PS. The therapeutic potential of microRNA modulation. Discov med. 2010;9:311-318.

21. Darling A, et al. The Design Implementation, and Evaluation of mpiBLAST. In: 4th International Conference on Linux Clusters: The HPC Revolution 2003 in Conjunction with the Cluster World Conference Expo. 2003;San Jose,CA.

22. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-3402.

23. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406-3415.

24. Dai X and Zhao PX. PsRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011;39:W155-W159.

25. Zhang BH, et al. Identification of cotton miRNA and their targets. Gene. 2007; 397:26-37.

26. Jones Rhoades MW and Bartel DP. Computational identification of plant microRNAs and their targets, including a stress induced miRNA. Mol Cell. 2004;14:787-799.

27. Shahmuradov IA, et al. Plant promoter prediction with confidence estimation. Nucleic Acids Res. 2005;33:1069-1076.

28. 28 Liu YX, et al. Genomic analysis of microRNA promoters and their *cis*-acting elements in soybean. Agric Sci China. 2010; 9:1561-1570.

29. Lescot M, et al. PlantCARE: A database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325-327.

30. Varkonyi Gasic E, et al. Protocol, a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods. 2007;3:12.

31. Tang F, et al. MicroRNA expression profiling of single whole embryonic stem cells. Nucleic Acids Res. 2006;34:e9.

32. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. MiRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:D140-D144.

33. Bonnet E, et al. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics. 2004;20:2911-2917.

34. Lai X, et al. Computational analysis of target hub gene repression regulated by multiple and cooperative miRNAs. Nucleic Acids Res. 2012; 40:8818-8834.

35. Gohar T, et al. MAPK machinery in plants, Recognition and response to different stresses through multiple signal transduction pathways. Plant Signal Behav. 2010;5:1370-1378.

36. Golldack D, et al. Expression of subtilisin-like proteases in Arabidopsis thaliana is cell specific and responds to jasmonic acid and heavy metals with developmental differences. Physiol Plant. 2003;118:64-73.

37. Kim MH, et al. Identification of Arabidopsis BAK1-associating receptor-like kinase 1 (BARK1) and characterization of its gene expression and brassinosteroid-regulated root phenotypes. Plant Cell Physiol. 2013;54:1620-1634.

38. Ceulemans H, Bollen M. Functional diversity of protein phosphatase-1, a cellular economizer and reset button. Physiol Rev 2004; 84:1-39.

39. Yanagisawa S. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol. 2004;45:386-391.

40. Sharma TR,et al. Resistance gene analogues as a tool for rapid identification and cloning of disease resistance genes in plants - a review. J Plant Biochem Biotechnol. 2009;18:1-11.

41. Cui M, et al. Genes involved in pre-mRNA 3′-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen. Proc Natl Acad Sci. 2009;105:16665-16670.

42. Rizhsky L, et al. The zinc-finger protein Zat12 is required for cytosolic ascorbate peroxidase 1 expression during oxidative stress in *Arabidopsis*. J Biol Chem. 2004;279:11736-11743.

43. Meng W, Su YCF, Saunders RMK, Chye ML. The rice acyl-CoAbinding protein gene family, phylogeny, expression and functional analysis. New Phytol 2011; 189:1170–1184.

44. Hichri I, et al. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot. 2011; 62:24650-2483.

45. Xie Z,et al. Expression of microRNAs and its regulation in plants. S*emin.* Cell Dev Biol. 2010;21:790-797.

46. Jones Rhoades MW, et al. MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol. 2006;57:19-53.

47. Frazier TP, et al. Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). Planta. 2010;232:1289-1308.

48. Ambros V. The functions of animal microRNAs. Nature. 2004;431:350-355.

49. Ambros V and Chen XM. The regulation of genes and genomes by small RNAs. Development. 2007;134:1635-1641.

50. Dhananjay K, et al. Discovery of novel leaf rust responsive microRNAs in wheat and prediction of their target genes. J Nucl Acids. 2014;2014:12.

51. Jain M, Narayana Chevala VVS, Garg R. Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. J Exp Bot. 2014;7:1-14.

52. Sunkar R and Zhu JK. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell 2004;16:2001-2019.

53. Yao Y, et al. Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). Genome Biol. 2007;8:R96.

54. Matousek J, et al. HlMyb3, a putative regulatory factor in hop (*Humulus lupulus* L.), shows diverse biological effects in heterologous transgenotes. J Agric Food Chem. 2007;55:77670-7776.

55. Lee Y, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004; 23:4051-4060.

56. Ozsolak FLL, Poling Z, Wang H, Liu XS, Liu RG, Roeder RGM, Zhang X, Song JS, Fisher DE. Chromatin structure analyses identifies miRNA promoters. Genes Dev. 2008;22:172-3183.

57. Dao TTH, et al. Chalcone synthase and its functions in plant resistance. Phytochem Rev. 2011;10:397-412.