TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2018) 42: 382-399

doi:10.3906/bot-1712-37

Bioinformatics prediction and annotation of cherry (*Prunus avium* L.) microRNAs and their targeted proteins

Iftekhar Ahmed BALOCH, Muhammad Younas Khan BAROZAI, Muhammad DIN*
Department of Botany, University of Balochistan, Quetta, Pakistan

Received: 18.12.2017 • **Accepted/Published Online:** 25.04.2018 • **Final Version:** 24.07.2018

Abstract: MicroRNAs (miRNAs) are important noncoding regulatory RNAs. They are expressed endogenously and are 18–26 nucleotides in length. These small molecules are conserved evolutionarily within the same kingdom and their conserved nature becomes an important logical tool for the quest of conserved miRNAs in other species by homology search using bioinformatics tools. Cherry (*Prunus avium* L.) is one of the important and nutritious species of the family Rosaceae, mainly distributed in temperate climate. This research is an attempt to identify and characterize conserved miRNAs in cherry from express sequence tags (ESTs). Bioinformatics analysis of cherry ESTs resulted in the identification of 91 conserved miRNAs belonging to 88 miRNA families. Among the identified miRNAs, two of the conserved miRNAs (pav-miR482 and pav-miR535) were found to be transcribed in the opposite direction of the same genomic locus (sense/antisense orientation). A miRNA (pav-miR482a) was identified as a precursor miRNA (pre-miRNA) cluster while another (pav-mir161) was identified with two overlapping mature miRNA sequences. In addition, highly conserved pre-miRNA was also found, i.e. pav-mir535, which showed 100% query coverage and 98% identity with the peach pre-miRNA ppe-miR535a. Moreover, 14 predicted miRNAs were selected randomly for experimental validation through RT-PCR. Experimental validation of these 14 microRNAs endorses the powerfulness of bioinformatics prediction of miRNAs. The 91 miRNAs were also subjected to study of their 211 protein targets. These were involved in various biological processes including cell signaling, growth and development, transcription factors, and stress management. The results of this research will contribute in understanding the miRNA-mediated life processes in cherry.

Key words: Cherry (*Prunus avium* L.), conserved precursor microRNA, RT-PCR

1. Introduction

MicroRNAs (miRNAs) are important, noncoding, small regulatory molecules. They are endogenously expressed and range from 18 nt to 26 nt in length. They have a significant posttranscriptional role in gene regulation (Ambros et al., 2003). Commonly RNA polymerase II is responsible for miRNA gene transcription. The resulting product, known as primary miRNA (pri-miRNA), may range up to thousands of nucleotides in length and may contain many miRNA hairpin stem loops. The 5' end of the pri-miRNA transcript is capped while its 3' end is polyadenylated and spliced (Meyers et al., 2008). Later the pri-miRNA self-hybridizes into a hairpin structure, creating precursor miRNAs (pre-miRNAs). The pre-miRNAs are transported to the cytoplasm where a special RNase III protein, Dicer, generates an unstable duplex miRNA structure. In plants this process is accomplished by the Dicer like 1 enzyme (DCL1) (Meyers et al., 2008). This duplex is later assimilated into the RNA-induced silencing complex (RISC). RISC along with miRNA acts as a negative regulator by hindering the translation process or by causing messenger RNA (mRNA) destruction. miRNAs perform important functions in plants during growth, organ development (Barozai, 2012a, 2012b), abiotic stresses, signaling processes, transgene inactivation, and defense against offensive microorganisms (Barozai et al., 2013).

Rosaceae is a medium-sized angiosperm family comprising over 100 genera and 3000 species. It derives its name from the type genus *Rosa*. Plants of this family have a cosmopolitan distribution, except for Antarctica. Sweet cherry (*Prunus avium* L.) is an important and nutritious species of the family Rosaceae geographically distributed in temperate climates. Turkey is the top cherry-producing country in the world (Campoy et al., 2016). In Pakistan, Balochistan is famous for the production of delicious cherries, grown on 1085 ha on a commercial basis with an annual production of about 2027 t (http://www.balochistan.gov.pk/)

* Correspondence: drmdin75@gmail.com

miRNA researchers strive to find miRNAs in the plants of Rosaceae by using both bioinformatics and wet-lab approaches, but still there is a need to identify more new conserved miRNAs in plants of this family. The increasing genomic data of this family, new expressed sequence tags (ESTs), and the sequenced genomes invite researchers to explore new miRNAs. Although miRNA researchers have reported miRNAs in peach (Zhang et al., 2013) and apple (Ma et al., 2014), until now cherry has remained neglected. Therefore, the need is felt to identify and annotate new conserved miRNAs in cherry.

2. Materials and methods

Use of bioinformatics tools is a widely used method to forecast new conserved miRNA orthologs by a comparative approach (Barozai, 2012a, 2012b). Barozai's (2013) methodology with slight alteration was used to predict new potential miRNAs in cherry. The new conserved miRNAs in cherry were identified and characterized using a variety of bioinformatics tools including BLASTn, BLASTx, Mfold, psRNA Target, Clustal W, Primer 3, and WebLogo.

2.1. Survey of miRNA databases and available literature for reported miRNAs in cherry

The famous miRNA repositories, i.e. miRBase (Griffiths-Jones, 2004) and the Plant microRNA Database (PMRD) (Zhang et al., 2010), and available miRNA literature were surveyed for the reported and nonreported miRNAs in cherry, which indicated that there was not even a single miRNA reported for cherry. Thus, an attempt was made to profile the new conserved miRNAs in cherry.

2.2. Survey of the Expressed Sequence Tags Database (dbEST)

A survey of dbEST was made to check the availability of cherry ESTs. The latest release of dbEST (release 130101, 1 January 2013) contains a total of 6035 ESTs for cherry. These cherry ESTs were computationally screened against the known plant miRNAs for the homology search.

2.3. Fetching of miRNA sequences

Since this research is logically based on a homology/similarity search approach, previously reported mature and pre-miRNA sequences of diverse plant species (*Prunus persica*, *Malus domestica*, *Brassica napus*, *Brassica rapa*, *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Oryza sativa*, *Cynara cardunculus*, *Helianthus annuus*, *Panax ginseng*, *Carica papaya*, *Ricinus communis*, *Brachypodium distachyon*, *Cucumis melo*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Glycine max*, *and Picea abies*) were fetched from miRBase (Griffiths-Jones, 2004) and PMRD (Zhang et al., 2010). These sequences were used as reference miRNAs to find their new potential orthologs in cherry.

2.4. Identification of potential candidate miRNA sequences

The reference miRNA sequences from the different plant species were used as queries and subjected to BLAST (Altschul et al., 1990) against the publicly available ESTs of cherry in the NCBI GenBank by using the BLASTn program. To find the candidate homolog sequences, the homology-based search was started with the miRNA sequences of the family Rosaceae. The candidate sequences with a maximum of 4 mismatches were saved.

2.5. Creation of single tone EST

Since many discrete ESTs are often partial sequences that correspond to the same mRNA of an organism, a single representative EST was generated to avoid false positive results. This was accomplished by using BLASTn (Altschul et al., 1990).

2.6. Validation of potential miRNA candidates as nonprotein-coding sequences

The newly predicted candidate pre-miRNAs sequences were subjected to BLAST against the NCBI protein database using BLASTx (Altschul et al., 1997) with default parameters to validate them as nonprotein-coding RNAs. The protein-coding pre-miRNA sequences were discarded.

2.7. Prediction of hairpin structures of potential miRNA candidates

The hairpin structures of candidate sequences were created by using MFOLD (version 3.6) (Zuker, 2003) with default parameters.

2.8. Physical inspection of hairpin structures

The predicted structures having the lowest free energy were selected for physical inspection. The stem portions of the miRNAs were manually checked for mature sequences with either 10 base pairs or equal to the reference miRNAs involved in Watson–Crick and non-Watson–Crick (G/U wobble) pairing between the mature miRNA and its opposite strand (miRNA*) in the duplex. The threshold values for the selection of a miRNA were the same as descried by Zhang et al. (2006).

2.9. Conservation analysis and phylogenetic analysis of newly identified miRNAs

Many miRNA families are conserved evolutionarily in plants (Zhang et al., 2006), which offers a sound approach for the bioinformatics identification of new conserved miRNAs (Barozai et al., 2008; Baloch et al., 2015). Therefore, one of the newly identified conserved miRNAs from cherry (pav-miR172) was subjected to conservation analysis with its orthologs in different plant species. For this purpose, WebLogo software (Crooks et al., 2004) was used. The WebLogo result was saved and scrutinized for conservation of precursor and mature miRNA sequences. One of the newly identified miRNAs from cherry (pav-miR169) was selected for phylogenetic studies. The

cladogram was created by neighbor-joining clustering method. The result was saved.

2.10. RT-PCR validation

Fourteen of the cherry miRNAs were selected randomly for reverse transcription polymerase chain reaction (RT-PCR) experimental validation. The Primer-3 algorithm was used to design the primers against the stem-loop sequences of the selected miRNAs from their ESTs (Table 1). Total RNA was extracted from the leaves of cherry using a CTAB-based optimized method (Wahid et al., 2015). Complementary DNA was synthesized using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas), as per the supplier's protocol. cDNA (100 ng) was used as a template for the PCR. The PCR was automated as follows: initial denaturation at 95 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 35 s, annealing at 60 °C for 35 s, and extension at 72 °C for 30 s with a final elongation step at 72 °C for 10 min. Later the PCR products were separated on 1.8% (w/v) agarose gel.

2.11. Prediction of miRNA targeted genes

The finding of new conserved miRNA targets is an important phase for the confirmation of identified miRNAs. To predict miRNA targets, the mature sequences of miRNAs were subjected to psRNATarget (Dai and Zhao, 2011). As no Refseq-RNAs are available for cherry from the NCBI and it is not among the listed organisms

Table 1. Forward and reverse primers for cherry miRNAs designed against precursor sequences using primer3.

Cherry miRNAs	Tm	Primers	Product size	EST
pav-miR156	59.99 60.03	AGTCTGGAAGCTGACCTGGA TCCAACAACTGCTGCTTCAC	179	HE641644.1
pav-miR160	60.39 59.19	CGCATGTTTTCAGCCTTTCT TGCAGCCACTGTAGAAGGAG	168	HE645242.1
pav-miR172	59.59 59.91	GCTGATGGAGGAGCTGTAGAA ACAACAGAGCCCAAGCTGTT	154	HE644209.1
pav-miR395	60.37 59.94	TTGGGGGAGTAAATGGGAAT AAATTCCCTGCAAAGGAAAAA	157	HE643141.1
pav-miR398	60.07 59.98	AATGGGCCCTGTACCCTTAC CTGGTGGGTACGAATCTCAAA	106	HE645551.1
pav-miR403	59.77 60.26	GGGTTAAGATTCGGGTCTCC CCAGTGTTCCTGATCCTGCT	183	HE644875.1
pav-miR413	59.75 59.99	TTGCTTCCAACACAATGTCTG TGCAAAGTTGGCTGAATCTG	179	HE642851.1
pav-miR414	59.94 59.84	CTCCTTCATCGTCAGCATCA GATTCGTTTCCTCAAGGCTCT	104	HE642574.1
pav-miR435	60.29 59.09	CAGCATCTCTCACTGCCTGA AGGGTCATCCTTCTGATTCG	183	HE645656.1
pav-miR482	59.93 59.97	GGGGAATGGAAGTTGCAGTA ACGCAACAAAATGGGAAAAG	176	HE644421.1
pav-miR530	60.46 60.05	TCATGGGCTTGAAACCATCT TGATTCCACCTTGCAAATGA	197	HE644656.1
pav-miR775	59.72 60.43	GGGGCCATATTGATTTCTCA ACGTGGTGTGTCCGAAGAAT	167	HE641063.1
pav-miR858	60.32 60.48	TTTGATGATCCGGTGGAAGT AGAGGCAACATATCCGGTGA	182	HE641269.1
pav-miR1507	60.43 59.85	TTGGGGAGATTTCTTGACCA CCAATTTAGCCGACAAGCTC	173	HE642607.1

in psRNATarget, the targets for the miRNAs identified in cherry were predicted by using preloaded transcript/genomic library sequences of peach (a closely related plant to cherry) and the model plant, i.e. *Arabidopsis thaliana*. Gene Ontology functional and enrichment analyses were conducted through GOrilla (Eden et al., 2009), while Cytoscape 3.3 (Shannon et al., 2003) was used for network interaction of miRNAs and their target genes in cherry. The target validation studies of some selected cherry miRNAs were done through RNAhybrid (Kruger and Rehmsmeier, 2006).

3. Results

3.1. The new conserved cherry miRNAs

Comparative genomics based on homology is a well-known approach for interesting findings in various organisms (Wahid and Barozai, 2016; Jahan et al., 2017; Shah et al., 2017). Bioinformatics analysis of cherry ESTs resulted in the identification of 91 new conserved miRNAs. The 91 potential cherry miRNAs belonged to 88 miRNA families, i.e. miR156, 157, 158, 160, 161, 164, 166, 167, 169, 171, 172, 391, 394, 395, 396, 398, 403, 408, 413, 414, 418, 426, 435, 444, 477, 482, 528, 530, 535, 537, 773, 774, 775, 780, 838, 840, 844, 845, 854, 856, 858, 860, 1088, 1089, 1096, 1098, 1139, 1507, 1512, 1523, 1535, 1888, 2079, 2082, 2084, 2085, 2111, 2118, 2275, 2933, 2938, 3440, 3627, 3704, 3709, 3710, 3932, 5019, 5057, 5069, 5185, 5643, 5656, 6029, 6034, 6114, 6135, 6260, 6290, 7122, 8122, 8131, 8135, 8148, 8154, 9554, 9559, and 9563. To consider the newly predicted cherry miRNAs as valid candidates, the empirical formula of biogenesis and expression of miRNAs suggested by Ambros et al. (2003) was used as a criterion. All the identified cherry pre-miRNAs satisfied criteria B, C, and D. Criterion D alone is enough for orthologous sequences to be validated as new conserved miRNAs in different species (Ambros et al., 2003). This was further confirmed by Meyers et al. (2008) in favor of plants.

3.2. Characterization of cherry miRNAs

The newly identified conserved cherry miRNAs were characterized for their reference miRNA precursor length (PL), minimum free energy (MFE), mature sequences (MS), mature sequence arms (MSA), mature sequence length (ML), number of mismatches (NM), source ESTs (SE), and strand orientation (Table 2).

The new cherry miRNAs showed similarity to the miRNAs of different plant families, signifying the conserved nature of miRNAs across the diverse plant species of different families (Table 3).

A lower MFE shows the higher thermodynamic stability of the secondary structure of a miRNA. Therefore, the MFE of the newly identified cherry pre-miRNAs is a key feature of miRNA characterization. As predicted by Mfold (Zuker, 2003), the MFE of the new conserved cherry miRNAs ranged from –6 Kcal mol^{-1} to –59.9 Kcal mol^{-1} with an average of –26 Kcal mol^{-1}.

The pre-miRNAs give rise to mature miRNAs (Bartel, 2004). In contrast to animal pre-miRNA, plant pre-miRNAs are wide-ranging in structure and size (Zhang et al., 2006). The conserved cherry pre-miRNAs also showed a wide range of lengths, ranging from 46 to 257 nt with an average of 102 nt.

The new conserved mature cherry miRNA sequence lengths range from 18 to 23 nt. The majority (63%, i.e. 58 out of 91) of the cherry miRNAs have length of 21 nt, followed by 20 nt (19%, i.e. 16 out of 91), 22 nt (13%, i.e. 12 out of 91), 19 and 23 (2% each, i.e. 2 each out of 91), and 18 nt (1%, i.e. 1 out of 91).

New conserved cherry miRNAs were also characterized for the permissible range (Zhang et al., 2006) of mismatches between the reference miRNA and newly identified miRNAs. A majority (49%, i.e. 45 out of 91) of cherry miRNAs were observed to show 4 mismatches with their orthologs, followed by 3 (31%, i.e. 28 out of 91), 2 (12%, i.e. 11 out of 91), 1 (5%, i.e. 5 out of 91), and 0 (2%, i.e. 2 out of 91) mismatches.

Mature cherry miRNA characterization was also done for the location in the pre-miRNA. The cherry miRNAs showed residence on both the 5' and 3' arms of the pre-miRNAs. A majority (76%, i.e. 69 out of 91) of cherry miRNAs are located on the 5' arms and the remaining (24%, i.e. 22 out of 91) are on the opposite 3' arms of the pre-miRNA secondary structures.

The new conserved cherry miRNAs were further characterized for Watson–Crick or G/U base pairing. All the predicted cherry miRNA stem-loop structures showed at least 10 nt engaged in Watson–Crick or G/U base pairings between the mature miRNA and the opposite arms (miRNAs*) in the stem region. No large internal loops or bulges were observed in hairpin precursors.

The new conserved cherry miRNAs were also characterized in terms of organ of expression. All of the new miRNAs were predicted in different ESTs of cherry leaf at the swollen bud stage of leaf.

Similar ranges of MFE, number of mismatches, mature miRNA length, pre-miRNA length, Watson–Crick or G/U pairing, and 3'/5' residency have been reported in various plants by different researchers in *Phaseolus* (Barozai et al., 2013), tomato (Din and Barozai, 2014a), eggplant (Din and Barozai, 2014b), coffee (Bibi et al., 2017), switchgrass (Barozai et al., 2018a), and *Porphyridium cruentum* (Barozai et al., 2018b). The agreement of results with previous reports reinforces the validation of the cherry miRNAs.

The cherry pre-miRNAs were subjected to BLASTx (Altschul et al., 1997) against the protein database of the NCBI and no homology was found with the proteins. The result confirms the newly identified pre-miRNAs as strong candidate miRNAs in cherry.

Table 2. Characterization of the newly identified conserved cherry miRNAs. The cherry miRNAs were characterized in terms of precursor miRNA length (PL), minimum free energy (MFE), mature sequence (MS), number of mismatches (NM), mature sequence length (ML), source EST (SE), mature sequence arm (MSA), GC content percentage (GC%), and organ of expression (OE).

Cherry miRNAs	Ref. miRNAs	PL	MFE kcal/mol	MS with their positions in precursor	MSA	ML	NM	SE	SO	OE
pav-miR156	ppe-miR156a	92	–22.04	3-TGACAGAAGA**T**AGAGAGCAC-22	5’	20	1	HE641644	Minus	Leaf bud
pav-miR157	par-miR157	84	–56.90	3-TTGAC-GA**C**GAGAGAGAGCACG-23	5’	21	2	HE644699	Plus	Leaf bud
pav-miR158	ath-miR158a	60	–9.10	31-TCCCAAATGTAGACA**CT**G**A**A-50	3’	20	3	HE641993	Minus	Leaf bud
pav-miR160	cca-miR160a	70	–12.40	7-**A**GA**C**T**T**GCT**G**CCTGGATGCCA-27	5’	21	4	HE645242	Minus	Leaf bud
pav-miR161, overlapping	aly-miR161-5p.1	68	–18.40	1-T**T**AA**G**GAATTGAAAGTGA**AA**A-21	5’	21	4	HE641740	Minus	Leaf bud
	aly-miR161-5p.2			9-TTGAAAGTGA**AA**ACATC**A**GG**C**-29						
pav-miR164	cca-miR164	257	–58.44	11-TGGAGAAGCAG**A**G**C**ACG**A**G**T**A-31	5’	21	4	HE644349	Plus	Leaf bud
pav-miR166	osa-miR166a	64	–16.30	1-**CA**AATGTTGTCTGGTTC**T**AGG-21	5’	21	3	HE643064	Plus	Leaf bud
pav-miR-167	mdm-miR167c	180	–46.50	60-**AC**AAGCTGCCAGC**G**TGATCT**C**-80	3’	21	4	HE640760	Minus	Leaf bud
pav-miR169	psl-miR169	126	–24.90	13-AGCCA**TTG**ATGACTTGCTG**G**-32	5’	20	4	HE640542	Plus	Leaf bud
pav-miR171	bna-miR171f	90	–27.30	3-TG**GC**TGAG**A**CGCGCCAATATC-23	5’	21	3	HE644266	Minus	Leaf bud
pav-miR172	ppe-miR172d	79	–14.20	5- GGAATC**C**TGATGATGCTGCA**A**-25	5’	21	2	HE644209	Minus	Leaf bud
pav-miR391	aly-miR391	108	–51.00	17-T**A**CGCAGGAGAGAT**G**GCGC**TG**-37	5’	21	4	HE640724	Plus	Leaf bud
pav-miR394	bna-miR394a	141	–37.95	1-TTGGC-TTCT**TTG**CACCTCCTT-21	5’	21	3	HE645729	Minus	Leaf bud
pav-miR395	aly-miR395c	94	–45.53	6-C**C**GA**C**GTGTTTGGGGGGA**T**TT-26	5’	21	3	HE643141	Minus	Leaf bud
pav-miR396	ppe-miR396a	110	–24.70	78-TTCCACA**G**GCTTTCTTGAACG-98	3’	21	1	HE645733	Minus	Leaf bud
pav-miR398	rco-miR398a	56	–12.20	1-TGTGTTCTCA**AA**TCACCCC**A**T-21	5’	21	3	HE645551	Plus	Leaf bud
pav-miR403	ath-miR403	71	–14.40	5-**G**GTTTTGTGCTTGAA**CT**TCTAA-26	5’	22	3	HE644875	Minus	Leaf bud
pav-miR408	ppe-miR408	117	–29.50	4-**T**TGCACTGC**G**TC**G**TCCCTG**T**C-24	5’	21	4	HE642547	Minus	Leaf bud
pav-miR413	ath-miR413	65	–12.70	1-**TA**AG**G**TT**T**TCTTGTTCTGCAC-21	5’	21	4	HE642851	Minus	Leaf bud
pav-miR414	osa-miR414	59	–12.80	7-TCATCCTCATCATCATC**A**TCC-27	5’	21	1	HE642574	Plus	Leaf bud
pav-miR418	osa-miR418	61	–14.40	1-**C**AATGTGATGATGA**T**ATGA**AG**-21	5’	21	4	HE646011	Plus	Leaf bud
pav-miR426	ath-miR426	71	–11.70	5-TTTTGGAAATTT**A**T**GT**TTAC**T**-25	5’	21	4	HE641919	Plus	Leaf bud
pav-miR435	osa-miR435	129	–18.65	14-**G**TATCCGGTATTGGAGTT**T**A-33	5’	20	2	HE645656	Minus	Leaf bud
pav-miR444	bdi-miR444a	74	–20.10	41-TT**T**CTGCCTCAAGCTTG**ATC**C-61	3’	21	4	HE645776	Minus	Leaf bud
pav-miR477	cpa-miR477	68	–18.30	3-ATTGGAGGACTTTGGGGG**GCC**-23	5’	21	3	HE643867	Minus	Leaf bud
pav-miR482	ppe-miR482a	118	–44.05	87-**C**TTCC**C**AAACCTCCCATTCC**T**A-108	3’	22	3	HE644421	Plus	Leaf bud
pav-miR482 antisense		115	–45.40	86-TTTCC**C**AATCCTCCCATTCC**CC**-107	3’	22	3		Minus	Leaf bud
pav-miR482a cluster	ppe-miR482b	136	–52.27	21-GGAATGGGAGGATTGGGAAAA-41	5’	21	0	HE644421	Plus	Leaf bud
				95-CTTCCCAAACCTCCCATTCCTA-116	3’	22	0			Leaf bud
pav-miR528	osa-miR528	152	–48.00	3-GGGA**T**GGGG-ATGCAGAGGAG-23	5’	20	3	HE644751	Minus	Leaf bud
pav-miR530	cme-miR530b	129	–30.64	5-T**T**CAT**CA**GCACCTACACCTT-24	5’	20	3	HE644656	Minus	Leaf bud
pav-miR535	ppe-miR535a	89	–59.90	6-TGAC**G**ACGAGAGAGAGCACGC-26	5’	21	1	HE644699	Plus	Leaf bud
pav-miR535 antisense		89	–47.60	6-TGACAAC**A**AGAGAGAGCACGC-26	5’	21	1		Minus	Leaf bud
pav-miR537	ppt-miR537a	140	–37.30	7-TTGAGGTGTTTCT**C**CA**C**GCT**C**-27	5’	21	3	HE640748	Minus	Leaf bud
pav-miR773	ath-miR773b	139	–18.98	2-**A**GC**T**ATAACTTGAGCA**GC**CA-21	5’	20	4	HE643843	Minus	Leaf bud
pav-miR774	ath-miR774b	69	–18.60	48-TGA**A**ATGAAGATATGGG**AC**AT-68	3’	21	3	HE640413	Plus	Leaf bud
pav-miR775	ath-miR775	113	–24.90	1-TTC**C**AT**A**TCTAGCAGTGCCA-20	5’	20	2	HE641063	Plus	Leaf bud

Table 2. (Continued).

pav-miR780	ath-miR780	85	−22.50	50-TCA**A**GCA**C**CTGTTGAGCAG**TC**-70	3’	21	4	HE643642	Minus	Leaf bud
pav-miR838	ath-miR838	56	−8.40	33-**C**TTTCTTCTT**T**CTTCTTGCACA-53	3’	21	2	HE645366	Minus	Leaf bud
pav-miR840	ath-miR840	46	−9.50	5-TT**A**TTTAGGTCCCTT**G**GTTTT**T**-25	5’	21	3	HE641543	Plus	Leaf bud
pav-miR844	ath-miR844	63	−8.40	43-TTATAAGCCATCTT**C**CT--CT-61	3’	19	3	HE641759	Minus	Leaf bud
pav-miR845	bdi-miR845	57	−16.00	3-**TT**CT**T**TGATACCAATTGTTGG-23	5’	21	3	HE643435	Plus	Leaf bud
pav-miR854	ath-miR854a	153	−44.20	3-GAT**A**AGGATAGGGAGGAGG**CA**-23	5’	21	3	HE641763	Plus	Leaf bud
pav-miR856	ath-miR856	80	−14.00	1-T**G**ATCCTACCAATAA**T**TTC-**C**C-21	5’	21	4	HE641986	Plus	Leaf bud
pav-miR858	ath-miR858a	83	−30.40	1-TTTC**A**TTGTCTGTTCG**C**CCC**CG**-21	5’	21	4	HE641269	Minus	Leaf bud
pav-miR860	bra-miR860	71	−9.10	9-A**AA**TAG**G**CCAATCTATTGAAG-29	5’	21	3	HE642649	Minus	Leaf bud
pav-miR1088	smo-miR1088	71	−20.70	49-**AG**GAAGAAAGAGAGCA**T**GC-T-68	3’	20	4	HE641483	Plus	Leaf bud
pav-miR1089	smo-miR1089	170	−36.80	150-A**C**ATCA**GT**TAGGATTGTTTGC-170	3’	21	3	HE641278	Plus	Leaf bud
pav-miR1096	smo-miR1096	138	−41.10	9-CTG**T**CTCTTTGCTTC**T**GG**G**AT-29	5’	21	3	HE643882	Plus	Leaf bud
pav-miR1098	smo-miR1098	137	−28.50	1-TG-TGA**TG**TTGTGCTGAAAT-20	5’	20	3	HE645621	Minus	Leaf bud
pav-miR1139	bdi-miR1139	122	−14.00	8-**A**AGTA**T**CATACACTAGT**CA**A**A**-28	5’	21	4	HE640288	Plus	Leaf bud
pav-miR1507	gma-miR1507	100	−31.00	6-G**TT**GTGT**G**TG**A**GATGAGAGAA-26	5’	21	4	HE642607	Minus	Leaf bud
pav-miR1512	gma-miR1512b	105	−16.50	11-T**TT**C**A**GGAAATTCTTAAA**T**CAT-32	5’	22	4	HE640717	Plus	Leaf bud
pav-miR1523	gma-miR1523a	84	−13.40	58-**G**TG**A**GA**G**AAAGGTGAGCTCA-77	3’	20	3	HE643736	Minus	Leaf bud
pav-miR1535	gma-miR1535a	65	−13.50	1-**AA**TGTTTGTGGTGATG-C**A**-18	5’	18	4	HE643621	Minus	Leaf bud
pav-miR1888	ath-miR1888a	61	−8.10	1-TAAGT**AT**A**T**A**G**TTGTGAAGAA-21	5’	21	4	HE645055	Minus	Leaf bud
pav-miR2079	ppt-miR2079	60	−17.90	1-AGAG**G**TGATGTTGATG-C**TG**A-20	5’	20	4	HE641817	Plus	Leaf bud
pav-miR2082	ppt-miR2082	102	−22.80	75-TGTGTGTTCC**AAG**TCTTCTTT-95	3’	21	3	HE643152	Plus	Leaf bud
pav-miR2084	ppt-miR2084	170	−33.90	151-CCTGCATTGTTGGA-TGTGG**C**-170	3’	20	2	HE643146	Plus	Leaf bud
pav-miR2085	ppt-miR2085	85	−19.30	1-ACA-TCCAACAATGCAGG**TTG**-20	5’	20	4	HE643146	Minus	Leaf bud
pav-miR2111	ath-miR2111a	124	−19.08	6-**T**TCCTC**T**G**T**ATGCGGATTACC-26	5’	21	3	HE644450	Minus	Leaf bud
pav-miR2118	bdi-miR2118a	122	−52.40	88-**C**TTCC**CAAA**CCTCCCATTCCTA-109	3’	22	4	HE644421	Plus	Leaf bud
pav-miR2275	bdi-miR2275b	156	−35.90	1-**AT**A**A**GTTTCTTCTA**G**TAT**A**TCA-22	5’	22	4	HE645748	Plus	Leaf bud
pav-miR2933	ath-miR2933a	221	−52.10	4-GAAAT**GG**GAGAGGAAATT**TG**T**G**-25	5’	22	4	HE642097	Plus	Leaf bud
pav-miR2938	ath-miR2938	155	−32.10	134- GATCTTTTGAG**GG**GA**T**TC-A**C**-153	3’	21	4	HE642247	Plus	Leaf bud
pav-miR3440	han-miR3440	80	−19.90	6-TGGGTTGGT**G**AA**C**GGAAA-GC-25	5’	20	3	HE641126	Plus	Leaf bud
pav-miR3627	ppe-miR3627	110	−54.40	19-**A**CGCAGGAGAGATGGC**G**CTGTC-40	5’	22	2	HE640724	Plus	Leaf bud
pav-miR3704	pab-miR3704	79	−25.00	57-**T**GT**G**TAGGTGGAGTTGGAA**TA**C-78	3’	22	4	HE641903	Minus	Leaf bud
pav-miR3709	pab-miR3709	69	−13.20	7-TT**GG**GAT**T**CTTTAAATTCCC**T**-27	5’	21	4	HE644574	Minus	Leaf bud
pav-miR3710	pab-miR3710	167	−54.80	1-T**CT**GA**G**CCTGACGGGCCTCC**G**-21	5’	21	4	HE641240	Minus	Leaf bud
pav-miR3932	ath-miR3932b	82	−16.90	7-TTTGACGTGCTC**GG**T**T**CTGCT**G**-28	5’	22	3	HE645506	Plus	Leaf bud
pav-miR5019	ath-miR5019	105	−23.78	2-T**T**TTG**A**GAAAGAAAAACTC**CA**-22	5’	21	4	HE645902	Minus	Leaf bud
pav-miR5057	bdi-miR5057	115	−27.20	69-AA**T**TTT**A**AAATC**C**TTTTGAC**G**-89	3’	21	4	HE643591	Minus	Leaf bud
pav-miR5069	bdi-miR5069	83	−17.70	11-T**G**GGTT**T**TTGATTTGACC**CA**T-31	5’	21	4	HE642209	Plus	Leaf bud
pav-miR5185	bdi-miR5185a	106	−26.60	16-T**C**TGAGAATTGAAC**A**AGAAGC-36	5’	21	2	HE645377	Minus	Leaf bud
pav-miR5643	ath-miR5643a	111	−20.50	4-**TAT**CTTTTAAGATCTGGTTG**A**-24	5’	21	4	HE643378	Minus	Leaf bud
pav-miR5656	ath-miR5656	56	−6.00	1-A**TGC**AAGTAGAGATTG**T**GTTT-21	5’	21	4	HE645468	Plus	Leaf bud
pav-miR6029	bna-miR6029	130	−29.95	3-TGGGG**G**TGTGATTTCAGGC**AA**-23	5’	21	3	HE641640	Minus	Leaf bud
pav-miR6034	bna-miR6034	60	−6.20	9-TCT**C**AT**T**TATATAGCTTTG**TT**-29	5’	21	4	HE644864	Minus	Leaf bud
pav-miR6114	cca-miR6114	105	−21.50	1-**A**CACG**AGCC**TTGTCCCTTTCA-21	5’	21	4	HE644354	Minus	Leaf bud

Table 2. (Continued).

pav-miR6135	pgi-miR6135	140	–41.20	84-CC**TCC**AGTTGGTCAATTGGC-103	3’	20	4	HE642644	Minus	Leaf bud
pav-miR6260	ppe-miR6260	142	–37.39	1-TG**T**AGTGAGAGAATGGGA**ATG**-21	5’	21	4	HE643440	Minus	Leaf bud
pav-miR6290	ppe-miR6290	105	–21.90	2-TGAATGA**A**TACAGAGATCGTGT**T**-24	5’	23	2	HE645009	Plus	Leaf bud
pav-miR7122	ppe-miR7122	164	–43.20	11- **A**CCGTG**A**TTTCTTTGTATAAA-31	5’	21	2	HE643009	Plus	Leaf bud
pav-miR8122	ppe-miR8122	70	–8.30	11-T-AAGGAAGATTT**C**TG**A**AAA**T**-30	5’	20	4	HE644221	Plus	Leaf bud
pav-miR8131	ppe-miR8131	89	–24.70	1-AT**A**TCAG**AA**AA**T**TTGAGTTGT-21	5’	21	4	HE642772	Minus	Leaf bud
pav-miR8135	cpa-miR8135	80	–13.45	4-**T**GGATTTTGCAGGGTT**C**AT-22	5’	19	2	HE643762	Plus	Leaf bud
pav-miR8148	cpa-miR8148	81	–28.90	54-TGACTGGGTCTGC-GA**G**GTGGC**CA**-76	3’	23	4	BI203082	Plus	Leaf bud
pav-miR8154	cpa-miR8154	69	–11.67	1-**T**AGAGGAGGAGAAGAAG**C**GG**CT**-22	5’	22	4	HE642126	Plus	Leaf bud
pav-miR9554	bra-miR9554	74	–12.00	46-**CC**ATGATA**AA**TGGATATAATC-66	3’	21	4	HE642024	Minus	Leaf bud
pav-miR9559	bra-miR9559	131	–40.23	13-TT**GT**GATTTTGGTCATT**TC**TG-33	5’	21	4	HE646197	Plus	Leaf bud
pav-miR9563	bra-miR9563b	121	–21.28	2- A**GA**G**T**AAGAGATGAATT**GA**TTA-22	5’	21	4	HE641358	Plus	Leaf bud

Table 3. Reference miRNAs from different plant families to which cherry miRNAs showed similarity/homology.

Family	Plant	Common name	Plant lineage group	No. of miRNAs
Rosaceae	*Prunus persica* (ppe)	Peach	Angiosperm (dicot)	15
	Prunus armeniaca (par)	Apricot	Angiosperm (dicot)	1
	Prunus salicina (psl)	Plum	Angiosperm (dicot)	1
	Malus domestica (mdm)	Apple	Angiosperm (dicot)	1
Brassicaceae	*Brassica napus* (bna)	Turnip	Angiosperm (dicot)	4
	Brassica rapa (bra)	Rapeseed	Angiosperm (dicot)	4
	Arabidopsis thaliana (ath)	Thale cress, mouse-ear cress	Angiosperm (dicot)	22
	Arabidopsis lyrata (aly)	-	Angiosperm (dicot)	3
Poaceae	*Oryza sativa* (osa)	Rice	Angiosperm (monocot)	5
Asteraceae	*Cynara cardunculus (cca)*	Artichoke thistle	Angiosperm (dicot)	3
	Helianthus annuus (han)	Sunflower	Angiosperm (dicot)	1
Araliaceae	*Panax ginseng* (pgi)	Ginseng	Angiosperm (dicot)	1
Caricaceae	*Carica papaya* (cpa)	Papaya	Angiosperm (dicot)	4
Euphorbiaceae	*Ricinus communis* (rco)	Castor oil plant	Angiosperm (dicot)	1
Poaceae	*Brachypodium distachyon* (bdi)	Stiff brome	Angiosperm (monocot)	8
Cucurbitaceae	*Cucumis melo* (cme)	Muskmelon	Angiosperm (dicot)	1
Funariaceae	*Physcomitrella patens* (ppt)	Spreading earthmoss	Moss	5
Selaginellaceae	*Selaginella moellendorffii* (smo)	Selaginella	Lycophyte	4
Fabaceae	*Glycine max* (gma)	Soybean	Angiosperm (dicot)	4
Pinaceae	*Picea abies* (pab)	Norway spruce	Gymnosperm	3

3.3. Sense/antisense miRNAs in cherry

The sense/antisense miRNAs are transcribed from both sense and antisense strands of the same genomic loci. In the current study, two of the new conserved cherry miRNAs, i.e. pav-miR482 (Figures 1a and 1b) and pav-miR535 (Figures 1c and 1d), were found to be transcribed in the opposite direction of the same genomic loci.

3.4. Cluster miRNAs in cherry

Animal genomes contain clusters of miRNAs (Altuvia et al., 2005), but plant miRNA clusters are not been frequently

observed. In this study pav-miR482a was identified as a pre-miRNA cluster (Figure 1e) with two mature miRNA sequences. The miR482 family has been reported as cluster miRNA in many plants, i.e. *Phaseolus vulgaris* (Arenas-Huertero et al., 2009), *Zea mays* (Zhang et al., 2009), *Citrus sinensis* (Xu et al., 2010), *Prunus persica* (Zhu et al., 2012), and *Malus domestica* (Xia et al., 2012).

3.5. Overlapping cluster miRNA in cherry

pav-miR161 has been identified with two overlapping mature miRNA sequences in cherry. The precursor sequence contains two different mature miRNA sequences from positions 1–21 and 9–29 (Figure 1f) with an overlapping sequence 13 nt long. The same overlapping pattern is observed in its reference miRNA, i.e. aly-miR161.

3.6. Conservation study of mature cherry miRNAs

The newly identified cherry miRNA (miR172) was selected for conservation studies. The cherry miRNA (pav-miR172) showed conservation with the mature miRNA sequences of miR172 of *Arabidopsis thaliana* (ath), *Prunus persica* (ppe), and *Oryza sativa* (osa), as shown in Figure 2a.

3.7. Highly conserved pre-miRNA in cherry

Usually mature sequences of miRNAs are conserved in plants, but the pre-miRNAs are not found to be conserved in plants (Bartel, 2004). Here, however, a highly conserved miRNA precursor has been predicted in cherry, i.e. pav-mir535, which showed 100% query coverage and 98% identity with the peach pre-miRNA ppe-mir535a. The result is shown in Figure 2b.

3.8. Phylogenetic study of cherry miRNAs

The phylogenetic analysis of one of the newly identified miRNAs, i.e. pav-miR169, was done with the help of Clustal W by using the neighbor-joining clustering method. It suggested that pav-miR169 is closer to *Gossypium hirsutum* and *Gossypium herbaceum* as compared to others as shown in Figure 3.

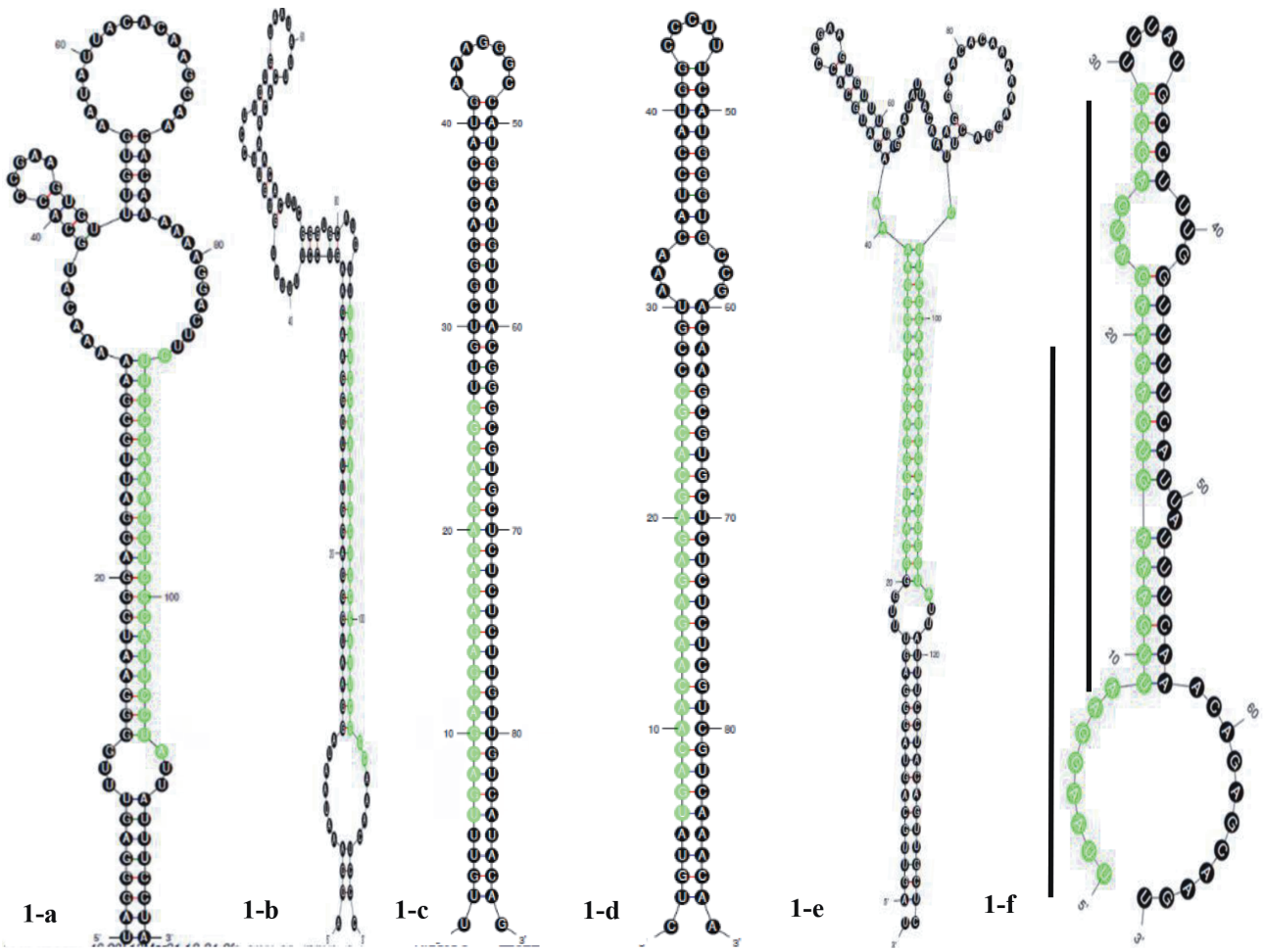


Figure 1. The secondary structures of the newly identified cherry miRNA showing mature sequences highlighted in green: a) pav-miR482 sense, b) pav-miR482 antisense, c) pav-miR535 sense, d) pav-miR535 antisense, e) pav-miR482a pre-miRNA cluster with two mature miRNAs, f) pav-miR161 pre-miRNA cluster showing two mature overlapping miRNAs.

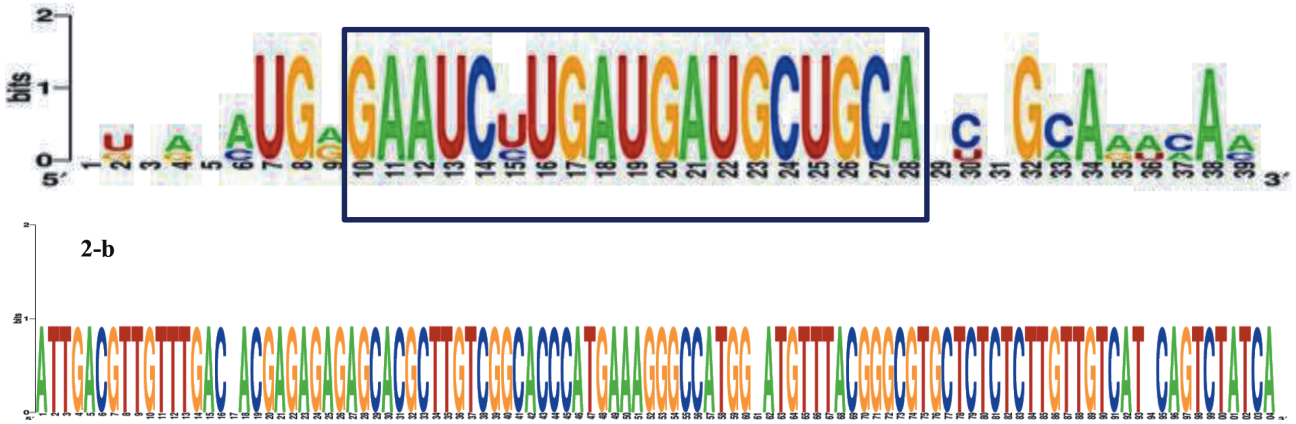


Figure 2. Alignment of the *Prunus avium* pre-miRNAs with their ortholog pre-miRNAs for sequence conservation analyses using WebLogo: a sequence logo generator, showing a) pvi-miR172 mature miRNA conservation with *Arabidopsis thaliana*, *Prunus persica*, and *Oryza sativa* highlighted in a box and b) alignment of the *Prunus avium* pre-miRNA (pav-mir535) with *Prunus persica* pre-miRNA (ppe-mir535) showing almost 100% conservation of pre-miRNAs in both species.

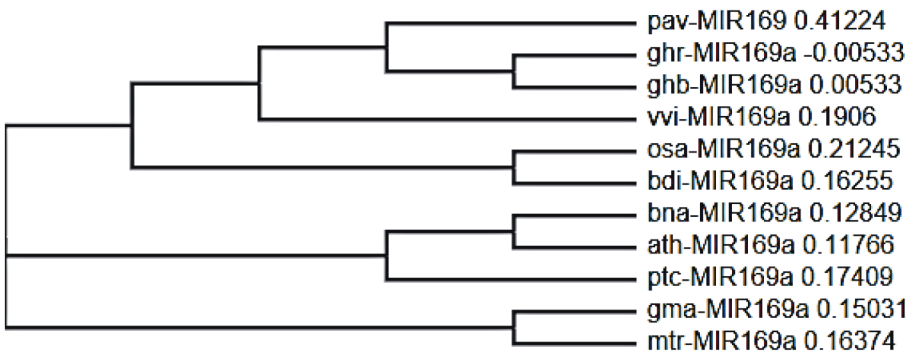


Figure 3. The phylogenetic analysis of the cherry (*Prunus avium*) pre-miRNAs (pav-miR169) with *Arabidopsis thaliana* (ath), *Vitis vinifera* (vvi), *Brassica napus* (bna), *Gossypium herbaceum* (ghb), *Populus trichocarpa* (ptc), *Medicago truncatula* (mtr), *Brachypodium distachyon* (bdi), *Gossypium hirsutum* (ghr), *Glycine max* (gma), and *Oryza sativa* (osa) miRNAs was done with the help of Clustal W and a cladogram tree was generated using the neighbor-joining clustering method. The phylogenetic tree shows that on the basis of pre-miRNA sequences, *Prunus avium* closer to *Gossypium hirsutum* (ghr) and *Gossypium herbaceum* (ghb) as compared to others.

3.9. RT-PCR validation of cherry miRNAs

RT-PCR analysis was conducted for the experimental validation of some of the new conserved cherry miRNAs. Fourteen randomly selected miRNAs, Pav-miR398, 414, 172, 395, 160, 775, 1507, 530, 858, 435, 482, 403, 413, and 156, were employed in RT-PCR validation studies. All of the selected miRNAs confirmed their experimental validation, as shown in Figure 4.

3.10. Cherry miRNA targets

The prediction and annotation of miRNA targets is a critical step to comprehend their regulatory functions. The 91 new conserved cherry miRNAs targeted a total of 211 mRNAs (Table 4). These miRNA targets are various proteins involved in numerous biological processes. Most (43%, 90 out of 211) of the identified cherry miRNAs target metabolism-related proteins, followed by transcription factors (18%, 39 out of 211), signaling (9%, 20 out of 211), biotic/abiotic stress-related proteins (8%, 16 out of 211), hypothetical proteins (8%, 16 out of 211), growth and development (5%, 11 out of 211), transport (4%, 9 out of 211), structural proteins (4%, 8 out of 211), and transposable elements (1%, 2 out of 211). GO-Biological Process (Supplementary Figure 1) revealed that the putative targets of the newly profiled cherry miRNAs are significantly involved in cellular biosynthesis (GO: 0031326), regulation of biosynthetic process (GO: 0009889), regulation of transcription (GO: 0006355), macromolecule biosynthetic process (GO: 0010556),

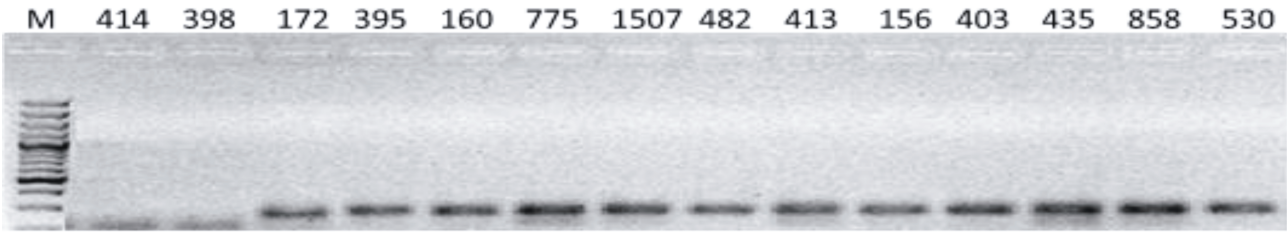


Figure 4. RT-PCR expressional validation for cherry miRNAs. The 14 cherry miRNAs were selected and subjected to RT-PCR expression analysis for the experimental validation. The product of each sample was separated on a 1.8% (w/v) agarose gel.

regulation of nitrogen compound metabolic process (GO: 0051171), regulation of gene expression (GO: 0010468), and anatomical structure development (GO: 0048856). GO-Molecular Function showed that the cherry miRNAs' targets are enriched in nucleic acid binding (GO: 0003676), transcription regulator activity (GO: 0140110), and DNA binding transcription factor activity (GO: 0003700). GO-Cellular Component found that the putative targets of the newly identified cherry miRNAs are engaged in the nucleus (GO: 0005634) and nuclear transcription factor complex (GO: 0044798). Some significant cherry miRNAs and their target networks (Figure 5) showed that Squamosa promoter-binding-like protein 6 (SPL6) is targeted by many miRNAs and one miRNA, such as pav-miR838, can target many genes. The validation of some miRNAs and their targets (Figure 6) was also done with the help of RNAhybrid.

Similar targets for different plant miRNAs were reported by various researchers (Barozai et al., 2008, 2012a, 2012b; Din and Barozai, 2014a, 2014b; Din et al., 2014; Xie et al., 2014; Gul et al., 2017).

4. Discussion

Most miRNA families are conserved evolutionarily through all major lineages of plants (Zhang et al., 2006). The conserved nature of miRNAs facilitates the bioinformatic hunt for new conserved miRNAs in other organisms (Barozai et al., 2008). The conserved mature miRNA sequence and secondary hairpin structure is sufficient to annotate miRNA homologs (Meyers et al., 2008). A majority of the identified miRNAs in this research have been categorized as highly conserved miRNAs by different researchers (Zhang et al., 2006). The newly identified conserved miRNAs in cherry have been predicted by the homology with mosses, pteridophytes, gymnosperms, monocots, and eudicots, which reconfirms their conserved nature across diverse plant groups. The data generated in this research could be helpful for researchers interested in studying the conservation of miRNAs from mosses, pteridophytes, gymnosperms, and monocots to dicot plants, especially in the family Rosaceae.

Many of the newly identified miRNA families in cherry have been profiled for their expression in various developmental processes, growth, and stress responses. The significance of some of the identified miRNAs is discussed below.

miR156 plays a role in flowering time regulation (Spanudakis and Jackson, 2014). It is known to target 11 of the 17 Squamosa promoter binding-like transcription factors that downregulate their expression (Yamaguchi and Abe, 2012). The identification of miR156 in cherry will be helpful in understanding the genetic temporal mechanisms of development in this plant.

The miR172 expression levels influence fruit size in apple (Ripoll et al., 2015; Yao et al., 2015). The newly identified miR172 in cherry would serve as a basis for discernment of this important agronomic trait, i.e. fruit size development of cherry plants.

miR396 plays an important role in salt stress in plants (Kohli et al., 2014; Tian et al., 2014). The newly identified miR396 in cherry will contribute in perceiving the management of this important abiotic stress.

Many miRNAs have been reported to target different transcription factors (Din and Barozai, 2014a, 2014b; Xie et al., 2014). The newly identified cherry miRNAs in this research also appear to target different transcription factors, e.g., pav-miR171 targets the GRAS family transcription factors**.** Similarly, pav-miR414 targets the MYB transcription factor. Such findings can help us comprehend the regulatory hierarchies of gene expression in cherry.

Serine/threonine (Ser/Thr) protein phosphatases are universal enzymes in almost all eukaryotes (Pais et al., 2009). They play important roles in signal transduction (Yu et al., 2005). The newly identified cherry miRNA pav-miR5069 targets this important enzyme. This finding may be helpful in fine-tuning the signal transduction and metabolic processes in cherry.

The bioinformatics methods of miRNA identification mainly rely on mature miRNA sequence conservation, but the identification of highly conserved miRNA cherry

Table 4. Cherry miRNAs targets: the cherry miRNA targets as predicted by psRNATarget. The cherry miRNAs, target GenBank accession numbers, and targeted descriptions are provided here.

Cherry miRNAs	Target accession	Target description
pav-miR156	AT1G27370.1	Squamosa promoter binding protein
	AT2G03750.1	P-loop containing nucleoside triphosphate hydrolases
	AT3G42792.1	Transposable element gene
	AT5G18590.1	Galactose oxidase/kelch repeat superfamily protein
	AT3G28690.1	Protein kinase superfamily protein
	AT5G38610.1	Plant invertase/pectin methylesterase inhibitor superfamily protein
	AT1G22000.1	FBD, F-box and leucine-rich repeat domains containing protein
	AT5G08620.1	STRS2, ATRH25 | DEA(D/H)-box RNA helicase family protein
pav-miR157	AJ827655	Ethylene-responsive transcription factor ERF094
pav-miR160	FC862386	Hypothetical protein
pav-miR161	DW340748	N-acetyltransferase
	DN554529	Protein kinase
	DY647191	Transaldolase
	DW345192	Riboflavin kinase / FMN adenylyltransferase
	DW345877	Haem lyase
	TC12057	Fertility restorer-like protein
	DN552827	Mitochondrial glycine decarboxylase
pav-miR164	TC12261	Hypothetical protein
pav-miR166	TC13481	Hypothetical protein
pav-miR-167	AT1G30330.2	Auxin response factor 6
	AT1G30330.1	Auxin response factor 6
	AT1G50580.1	UDP-Glycosyltransferase
pav-miR169	DY639834	F-box protein
pav-miR171	AT4G00150.1	GRAS family transcription factor
pav-miR172	DY643857	Transcription factor AHAP2
	TC10415	Transcription factor AHAP2
	BU048186	Transcription factor AHAP2
pav-miR391	TC8541	Profilin
	TC8633	Profilin
	TC10146	Profilin
	DT455513	Profilin
	AJ873085	Profilin
pav-miR394	TC9050	Signal recognition particle
pav-miR395	TC14027	Carboxypeptidase type III
pav-miR396	AT3G52910.1	Growth-regulating factor 4
	AT1G30610.1	Pentatricopeptide (PPR) repeat-containing protein
	AT3G53403.1	Zinc finger (C3HC4-type RING finger) family protein
	AT3G24020.1	Disease resistance-responsive (dirigent-like protein) family protein biotic
	AT5G27050.1	AGAMOUS-like 101
	AT1G22950.1	2-Oxoglutarate (2OG) and Fe(II)-dependent oxygenase
	AT5G04500.1	Glycosyltransferase family protein 47

Table 4. (Continued).

pav-miR398	TC9912	Histone H3
pav-miR403	AT1G68910.1	WPP domain-interacting protein 2
	AT1G68910.3	WPP domain-interacting protein 2
	AT5G23120.1	Photosystem II stability/assembly factor, chloroplast (HCF136)
	AT3G22400.1	LOX5 | PLAT/LH2 domain-containing lipoxygenase family protein
	AT2G25760.1	Protein kinase family protein
	AT5G55310.1	DNA topoisomerase 1 beta
	AT1G19520.2	Pentatricopeptide (PPR) repeat-containing protein
	AT1G61610.1	S-locus lectin protein kinase family
	AT1G27752.1	Ubiquitin system component Cue protein
	AT3G54520.1	Hypothetical protein
pav-miR408	AT5G25630.1	Tetratricopeptide repeat (TPR)-like superfamily protein
	AT5G25630.2	Tetratricopeptide repeat (TPR)-like superfamily protein
pav-miR413	DW346647	Unknown protein
pav-miR414	DY652644	Mutant TFIIF-alpha - *Arabidopsis thaliana* (mouse-ear cress), partial (30%)
	TC12811	Cellulose synthase-like protein D4
	AJ826447	Cellulose synthase-like protein D4
	DW342910	MYB transcription factor
	TC13524	Abscisic stress ripening-like protein
pav-miR418	DY645975	Cinnamoyl CoA reductase
pav-miR426	BU045113	ABC transporter related
pav-miR435	AT1G65780.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
	AT3G59600.1	RNA polymerase Rpb8
pav-miR444	TC14130	Similar to UniRef100_Q9LR00
pav-miR477	AT4G22485.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
	AT4G22485.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
	AT4G22485.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
	AT4G22485.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
	AT5G31927.1	Transposable element gene
	AT1G66710.1	AtPP protein
pav-miR482	DY654068	NBS-LRR resistance protein RGH2
	TC8712	CHZMRRNA Zea mays chloroplast rRNA-operon
pav-miR528	TC14323	Os09g0514600 protein
	TC15412	Beta-tubulin
	TC13784	MYB91
	TC10484	Sulfate transporter protein
	TC15195	F-box/kelch-repeat protein At3g63220
	TC11510	Cu/Zn-superoxide dismutase copper chaperone precursor
pav-miR530	TC9415	Eukaryotic translation initiation factor 5
	TC15621	Acyl-ACP thioesterase
pav-miR535	TC375829	RNA recognition motif (RRM)-containing protein
	TC396876	Defensin-like protein
	TC375392	Calcium-binding EF hand family protein
	BP851036	Outer membrane protein; n = 1; *Ralstonia eutropha* H16

Table 4. (Continued).

pav-MIR537	TC13383	Protein phosphatase type 2C
	BU044259	Otubain-like cysteine protease
	TC17842	Otubain-like cysteine protease
pav-miR773	AT2G45720	ARM repeat superfamily protein
	AT3G61690	Nucleotidyltransferases
pav-miR774	DW342137	Envelope glycoprotein precursor
	DW342679	Transcriptional regulator
	TC17074	Auxin-regulated protein
	AJ533666	TRAP C4-dicarboxylate transport system permease
pav-miR775	TC11232	Ubiquitin-conjugating enzyme 2
pav-miR780	AT2G41860	Calcium-dependent protein kinase 14
	AT2G41860	Calcium-dependent protein kinase 14
	AT4G35500	Protein kinase superfamily protein
	AT1G04580	Aldehyde oxidase 4
pav-miR838	TC10023	Core protein
	TC10262	CBL-interacting protein kinase
pav-miR840	DW347291	Hypothetical protein
pav-miR845	TC11802	Cysteine synthase
pav-miR854	TC14008	WD40 repeat protein, COMPASS complex protein
	DW346291	Uncharacterized protein
	DY645970	Pathogenesis-related protein 8
	DT454981	Pathogenesis-related protein 8
	TC10764	Arginine decarboxylase
pav-miR856	AT5G41610.1	Cation/H+ exchanger 18
	AT2G36570.1	Leucine-rich repeat protein kinase family protein
	AT1G08592.1	Potential natural antisense gene, locus overlaps with AT1G08590
	AT2G46800.1	Zinc transporter of *Arabidopsis thaliana*
	AT5G13870.1	Xyloglucan endotransglucosylase/hydrolase 5
	AT4G35460.1	NADPH-dependent thioredoxin reductase B
	AT4G01090.1	Hypothetical Protein
pav-miR858	AT5G59780.3	myb domain protein 59
pav-miR860	AT1G02880.2	Thiamin pyrophosphokinase1
	AT5G26030.2	Ferrochelatase 1
	AT1G74510.2	Galactose oxidase/kelch repeat superfamily protein
	AT2G44581.1	U-box superfamily protein
	AT2G43290.1	Calcium-binding EF-hand family protein
	AT2G25640.1	SPOC domain / Transcription elongation factor S-II protein
	AT4G19020.1	Chromomethylase 2
	AT4G16020.2	Transposable element gene
	AT3G60955.1	A cytochrome P450 pseudogene

Table 4. (Continued).

pav-miR1088	AJ822388	Ferritin
	TC17909	Auxin-induced protein AUX22
	TC8790	Auxin-induced protein 22C
	DY636721	UniRef100_Q10RG3 Cluster: Expressed protein; n = 3
	FC863389	Bax inhibitor
	TC8858	6,7-Dimethyl-8-ribityllumazine synthase
pav-miR1089	TC13608	Betaine-aldehyde dehydrogenase
	TC14713	Peroxisomal acyl-CoA oxidase 1A
pav-miR1096	TC8887	Endo-beta-1,4-glucanase
	DW343035	Ethylene-responsive transcription factor 1A
pav-miR1098	DY636277	Transcription elongation factor 1
pav-miR1139	AT1G30200.1	F-box family protein
	AT1G54850.1	HSP20-like chaperones superfamily protein
pav-miR1507	DW345717	NAD(P)H-quinone oxidoreductase chain 6,
	TC8682	NAD(P)H-quinone oxidoreductase chain 6,
	DY652730	UniRef100_Q9LRY5 Cluster: Gb|AAF16548.1
pav-miR1512	AT4G06485.1	Transposable element gene
	AT3G48500.2	Nucleic acid-binding, OB-fold-like protein
	AT4G20080.1	Calcium-dependent lipid-binding (CaLB domain) plant phosphoribosyltransferase
	AT1G29030.1	Apoptosis inhibitory protein 5
	AT2G31290.1	Ubiquitin carboxyl-terminal hydrolase family protein
pav-miR1523	AT1G16160.1	Wall-associated kinase-like 5
	AT1G60570.1	Galactose oxidase/kelch repeat superfamily protein
	AT2G30870.1	Glutathione S-transferase PHI 10
pav-miR1888	AT5G60850.1	OBF binding protein 4
pav-miR2079	AT2G06850.1	Xyloglucan endotransglucosylase/hydrolase 4
	AT3G02940.1	myb domain protein 107
	AT5G01250.1	Alpha 1,4-glycosyltransferase family protein
	AT2G39710.1	Eukaryotic aspartyl protease family protein |
pav-miR2082	TC13964	UniRef100_Q5M757 Cluster: At4g00840
pav-miR2084	AT4G35360.1	Uncharacterized conserved protein
pav-miR2085	AT1G05520.1	Sec23/Sec24 protein transport family protein
	AT2G23530.1	Zinc-finger domain of monoamine-oxidase A repressor R1
pav-miR2111	AT5G60710.1	Zinc finger (C3HC4-type RING finger) family protein
	AT1G09190.1	Tetratricopeptide repeat (TPR)-like superfamily protein
pav-miR2118	AT1G52440.1	Alpha/beta-hydrolases superfamily protein
	AT3G58980.1	F-box family protein
	AT1G63350.1	Disease resistance protein (CC-NBS-LRR class) family
pav-miR2275	AT1G14360.1	UDP-galactose transporter 3
	AT2G36770.1	UDP-glycosyltransferase superfamily protein
	AT1G62950.1	Leucine-rich repeat transmembrane protein kinase family protein
	AT3G51560.1	Disease resistance protein (TIR-NBS-LRR class) family

Table 4. (Continued).

pav-miR2933	TC15517	Os02g0674700 protein
	TC13771	TIR-NBS-LRR type R protein 7
	DW346969	Ribosomal protein L32
	TC16033	Chlorophyll synthase
pav-miR2938	AT4G08870.1	Arginase/deacetylase superfamily protein
pav-miR3627	TC8541	Profilin
	AJ873085	Profilin
pav-miR3704	AT4G36020.1	Cold shock domain protein 1
	AT3G49700.1	1-Aminocyclopropane-1-carboxylate synthase 9
pav-miR3709	TC13988	UniRef100_Q9SKM7 Cluster
	DN676796	Leucoanthocyanidin dioxygenase
	TC13644	UniRef100_A7QUK4 Cluster
pav-miR3710	AT4G16680.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
pav-miR3932	AT1G02840.2	RNA-binding (RRM/RBD/RNP motifs) family protein
pav-miR5019	TC8455	ATP synthase CF0 subunit III
	DW351203	UniRef100_A5BX39 Cluster: Histone H3
pav-miR5069	TC11428	Serine/threonine protein phosphatase
pav-miR5185	DY646879	Nuclear ribonuclease Z
	DY647054	Delta12-fatty acid desaturase
pav-miR5643	AT2G37360.1	ABC-2 type transporter family protein
	AT4G30890.1	Ubiquitin-specific protease 24
pav-miR6135	AJ822540	1-Aminocyclopropane-1-carboxylate oxidase
pav-miR6260	DY645958	UniRef100_A9T0G4 Cluster
pav-miR6290	DW340879	Disease resistance protein
	DY645998	Ubiquitin carrier protein
	TC12020	Ubiquitin carrier protein
	TC16702	Ubiquitin carrier protein
pav-miR8122	TC17119	AAA-metalloprotease FtsH
	TC8524	Homolog to emb|Y18934.1|CPY18934
	TC15227	UniRef100_Q38719 Cluster: Seed-specific protein of balanced nutritional quality
	DW341211	Proline-rich protein
	DY639675	Glycerol dehydrogenase
pav-miR8131	DY640475	Alpha-glucosidase 2
pav-miR8148	DN556098	Chlorophyll a/b-binding protein AB80
	TC17378	Chlorophyll a/b-binding protein precursor
	TC14618	UniRef100_P09756 Cluster: Chlorophyll a/b-binding protein 3
pav-miR8154	TC10692	Ethylene-responsive transcription factor 1A
	TC12529	Ethylene-responsive transcription factor 1A
	TC11034	Pyruvate dehydrogenase E1 alpha subunit
	TC14553	Cytochrome b5
	TC16903	Glycine dehydrogenase [decarboxylating], mitochondrial precursor
	TC15419	Proline-rich protein
pav-miR9554	FC865222	UniRef100_A9U4N5 Cluster: Predicted protein; n = 1
	DY635974	Abscisic stress ripening-like protein
pav-miR9563	DW345717	NAD(P)H-quinone oxidoreductase chain 6
	TC8682	NAD(P)H-quinone oxidoreductase chain 6

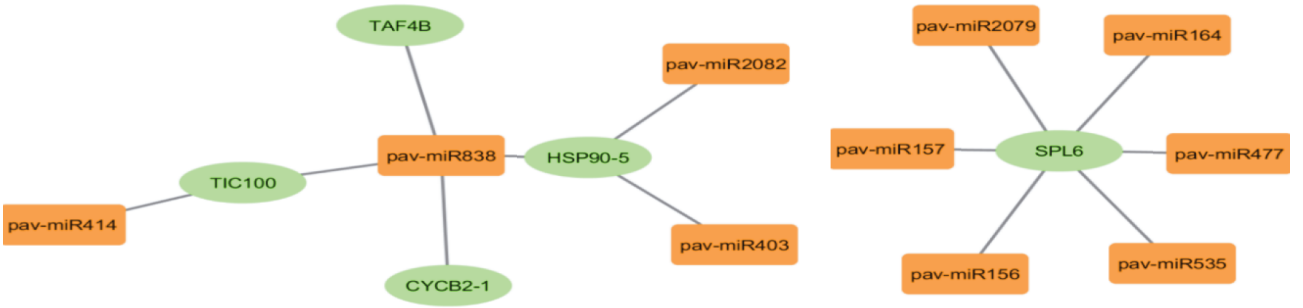


Figure 5. Cherry miRNAs and their target networks.

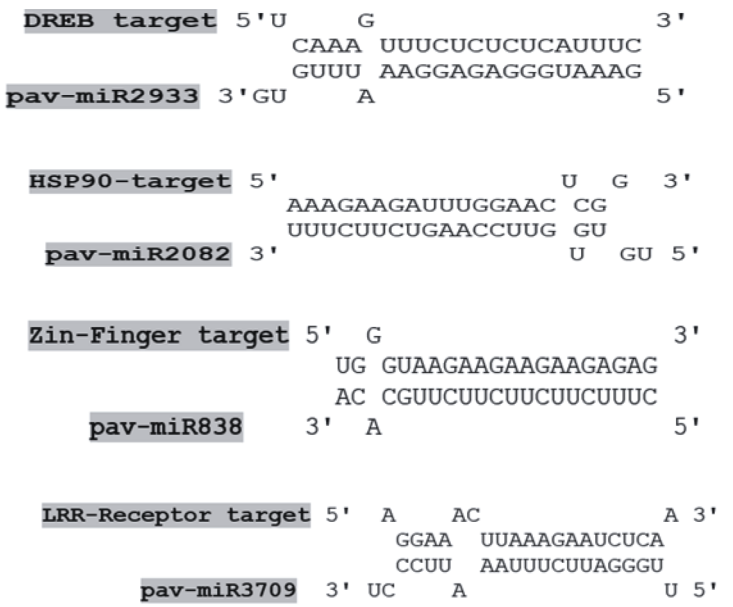


Figure 6. Some selected cherry miRNAs and their target validation by RNAhybrid.

(pav-miR535) is being reported in a plant of the family Rosaceae for the first time and this will open new vistas for the rosaceous research community.

Consequently, we have identified 91 new conserved miRNAs belonging to 88 miRNA families from sweet cherry EST sequences. All the families are being reported for the first time. The EST-based identification is the confirmation of the expression of identified miRNAs. Experimental validation of 14 miRNAs confirms the powerfulness of bioinformatic prediction of miRNAs. These findings will be helpful to comprehend the miRNA-mediated life processes in cherry.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215: 403-410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H (2005). Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 33: 2697-2706.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones SAM, Marshall M et al. (2003). A uniform system for microRNA annotation. RNA 9: 277-279.

Arenas-Huertero C, Pérez B, Rabanal F, Blanco-Melo D, De la Rosa C, Estrada-Navarrete G, Sanchez F, Covarrubias AA, Reyes JL (2009). Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. Plant Mol Biol 70: 385-401.

Baloch IA, Barozai MYK, Din M, Achakzai AKK (2015). Computational identification of 18 miRNAs and their targets in three species of rose. Pak J Bot 47: 1281-1285.

Barozai MYK (2012a). In silico identification of microRNAs and their targets in fiber and oil producing plant flax (*Linum usitatissimum* L.). Pak J Bot 44: 1357-1362.

Barozai MYK (2012b). Identification and characterization of the microRNAs and their targets in *Salmo salar*. Gene 499: 163-168.

Barozai MYK (2013). Identification of microRNAs and their targets in *Artemisia annua* L. Pak J Bot 45: 461-465.

Barozai MYK, Din M, Baloch IA (2013). Structural and functional based identification of the bean (*Phaseolus*) microRNAs and their targets from expressed sequence tags. J Struct Funct Genom 14: 11-18.

Barozai MYK, Irfan M, Yousaf R, Ali I, Qaisar U, Maqbool A, Zahoor M, Rashid B, Hussnain T, Riazuddin S (2008). Identification of micro-RNAs in cotton. Plant Physiol Bioch 46: 739-751.

Barozai MYK, Qasim M, Din M, Achakzai AKK (2018b). An update on the microRNAs and their targets in unicellular red alga *Porphyridium cruentum*. Pak J Bot 50: 817-825.

Barozai MYK, Ye Z, Sangireddy SR, Zhou S (2018a). Bioinformatics profiling and expressional studies of microRNAs in root, stem and leaf of the bioenergy plant switchgrass (*Panicum virgatum* L.) under drought stress. Agri Gene 8: 1-8.

Bartel DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.

Bibi F, Barozai MYK, Din M (2017). Bioinformatics profiling and characterization of potential microRNAs and their targets in the genus *Coffea*. Turk J Agric For 41: 191-200.

Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-García J, Dirlewanger E, Barreneche T (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. BMC Plant Biol 16: 1.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). WebLogo: a sequence logo generator. Genome Res 14: 1188-1190.

Dai X, Zhao PX (2011). psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39: W155-W159.

Din M, Barozai, MYK (2014a). Profiling microRNAs and their targets in an important fleshy fruit: tomato (*Solanum lycopersicum*). Gene 535: 198-203.

Din M, Barozai MYK (2014b). Profiling and characterization of eggplant (*Solanum melongena* L.) microRNAs and their targets. Mol Biol Rep 41: 889-894.

Din M, Barozai MYK, Baloch IA (2014). Identification and functional analysis of new conserved microRNAs and their targets in potato (*Solanum tuberosum* L.). Turk J Bot 38: 1199-1213.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48.

Griffiths-Jones S (2004). The microRNA registry. Nucleic Acids Res 32: D109-D111.

Gul Z, Barozai MYK, Din M (2017). In-silico based identification and functional analyses of miRNAs and their targets in cowpea (*Vigna unguiculata* L.). Aims Genetics 4: 138-165.

Jahan S, Barozai MYK, Din M, Achakzai H, Sajjad A (2017). Expressional studies of microRNAs in hepatitis B patients of Quetta, Pakistan. Pure Appl Biol 6: 1044-1052.

Kohli D, Joshi G, Deokar AA, Bhardwaj AR, Agarwal M, Katiyar-Agarwal S, Srinivasan R, Jain PK (2014). Identification and characterization of wilt and salt stress-responsive microRNAs in chickpea through high-throughput sequencing. PLoS One 9: 108851.

Kruger J, Rehmsmeier M (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res 34 (Suppl. 2): W451-W454.

Ma C, Lu Y, Bai S, Zhang W, Duan X, Meng D, Wang Z, Wang A, Zhou Z, Li T (2014). Cloning and characterization of miRNAs and their targets, including a novel miRNA-targeted NBS–LRR protein class gene in apple (Golden Delicious). Mol Plant 7: 218-230.

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ et al. (2008). Criteria for annotation of plant microRNAs. Plant Cell 20: 3186-3190.

Pais SM, Téllez-Iñón MT, Capiati DA (2009). Serine/threonine protein phosphatases type 2A and their roles in stress signaling. Plant Signal Behav 4: 1013-1015.

Ripoll JJ, Bailey LJ, Mai QA, Wu SL, Hon CT, Chapman EJ, Ditta GS, Estelle M, Yanofsky MF (2015). MicroRNA regulation of fruit growth. Nature Plants 1: 4.

Shah SQ, Din M, Barozai MYK, Achakzai AKK (2017). Codon usage bias analysis for abiotic stress genes in *Arabidopsis thaliana* and *Oryza sativa*. Pak J Bot 49: 15-23.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504.

Spanudakis E, Jackson S (2014). The role of microRNAs in the control of flowering time. J Exp Bot 65: 365-380.

Tian Y, Tian Y, Luo X, Zhou T, Huang Z, Liu Y, Qiu Y, Hou B, Sun D, Deng H et al. (2014). Identification and characterization of microRNAs related to salt stress in broccoli, using high-throughput sequencing and bioinformatics analysis. BMC Plant Biol 14: 1.

Wahid HA, Barozai MYK, Din M (2015). Optimization of total RNA extraction protocol from Ziarat *Juniper* (*Juniperus excelsa* M.Bieb). Pure Appl Biol 4: 275-279.

Wahid HA, Barozai MYK, Din M (2016). Functional characterization of fifteen hundred transcripts from Ziarat *Juniper* (*Juniperus excelsa* M.Bieb). Adv Life Sci 4: 20-26.

Xia R, Zhu H, An YQ, Beers EP, Liu Z (2012). Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol 13: R47.

Xie F, Stewart CN, Taki FA, He Q, Liu H, Zhang B (2014). High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. Plant Biotechnol J 12: 354-366.

Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X (2010). Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. BMC Genomics 11: 1.

Yamaguchi A, Abe M (2012). Regulation of reproductive development by non-coding RNA in *Arabidopsis*: to flower or not to flower. J Plant Res 125: 693-704.

Yao JL, Xu J, Cornille A, Tomes S, Karunairetnam S, Luo Z, Bassett H, Whitworth C, Rees-George J, Ranatunga C et al. (2015). A microRNA allele that emerged prior to apple domestication may underlie fruit size evolution. Plant J 84: 417-427.

Yu RMK, Wong MML, Jack RW, Kong RYC (2005). Structure, evolution and expression of a second subfamily of protein *phosphatase 2A* catalytic subunit genes in the rice plant (*Oryza sativa* L.). Planta 222: 757-768.

Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006). Conservation and divergence of plant microRNA genes. Plant J 46: 243-259.

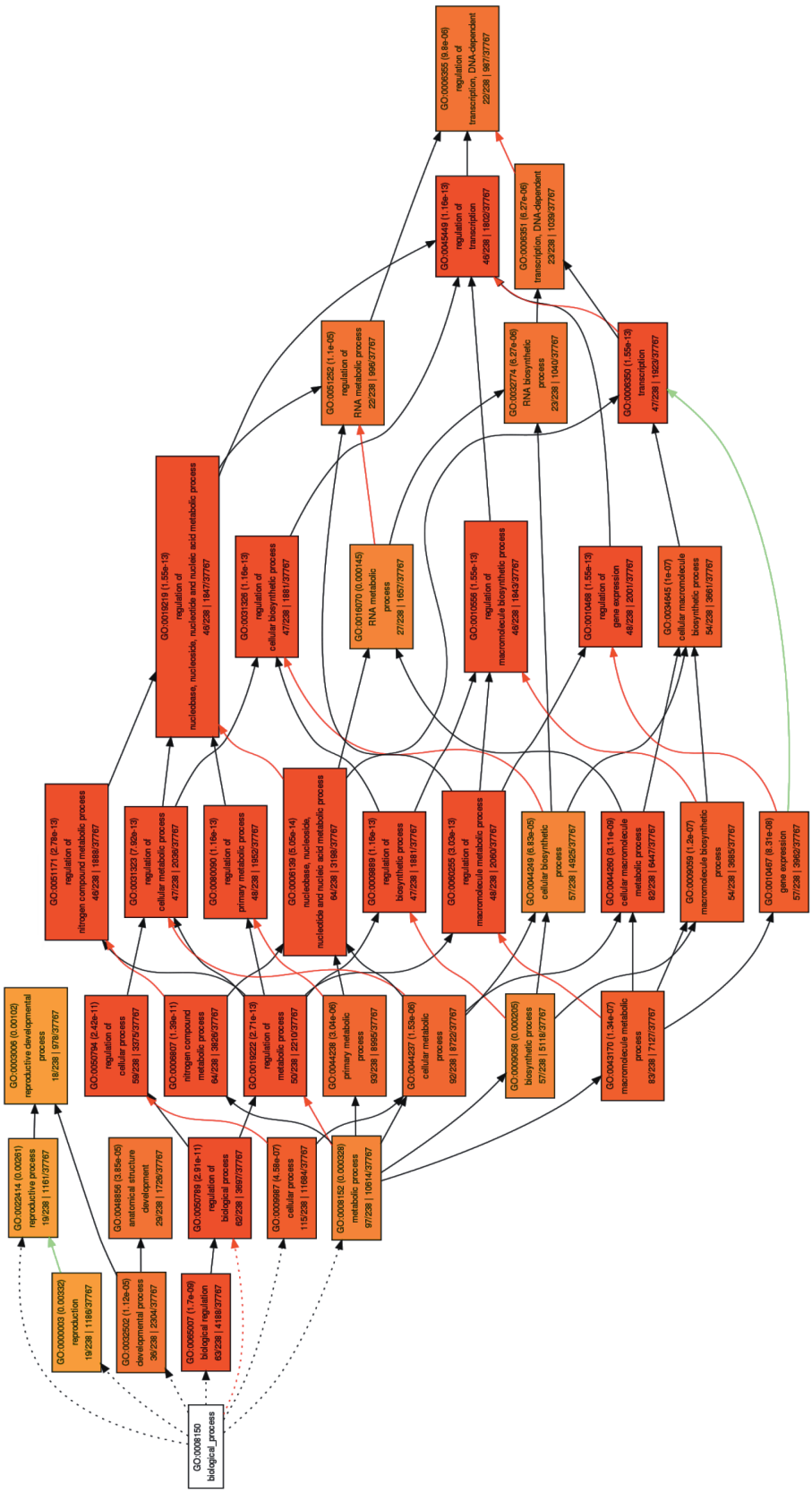
Zhang L, Chia JM, Kumari S, Stein JC, Liu Z, Narechania A, Ware D (2009). A genome-wide characterization of microRNA genes in maize. PLoS Genet 5: 1000716.

Zhang Y, Bai Y, Han J, Chen M, Kayesh E, Jiang W, Fang J (2013). Bioinformatics prediction of miRNAs in the *Prunus persica* genome with validation of their precise sequences by miR-RACE. J Plant Physiol 170: 80-92.

Zhang Z, Yu J, Li D, Zhang Z, Liu F, Zhou X, Wang T, Ling Y, Su Z (2010). PMRD: Plant MicroRNA Database. Nucleic Acids Res 38: D806-D813.

Zhu H, Xia R, Zhao B, An YQ, Dardick CD, Callahan AM, Liu Z (2012). Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol 12: 1.

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



Supplementary Figure 1. Based on GOrilla, more complicated enriched biological processes were built and cherry miRNAs are involved in many significant biological processes, with different colors representing P-values.