Pak. J. Bot., 45(2): 461-465, 2013.

IDENTIFICATION OF MICRORNAS AND THEIR TARGETS IN *ARTEMISIA ANNUA* L.

MUHAMMAD YOUNAS KHAN BAROZAI

Department of Botany, University of Baluchistan, Sariab Road Quetta, Pakistan
Corresponding author's e-mail: barozaikhan@gmail.com

Abstract

MicroRNAs (miRNAs), a class of regulatory RNAs, are tiny, non-protein coding and about 18 to 26 nucleotides long. They are involved in the posttranscriptional gene regulation. Their conserved nature among various organisms makes them a worthy foundation for the prediction of novel miRNAs by computational comparative genomics approach. This study resulted in 16 miRNAs belonging to 13 families; miR 156, 159, 160, 162, 166, 171, 172, 390, 395, 397, 535, 1310 and 4221 for the first time in *Artemisia annua*. All 16 miRNA precursors are observed with stable secondary structures and the mature miRNAs found in stem region of the stem loop structures. Their targets consist of transcription factors like; APETALA2, WRKY3, DELLA, MYB and hypothetical proteins.

Introduction

Artemisia L., of the family *Asteraceae*, found in the temperate regions of the northern hemisphere, in arid and semiarid climates areas.The genus has a great economic importance as medicinal resource, flavouring agent, antibacterial, antifedant and antimalarial activities (Ihsan-ul-haq *et al.,* 2012). MicroRNAs (miRNAs) are small about 18-26 nucleotides long, and non-protein coding (Mica *et al.,* 2006). They are conserved in various plant and animal species (Wang *et al.,* 2012; Barozai, 2012a). They are the gene regulation player to control the expression of messenger RNAs (mRNAs) (Carrington & Ambros, 2003). They generate from a folded stem-loop structures called as Precursor miRNAs (pre-miRNAs). A short double-stranded RNA (dsRNA) is created by detaching the loop of pre-miRNAs. The mature miRNA is one of the single strand of the dsRNA that later integrate into the RNA induced silencing complex (RISC) (Bai *et al.,* 2012). The RISC complex containing miRNA has the ability to negatively regulate the mRNA expression either by inhibiting translation process or by causing its destruction. This depends on the stringency of the miRNA complementarity to its mRNA target (Tang *et al.,* 2003).

In plant and animal species the miRNAs regulate various life processes like; growth & development, organ development, inactivation of the inserted gene, signaling pathway, environmental stresses and resist the attacking viruses (Aukerman & Sakai, 2003; Chen *et al.,* 2012).

Among majority of animals and plants, miRNAs showed conserved nature (Barozai *et al.,* 2011a; 2011b; Barozai, 2012b). This conservation of the miRNAs brings a chance for the prediction and identification of new homolog miRNAs in other species.

As according to the microRNA Registry Database (Version Rfam 16.0 released Sept. 2010) (Griffiths-Jones, 2004), there is no miRNA repository for the *Artemisia annua*. The current research is aimed to find potential miRNAs in *Artemisia annua*. This effort is resulted 16 novel miRNAs for the first time in *Artemisia annua*. These miRNAs belong to 13 miRNA families. Further their targets were also identified.

Materials and Methods

Identification of *Artemisia annua* candidate sequences: The Barozai *et al.,* (2008) methodology after some modification is used to predict the *Artemisia annua* potential miRNAs. The candidate sequences containing Pre-miRNAs of *A. annua* were identified, using known plant pre-miRNAs from the microRNA Registry Database (Version Rfam 16.0 released Sept. 2010) (Griffiths-Jones, 2004), with the help of Blastn algorithm (Altschul *et al.,* 1990). The *A. annua* Expressed Sequence Tags (ESTs) form the EST database publically available at http://blast.ncbi.nlm.nih.gov/Blast.cgi. were mined to find the candidate sequences. The candidate sequences in FASTA formats having a range of 0-4 mature miRNAs mismatches were saved. These sequences were further subjected to develop a single tone EST.

To validate as non-protein coding the initial candidate *A. annua* miRNA sequences were blast for protein homology search in the protein database at National Center for Biotechnology Information (NCBI) through Blastx program (Stephen *et al.,* 1997).

Creation of stem-loop structures for the *Artemisia annua* candidate sequences: The RNA folding algorithm, MFOLD (version 3.2) (Zuker, 2003), was used to predict the stem-loop structures for the candidate's sequences. The same parameters as used earlier (Barozai. 2012c) were applied. The stem portion of the secondary structures were checked for the mature sequences with at least 10 base pairs involved in Watson-Crick or G/U base pairing between the mature miRNA and the opposite strand (miRNA*).

Conservation and phylogenetic analysis: The *A. annua* miRNA (miR-156) conservation and Phylogenetic analysis with *Brassica napus*, *Arabidopsis thaliana* and *Populus trichocarpa* orthologues was carried out by applying weblogo program: a sequence logo generator (Crooks *et al.,* 2004) and ClustalW to initiate cladogram tree using neighbor joining clustering method (Larkin *et al.,* 2007) respectively.

Prediction of *A. annua* miRNA targets: The *A. annua* miRNA targets were predicted using the NCBI Blastn program (Altschul, 1990) and RNA-hybrid, a miRNA target

prediction tool (Kruger & Rehmsmeier, 2006) in similar way as reported for other organism (Barozai, 2012d). Briefly, the mature miRNA sequences as queries were subjected to Blast program and the sequences showing 75% query coverage were selected and subjected to RNA-hybrid for the confirmation. The results were saved.

Results and Discussion

The novel identified *A. annua* miRNAs: Total 16 novel miRNAs in *A. annua* were identified from the available ESTs applying the computational genomics approach. These miRNAs belong to 13 miRNA families; mir 156, 159, 160, 162, 166, 171, 172, 390, 395, 397, 535, 1310 and 4221. The computational comparative genomics approaches are proven techniques for the new interesting findings in plants (Barozai & Husnain, 2011; Barozai & Wahid, 2012; Barozai *et al.*, 2012). Maximum three miRNAs are found from miR 156 family followed by 2 in miR 390. All the novel *A. annua* miRNAs annotated as a valid candidate after satisfying the empirical formula for biogenesis and expression of the miRNAs, suggested by Ambrose *et al.,* (2003). The novel *A. annua* pre-miRNAs satisfied the three criteria (B, C and D) of the Ambrose *et al.,* (2003). In such type studies the fulfilling of the only criterion D is enough for homologous sequences to validate as new miRNAs in different species.

***A. annua* miRNAs Characterization:** According to MFOLD (Zuker, 2003), the novel identified *A. annua* pre-miRNAs have minimum folding free energies (mfe) ranges from -9.1 to -58.6 Kcal mol^{-1} with an average of about -31.8 Kcal mol^{-1}. The pre-miRNAs length ranges from 51-178 nt with an average of 88 nt. The mature miRNA sequences length ranges from 20-23 nt. Majority (62.5%) of the *A. annua* miRNAs have 21nt length, followed by 20nt (18.8%), 22nt (12.5%) and 19nt (6.2%). Majority (43.8%) of the *A. annua* miRNAs are observed perfectly conserved with 0 mismatches with their homologs, followed by 1 (31%), 2 (12.5%), 3 (6%) and 4 (6%) mismatches. Majority of *A. annua* miRNAs are located on the 5' (62.5%) followed by 3' (37.5%) arms of the pre-miRNAs as illustrated in Fig. 1. The *A. annua* miRNAs characterization such as source miRNAs, pre-miRNAs length (PL), minimum free folding energies (MFE), mature miRNA sequences (MS), number of mismatches (NM), mature sequence length (ML), source ESTs (SE), mature sequence arm (MSA) and GC percentage are summarized in Table 1. All the mature sequences of *A. annua* miRNAs are observed in the stem region of the stem-loop structures, as shown in Fig. 1. The novel *A. annua* miRNA stem-loop structures were observed with at least 11-21 nucleotides engaged in Watson-Crick or G/U base pairings between the mature miRNA and the opposite arms (miRNAs*) in the stem region. These results are in agreement with previously reported works by many researcher groups (Barozi *et al.,* 2008; Barozai *et al.,* 2011c; Wang *et al.,* 2012).

To confirm the new *A. annua* miRNAs as solid candidates of miRNAs, the connection between them and known protein is very important. The *A. annua* miRNAs have showed no homology with known proteins. Similar results were reported for plant and animal miRNAs (Barozai *et al.*, 2008; Barozai, 2012b).

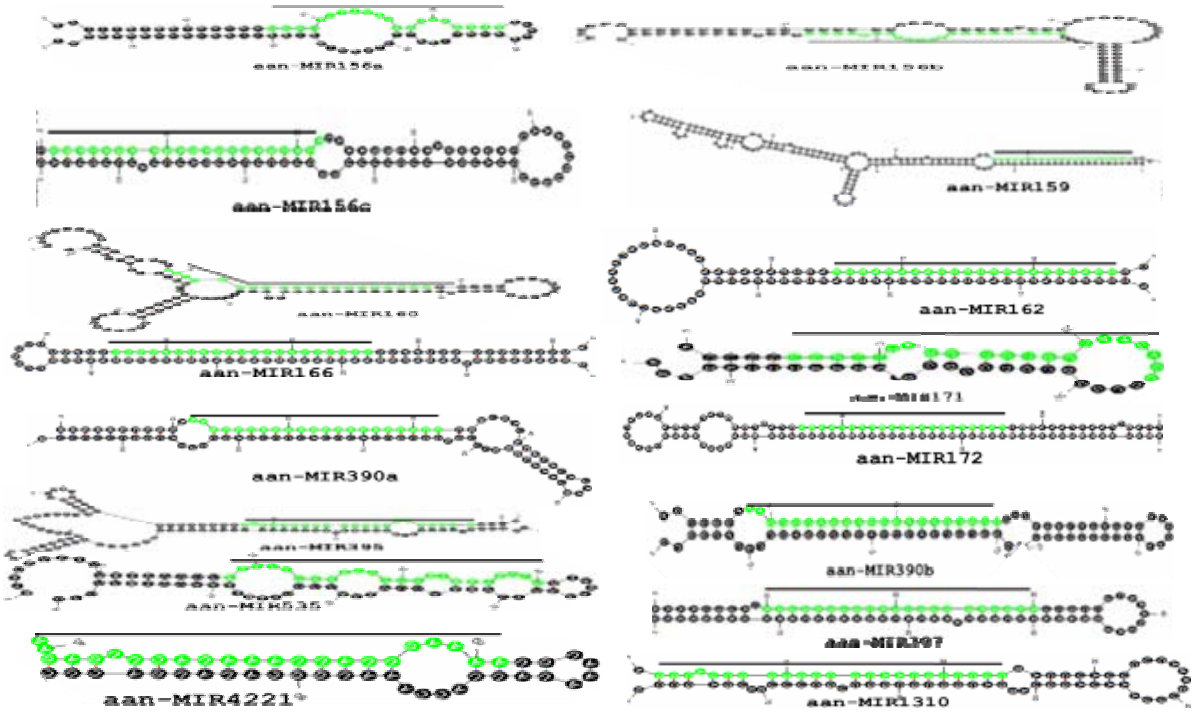

Fig. 1. The novel *A. annua* miRNA secondary structures
The *A. annua* pre-miRNAs secondary structures are predicted using Mfold algorithm. These structures are clearly showing that mature miRNAs in stem region of the stem-loop structures, highlighted with parallel lines and in color.

Table 1. Characterization of the novel identified *A. annua* miRNAs.

A. annua miRNAs	Source miRNAs	PL	MFE	MS	NM	ML	SE	MSA	GC%
aan-MIR156a	aly-MIR156a	60	-12.30	UGACAGAAGAG AG**A**GAGCAC	1	20	EY112709	5’	43.33
aan-MIR156b	aly-MIR156a	69	-20.30	UGACAGAAGAG AG**A**GAGCAC	1	20	EY094777	5’	59.42
aan-MIR156c	bna-MIR156a	86	-37.30	UGACAGAAGAGA GUGAGCACA	0	21	EZ144621	5’	39.53
aan-MIR159	pvu-MIR159a	178	-58.60	U**G**UGGAUUGAAGGG AGC**A**CU**G**	3	21	EZ205405	3’	42.70
aan-MIR160	vvi-MIR160e	51	-32.30	UGCCUGGCUCCC UGUAUGCC-	1	21	EY106936	5’	66.66
aan-MIR162	sly-MIR162	106	-37.50	UCGAUAAACCUC UGCAUCCAG	0	21	EZ141250	3’	42.45
aan-MIR166	ptc-MIR166q	92	-42.60	U**U**GGACCAGGC UUCAUUCCUU	1	21	EZ182127	3’	47.80
aan-MIR171	ccl-MIR171	54	-17.60	UGAUUGAGCCGCG CCAAUAUC	0	21	EY077207	5’	53.70
aan-MIR172	ath-MIR172a	112	-48.0	AGAAUCUUGAUG AUGCUGCAU	0	21	EZ199288	3’	35.71
aan-MIR390a	ath-MIR390a	96	-43.50	AAGCUCAGGAGGG AUAGCGCC	0	21	EZ304761	5’	40.63
aan-MIR390b	ath-MIR390a	87	-40.40	AAGCUCAGGAGGG AUAGCGCC	0	21	EZ215922	5’	43.68
aan-MIR395	osa-miR395o	121	-26.13	AUGAAG**A**GUUUGG AGGAACUC	1	21	EZ347884	3’	38.84
aan-MIR397	sly-MIR397	79	-37.40	AUUGAGUGCAGCG UUGAUGA	0	20	EZ179528	5’	39.23
aan-MIR535	pab-MIR535	80	-9.10	UGACA**G**AA**A**GAGAG AGAGCACGC	2	22	EY095515	5’	45.00
aan-MIR1310	pta-MIR1310	78	-29.40	GGCAUCGGGGGCG**C** AACGCCC-U	2	22	EZ285930	5’	57.70
aan-MIR4221	aly-MIR4221	59	-16.1	AAGAGUUCA**U**AAA**U**U A**U**UGAA**A**A	4	23	EY071372	3’	23.72

The novel identified *A. annua* miRNAs were characterized in terms of PL= Precursor miRNA length, MFE= Minimum free energy, MS= Mature sequence, NM= Number of mismatches (represented in bold & enlarged font size), ML= Mature sequence length, MSA= Mature sequence arm, SE= Source EST and GC % = GC percentage

Conservation and Phylogenetic analysis of *A. annua* miRNAs: The novel *A. annua* miRNAs were further annotated in terms of conservation and phylogenetic analysis. Conservation was observed for the *A. annua* (aan) (miR-156) with *Arabidopsis thaliana* (ath), *Brassica napus* (bna) and *Populus trichocarpa* (ptc) miRNAs as shown in Fig. 2. Similar findings were given by Barozai *et al.*, in plants (Barozai *et al.*, 2008). The Phylogenetic analysis of the same miRNA (miR-156) sequences have showed that the *A. annua* is more closed to *Brassica napus* (bna) than the *Arabidopsis thaliana* (ath) and *Populus trichocarpa* (ptc) as shown in Fig. 3.

***A. annua* miRNA targets:** The prediction of the newly identified *A. annua* miRNA targets is a crucial step for strong validation of computationally identified miRNAs. Total 14 targets were annotated for the novel identified *A. annua* miRNAs, using a combination of Blast and RNA hybrid algorithms as shown in Table 2. Almost all of these targets are already reported as miRNA targets in other plant and animal species (Barozai *et al.*, 2011c; Barozai. 2012c).

Many targeted proteins of plant miRNAs belong to the transcription factors (Wang *et al.*, 2012; Barozai *et al.*, 2013). Same targets family is predicted for *A. annua* miRNAs. The newly identified *A. annua* miRNAs 156, 160, 162 and 172 target APETALA2, WRKY3, DELLA, MYB-related and WRKY1 transcription factors respectively. Other *A. annua* miRNAs target hypothetical proteins. Similar findings are reported for various plants (Barozai *et al.*, 2008; Chen *et al.*, 2012; Wang *et al.*, 2012).

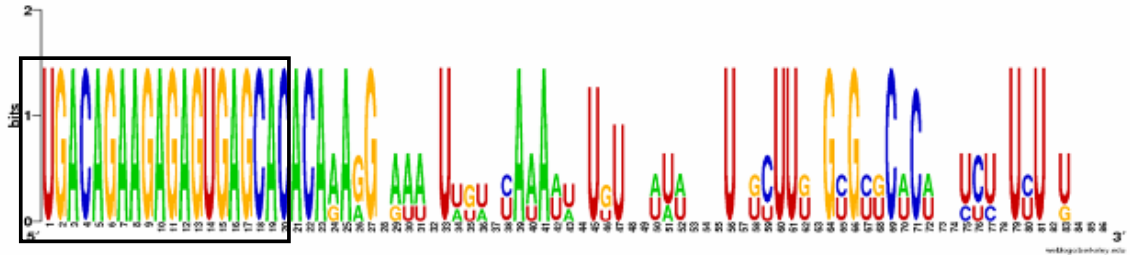

Fig. 2. The *A. annua* miRNA conservation studies.
Alignment of pre-miRNAs (156) of *A. annua* with *Arabidopsis thaliana* (ath), *Brassica napus* (bna) and *Populus trichocarpa* (ptc) miRNAs, using Weblogo: a sequence logo generator, showing miRNA sequences conservation. The mature sequences highlighted in a box.

Table 2. Putative *A. annua* miRNA target. The *A. annua* miRNA families and their putative targets, predicted with the help of Blastn and RNA- hybrid tools are represented. The targeted proteins function, Genbank Acc. And RNA-hybrid results are also provided.

A. annua Micro RNA family	Targets		
	Function	Genbank Acc.#	RNA-Hybrid Result
156	APETALA2 transcription factor	ACY74336	target 5' A U GCUG G AA C 3' GUG CUCAU CU CU UGUCA CAC GAGUG GA GA ACAGU miRNA 3' A A AG 5'
156	WRKY3 transcription factor	GU299481	target 5' A GAGA G AA U 3' GCUUACU U UU CUGUCA CGAGUGA A AA GACAGU miRNA 3' ACA G G 5'
159	Hypothetical protein	EZ320956	target 5' U A UC A A 3' GG GCUCCCUUU UC ACA UC CGAGGGAAG AG UGU miRNA 3' G A UU G 5'
160	DELLA protein	GQ468552	target 5' U GU A U A 3' GGCGU GCAG AGCCG GCA CCGUA UGUC UCGGU CGU miRNA 3' CC C 5'
162	MYB-related transcription factor	GQ468553	target 5' A CCA UGUU GCC C 3' UGG AUGU GGAGG GUC ACC UACG UCUCC UAG miRNA 3' G AAA CU 5'
166	Hypothetical protein	EZ211664	target 5' C GG UC A G 3' G UG GUC GG CCAG C AC CGG CC GGUU miRNA 3' UUC UU UU A A 5'
171	Hypothetical protein	EZ349102	target 5' G A C A 3' AUUG GC CG GC CAAU UAAC CG GC CG GUUA miRNA 3' CUA C A GU 5'
172	WRKY transcription factor 1	FJ390842	target 5' C C G 3' UAGC CGUCAAGG UUCU GUCG GUAGUUCU AAGA miRNA 3' UAC UA 5'
390	Hypothetical protein	EZ283822	target 5' A C G C 3' GC UUGUU UUCUUG GCUU CG GAUAG GAGGAC CGAA miRNA 3' C C G U 5'
395	Hypothetical protein	EZ274318	target 5' A UUUU C U 3' UUCCUC CAAACUC UC AAGGAG GUUUGAG AG miRNA 3' CUC A UA 5'
397	Hypothetical protein	EZ182168	target 5' G AUGAUU C 3' CG CAGCGUUG CUCAG GU GUUGCGAC GAGUU miRNA 3' A A GU A 5'
535	Hypothetical Protein	EZ304086	target 5' U GUACC GGA G C 3' CG GCUCUC CUU UG CA GC CGAGAG GAG AC GU miRNA 3' C A A AAG A 5'
1310	Hypothetical Protein	EZ345678	target 5' G U 3' AGGGCGUUGCGCCCCCGAUGCC UCCCGCAACGCGGGGGCUACGG miRNA 3' 5'
4221	Hypothetical protein	EZ208022	target 5' A G A 3' UUUCAAUAAUUUA GAACU AAAGUUAUUAAAU CUUGA miRNA 3' A A GAA 5'

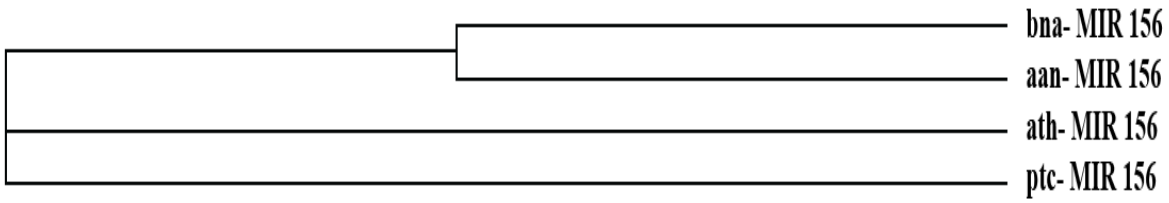

Fig. 3. The *A. annua* miRNA phylogenetic analysis.
The Phylogenetic analysis of the pre-miRNA (156) of *A. annua* (aan) with *Arabidopsis thaliana* (ath), *Brassica napus* (bna) and *Populus trichocarpa* (ptc) miRNAs, was done with the help of ClustalW and cladogram tree was generated using neighbor joining clustering method. The Phylogenetic tree showed that on the basis of pre-miRNA sequences, the *A. annua* (aan) is more closed to *Brassica napus* (bna) than the *Arabidopsis thaliana* (ath) and *Populus trichocarpa* (ptc).

Conclusion

Total 16 miRNAs belonging to 13 families are reported here in *Artemisia* species. These results will be a valuable resource in elucidating the gene regulation mechanism in the *Artemisia*. It also supports the bioinformatics method for new pre-miRNAs documentation from plant species whose genome is not yet sequenced. These miRNAs can be utilized in the medicinal aspect of this plant species.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.

Ambros, V., B. Bartel and D.P. Bartel. 2003. A uniform system for microRNA annotation. *RNA.*, 9: 277-279.

Aukerman, M.J. and H. Sakai. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-Like target genes. *Plant Cell.*, 15: 2730-2741.

Bai, M., G.S. Yang, W.T. Chen, Z.C. Mao, H.X. Kang, G.H. Chen, Y.H. Yang and B.Y. Xie. 2012. Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene*., 501(1): 52-62.

Barozai, M.Y.K. 2012a. The novel 172 sheep (*Ovis aries*) microRNAs and their targets. *Mol. Biol. Rep*., 39(5): 6259-6266.

Barozai, M.Y.K. 2012b. Identification and characterization of the microRNAs and their targets in *Salmo salar. Gene*., 499(1): 163-168.

Barozai, M.Y.K. 2012c. The MicroRNAs and their targets in the channel catfish (*Ictalurus punctatus*). *Mol. Biol. Rep*., 39(9): 8867-8872.

Barozai, M.Y.K. 2012d. Insilico identification of microRNAs and their targets in fiber and oil producing plant Flax (*Linum usitatissimum* L.) *Pak. J. Bot.*, 44(4): 1357-1362.

Barozai, M.Y.K. and H.A. Wahid, 2012. Insilico identification and characterization of cumulative abiotic stress responding Genes in Potato (*Solanum tuberosum* L.). *Pak. J. Bot*., 44(SI): 57-69.

Barozai, M.Y.K. and T. Hussnain. 2012. Identification of biotic and abiotic stress up-regulated ESTs in Gossypium arboreum. *Mol. Biol. Rep*., 39(2): 1011-1018.

Barozai, M.Y.K., A.G. Kakar and M. Din. 2012. The relationship between codon usage bias and salt resistant genes in *Arabidopsis thaliana* and *Oryza sativa*. *Pure Appl. Bio*., 1(2): 48-51.

Barozai, M.Y.K., I.A. Baloch and M. Din. 2011a. Computational Identification of MicroRNAs and their Targets in two Species of Evergreen Spruce Tree (*Picea*). *P. Waset.* 75: 413-418.

Barozai, M.Y.K., I.A. Baloch and M. Din. 2011c. Identification of MicroRNAs and their targets in Helianthus. *Mol. Biol. Rep*., 39(3): 2523-2532.

Barozai, M.Y.K., M. Din and I.A. Baloch. 2011b. Computational identification of microRNAs and their targets in ecological model plant *Mimulus*. *J. Biophy. Chem*., 2(3): 322-331.

Barozai, M.Y.K., M. Irfan, R. Yousaf, I. Ali, U. Qaisar, A. Maqbool, M. Zahoor, B. Rashid, T. Hussnain and S. Riazuddin. 2008. Identification of micro-RNAs in cotton. *Plant Physiol. and Biochem.*, 46(8-9): 739-51.

Barozai, M.Y.K., S. Kakar and A.M. Sarangzai. 2013. Profiling the Carrot (*Daucus carota* L.) microRNAs and their targets. *Pak. J. Bot.*, 45(S1): 353-358.

Carrington, J.C. and V. Ambros. 2003. Role of microRNAs in plant and animal development. *Sci.*, 301: 336-338.

Chen, L., Y. Ren, Y. Zhang, J. Xu, F. Sun, Z. Zhang and Y. Wang. 2012. Genome-wide identification and expression analysis of heat-responsive and novel microRNAs in *Populus tomentosa*. *Gene*., 504(2): 160-165.

Crooks, G.E., G. Hon, J.M. Chandonia and S.E. Brenner. 2004. WebLogo: A sequence logo generator, *Gen. Research.*, 14: 1188-1190.

Griffiths-Jones, S. 2004. The microRNA Registry. *Nucleic Acids Res.*, 32D: 109-111.

Ihsan-ul-haq., A. Mannan, I. Ahmed, I. Hussain, M. Jamil and B. Mirza. 2012. Antibacterial activity and brine shrimp toxicity of *Artemisia dubia* extract. *Pak. J. Bot.*, 44(4): 1487-1490.

Kruger, J. and M. Rehmsmeier. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible *Nucl. Acids. Res.*, 34(suppl_2): W451-454.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. Gettigan, H. William, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins 2007. ClustalW and ClustalX version 2. *Bioinform.*, 23(21): 2947-2948.

Mica, E., L. Gianfranceschi and M.E. Pe. 2006. Characterization of five microRNA families in *maize. J. Exp. Bot.*, 57(11): 2601-2612.

Stephen, F.A., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST, A new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389-3402.

Tang, G., B.J. Reinhart, D.P. Bartel and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes Dev.*, 17: 49-63.

Wang, J., X. Yang, H. Xu, X. Chi, M. Zhang and X. Hou. 2012. Identification and characterization of microRNAs and their target genes in *Brassica oleracea*. *Gene*., 505(2): 300-308.

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

(Received for publication 14 February 2011)