Pak. J. Bot., 44(4): 1357-1362, 2012.

IN SILICO IDENTIFICATION OF MICRORNAS AND THEIR TARGETS IN FIBER AND OIL PRODUCING PLANT FLAX (*LINUM USITATISSIMUM* 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) are a class of small, non-coding RNA. They are about 21 nucleotides in length, produced endogenously and negatively regulate gene regulation. The most miRNAs showed conserved nature and this conservation leads to the novel miRNAs discovery applying insilico comparative genomics. Total 26 novel miRNAs from 19 families are identified in Flax (*Linum usitatissimum* L.). All the miRNA families (miR 156, 159, 162, 166, 167, 171, 172, 319, 394, 397, 408, 783, 1316, 1535, 2108, 3712, 3932, 5014 and 5172) are being reported here for the first time in Flax. All 26 miRNA precursor sequences showed thermodynamically stable structures and the mature miRNAs are appeared in the stem of the stem loop structure. These novel miRNAs in Flax targets transcription factors, retrotransposon, rust resistance genes, cell signaling proteins, maturase K, cellulose synthase, male sterility-related protein and many hypothetical proteins.

Introduction

Flax *(Linum usitatissimum* L.) is an annual, self-pollinating plant belonging to family Linaceae. It is known for its fibers (fiber flax) and seed oil (linseed) (Kurt *et al.,* 2005). Beside this the Flax plant has also medicinal values (Muhammad & Hussain, 2010).

MicroRNAs (miRNAs), is a class of small non-coding RNA, endogenous in nature, about 21 nucleotides long and act as gene regulators. They play central role in post transcriptional gene regulation (Carrington & Ambros, 2003). They are generated from long single strand RNA, folded into a stable hair-pin / stem-loop structures known as Precursor miRNAs (pre-miRNAs). Later the stem and loop of the pre-miRNA detach from each other to develop a small double-stranded RNA (dsRNA). A single strand of the dsRNA acts as mature miRNA and integrates into the RNA induced silencing complex (RISC) (Hammond *et al.,* 2000). The RISC complex containing miRNA negatively regulates genes either by suppressing or degrading its mRNA. This depends on the degree of complimentary of miRNA within its target mRNA (Tang *et al.,* 2003).

The miRNAs are involved almost in each stage of growth, development, stress, diseases, transgene suppression, signalling pathway and defence against the viruses (Aukerman & Sakai, 2003; Fuliang *et al.,* 2010). Another important property of miRNAs is its conservation among the different organisms (Weber, 2005; Barozai *et al.,* 2008; Barozai, 2012a; 2012b: 2012c). This conservation leads to a logical approach for identification of novel homologs using insilico genomics in other species.

The present study is reported 26 novel miRNAs for the first time in Flax (*Linum usitatissimum* L.), fiber and oil producing plant. These miRNAs belong to 19 families (miR 156, 159, 162, 166, 167, 171, 172, 319, 394, 397, 408, 783, 1316, 1535, 2108, 3712, 3932, 5014 and 5172). All 26 miRNA precursors form stable minimum free energy stem loop structures as their orthologues form and the mature miRNAs reside in the stem portion of the stem loop structure.

Materials and Methods

Identification of candidate miRNAs: A similar approach as used earlier (Barozai *et al.,* 2011a) was applied. The sequences containing Pre-miRNA of Flax with a range of 0-4 mismatches in the mature miRNAs were identified and saved as candidate miRNAs. Briefly, the known plant miRNAs from the microRNA Registry Database (Version Rfam 17.0 released April 2011) (Griffiths-Jones, 2004) were downloaded and homology search was performed using Blast algorithem (Altschul *et al.,* 1990) from the publicly available Flax Expressed Sequence Tags (ESTs) database at http://blast.ncbi.nlm.nih.gov/Blast.cgi. The repeated ESTs for the same gene were removed.

Further the protein coding sequences were removed using Blastx with default parameter (Stephen *et al.,* 1997) for protein homology search against the NCBI protein database.

Stem-loop structures generation: The stem-loop structures were generated for the candidate miRNA sequences. For this purpose the RNA folding algorithm, MFOLD (version 3.2) (Zuker, 2003) with the same parameters as described earlier (Barozai *et al.,* 2011b), was applied. The stem-loop structures showing at least 12 base pairs, for the mature sequences, engaged in Watson-Crick or G/U base pairing between the mature miRNA and the opposite strand (miRNA*) were selected for the down-stream analysis.

Conservation and phylogenetic studies: The Flax miRNA (miR-156 & miR-159) conservation and miR-156 Phylogenetic analysis with *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa* orthologues were performed applying the publically available weblogo: a sequence logo generator (Crooks *et al.,* 2004) and ClustalW to generate cladogram tree using neighbor joining clustering method (Larkin *et al.,* 2007) respectively. The results were saved.

Prediction of Flax miRNA targets: The same method as described by Barozai *et al.,* (2011c) was used to predict the miRNA targets. The NCBI Blastn program (Altschul *et al.,* 1990) and RNA-hybrid, a miRNA target prediction tool (Kruger & Rehmsmeier, 2006) were applied to predict the Flax miRNA targets. Briefly the Flax mature miRNA sequences were subjected as queries. The mRNA sequences having 70% query coverage were selected and subjected to RNA-hybrid for the confirmation. The results were saved.

Results and Discussion

The Flax miRNAs: The homology based studies using comparative genetics is a valid approach to find new interesting findings in plants (Barozai & Husnain, 2011; Barozai & Wahid, 2012). In this study a total of 26 novel miRNAs were identified for the first time in Flax by subjecting Flax ESTs to the insilico comparative genomics approach using bioinformatics' tools. The 26 miRNAs were identified in 19 miRNA families (miR 156, 159, 162, 166, 167, 171, 172, 319, 394, 397, 408, 783, 1316, 1535, 2108, 3712, 3932, 5014 and 5172). For families miR-156, 159, 166, 167, 172 and 397 two miRNAs are identified and only one is identified in the remaining families. The miR-408 is identified as pre-miRNA cluster with two miRNAs (Fig. 1). All the novel Flax miRNAs annotated as a valid candidate after satisfying the empirical formula for biogenesis and expression of the miRNAs, suggested by Ambros *et al.*, (2003). The novel Flax pre-miRNAs satisfied the criteria B, C and D. According to Ambros *et al.*, (2003) only the criterion D is enough for homologous sequences to validate as new miRNAs in different species.

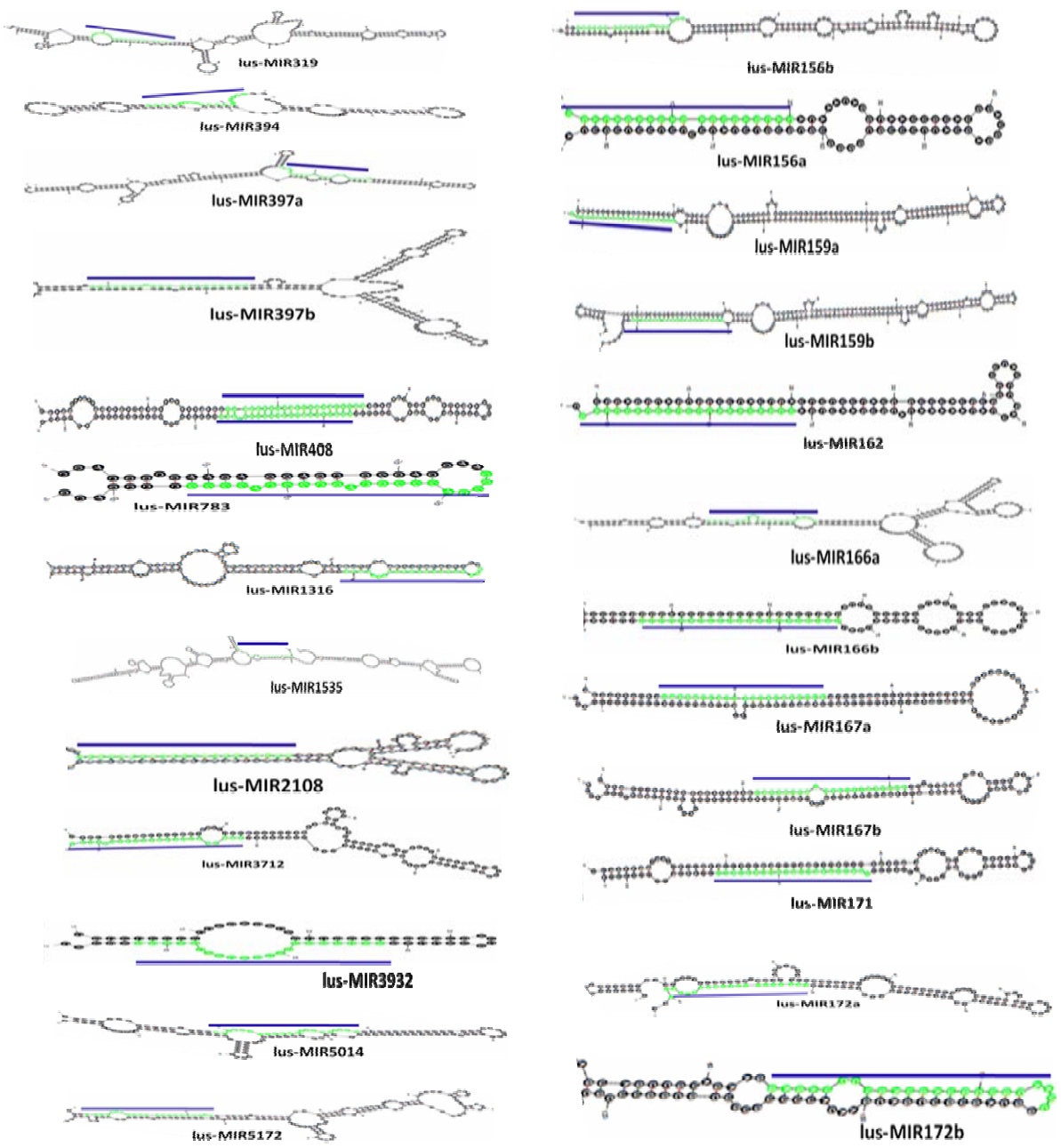

Fig. 1. The novel Flax miRNA secondary structures

The Flax 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 Flax miRNAs.

L. usitatissimum (Flax) miRNAs	Reference miRNAs	PL	MFE	MS	NM	ML	SE	MSA
lus-MIR156a	ath-MIR156a	83	-43	UGACAGAAGAGAGUGAGCACA	0	21	JG192097	5’
lus-MIR156b	ptc-MIR156f	172	-53	UGACAGAAGAGAGAGAGCAC	1	20	GW865886	5’
lus-MIR159a	ath-MIR159a	182	-95	UUUGGAUUGAAGGGAGCUCUU	1	21	JG106941	3’
lus-MIR159b	ath-MIR159a	210	-96	UUUGGAUUGAAGGGAGCUCUC	1	21	JG106559	3’
lus-MIR162	ath-MIR160a	93	-42	UCGAUAAACCUCUGCAUCCAG	0	21	GW866351	5’
lus-MIR166a	ath-MIR162a	210	-43	UCGGACCAGGCUUCAUUCCCC	0	21	JG069755	5’
lus-MIR166b	ath-MIR162a	100	-38	UCGGACCAGGCUUCAUCCCCC	1	21	JG026926	3’
lus-MIR167a	bna-MIR167c	124	-57	UGAAGCUGCCAGCAUGAUCUA	0	21	JG151333	5’
lus-MIR167b	bna-MIR167c	132	-62	UGAAGCUGCCAGCAUGAUCUA	0	21	JG160164	5’
lus-MIR171	ath-MIR171	123	-42	UGAUUGAGCCGUGCCAAUAUC	1	21	JG283626	3’
lus-MIR172a	ath-MIR172a	142	-35	AGAAUCCUGAUGAUGCUGCAU	1	21	JG221301	3’
lus-MIR172b	ptc-MIR172g	156	-40	GGAAUCCUGAUGAUGCUGCAA	2	21	JG222695	5’
lus-MIR319c	gma-MIR319c	260	-52	UUUGAAUGAAAGGAGCUCCA	3	20	JG050775	5’
lus-MIR394	ath-MIR394a	129	-26	UUGGCAUUCUGUCTACCUCC	1	20	JG119724	3’
lus-MIR397a	sly-MIR397	210	-62	GUUGAGUCCAGCGUUGAUGA	2	20	JG063140	5’
lus-MIR397b	sly-MIR397	210	-57	AUUGAGGGCAGCGUUGAUUA	2	20	JG170566	5’
lus-MIR408	aly-MIR408	143	-54	CUGGGAACAGACAGAGCAUGG AUGCACUGCCUCUUCCCUGGC	3 0	21 21	JG083782	5’ 3’
lus-MIR783	pta-MIR783	53	-17	GUGAUUUGCAGGUUCAUUUUC	4	21	JG243376	3’
lus-MIR1316	pta-MIR1316	350	-58	UAAAUACACAAACCAAUAGAA	4	21	EH791487	3’
lus-MIR1535	gma-MIR1535	392	-90	AGUGUCUGUGGUGAUGUCU	3	19	EB712490	3’
lus-MIR2108	gma-MIR2108a	389	-89	UUAAUGUGUAUUGUUUGUCGG	2	21	JG033103	3’
lus-MIR3712	pab-MIR3712	123	-37	GGUGAUCAAGAUCAACGACUACCA	3	24	JG075573	3’
lus-MIR3932	ath-MIR3932b	67	-19	AACUG-GUGAUGACAACGAAG	2	20	JG250466	3’
lus-MIR5014	ath-MIR5014	130	-21	UUGUACAAAUUGAAGUGUACU	2	21	JG102656	5’
lus-MIR5172	bdi-MIR5172	196	-43	GCUUCUACGAGCUCCUCGGCA	4	21	GW867740	5’

The novel identified Flax miRNAs were characterized in terms of PL=Precursor miRNA Length, MFE=Minimum Free Energy, MS=Mature Sequence, NM= Number of Mismatches (represented by highlights), ML=Mature sequence Length, MSA=Mature Sequence Arm and SE=Source EST

Table 2. Putative Flax miRNA targets.

Flax MicroRNA family	Targets	
	Function	**Genbank Acc.**
156	Caffeoyl-CoA 3-O-methyltransferase	EU926495
	Flax resistance gene	AJ310183
	Fis1 protein	X86733
	MYB1-2 mRNA	GQ374578
159	TIR-NBS-LRR protein	AJ310150
	M3 gene rust resistance	GQ141889
162	Koto' resistance-like protein P-B	AF310966
	Abyssinian' resistance-like protein P2-A	AF310960
	Akmolinsk' resistance-like protein P1-A	AF310958
166	Peroxidase gene	AF389350
	Oleosin low molecular weight isoform (G2) gene	DQ207616
	Pinoresinol-lariciresinol reductase	AJ849358.1
	Hoshangabad' resistance-like protein PH-B	AF310968.1
	Leona' resistance-like protein P3-B	AF310962
	Maturase (matR) gene	AY674533
167	Stearoyl-acyl carrier protein	DQ157258
	mRNA for conlinin	AJ414733
	LTR retrotransposon FL5	GU929875
171	Hypothetical protein	EU831071
	Maturase K	GQ845223
172	Hypothetical protein	EU829319
	ATP-binding cassette transporter mRNA	GU581038
319	putative amino acid transporter mRNA	GU581041
	Vacuolar processing enzyme mRNA	GU581045
394	Cellulose synthase-like protein D5	EF490437
	Rust resistance protein (L) gene	AF093646
397	Hypothetical protein	EU831060
	M1 gene	GQ141888
408-5p	Zeta-carotene desaturase mRNA	FJ169886
	Putative GDP-mannose pyrophosphorylase	DQ487209
408-3p	Putative protein kinase atsik mRNA	GU581026
	Zinc finger protein	GU581017
783	Rust resistance protein M gene	U73916
	Hypothetical protein	EU828958
1316	Nbi-D gene TIR-NBS-LRR	AJ310163
	Ngc-C gene	AJ310151
1535	Male sterility-related protein	EU365361
	Site-specific insertion sequence	AJ131991
2108	Hypothetical protein	EU831070
	LTR retrotransposon FL6	GU929876
3712	Hypothetical protein	EU828820
	Retrotransposon FL1a	GU735098
5014	M1 rust resistance gene	GQ141890
	Hypothetical protein	EU830844
3932	S-adenosyl-L-methionine-dependent coniferyl alcohol 9-O-methyltransferase	EU272821
	Actin	AY857865
5172	MYB1-1 mRNA	GQ374577
	Omega-3 fatty acid desaturase	AB457843

The Flax miRNA families and their putative targets, predicted with the help of Blastn and RNA-hybrid tools are represented. The targeted proteins function and Genbank Acc. are provided here

The characterization of Flax miRNAs: The novel identified Flax miRNAs were characterized in terms of minimum free folding energy (mfe), mature miRNAs length, mature sequence arm and mismatches with their source orthologues miRNAs. The minimum folding free energies (mfe) for the Flax pre-miRNAs ranges from -17 to -96 Kcal mol^{-1} with an average -50.8 Kcal mol^{-1}. The pre-miRNAs length ranges from 53-392 nt with an average of 175 nt. The mature miRNA sequences length ranges from 19-24 nt. Of 26, the most (69.23%) of the Flax miRNAs have 21 nt length, followed by 20 nt (23%), 24 nt (3.85%) and 19 nt (3.85%). The maximum (27%) of the Flax miRNAs have shown 1 mismatches with their reference miRNA followed by perfectly conserved (23%), 2 (23%), 4 (15.4%) and 3 (11.5%) mismatches. Equal Flax miRNAs are located on the 5' (50%) and 3' (50%) arms of the pre-miRNAs (Fig. 1). The Flax miRNAs characterization in parameters of reference 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) and mature sequence arm (MSA) are given in Table 1. The stem regions of the stem-loop structures have contained the Flax mature sequences, as illustrated in Fig. 1. At least 12-21 nucleotides engaged in Watson-Crick or G/U base pairings between the mature miRNA and the opposite arms star sequence (miRNAs*) as illustrated in the stem-loop structures. Number of researcher groups reported similar results for the miRNAs in other plant species (Mica *et al.*, 2006; Barozi *et al.*, 2008; Fuilang *et al.*, 2010; Barozai *et al.*, 2011a).

The insilico identified miRNAs also need validation as a non-protein coding sequences. The same validation test was applied on the novel identified Flax miRNAs. No homology was observed in the Flax pre- miRNAs with known proteins. Similar results were reported for cotton and Picea (Barozai *et al.*, 2008; Barozai *et al.*, 2011b).

Phylogenetic and conservation studies: The phylogenetic and conservation studies were performed for the newly identified Flax miRNAs. According to the conservation studies the Flax miRNAs; miR-156 & miR-159 showed maximum 100% conserved nature with *Arabidopsis lyrata* (aly), *Arabidopsis thaliana* (ath), *Oryza sativa* (osa) and *Populus trichocarpa* (ptc) orthologues miRNAs as revealed in Fig. 2. Similar findings were given by Barozai *et al.*, in plants (Barozai *et al.*, 2008). The miRNA (miR-156) phylogenetic analysis showed that the Flax is more close to *Populus trichocarpa* (ptc) than the *Arabidopsis* (aly & ath) and *Oryza sativa* (osa) (Fig. 3).

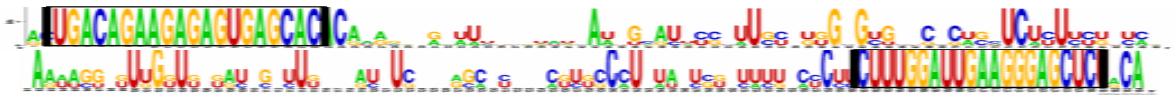

Fig. 2. The Flax miRNA conservation studies.
The Flax miRNAs' (miR-156 & miR-159; upper & lower illustrations respectively) showed 100% conservation (rectangle boxes) with *Arabidopsis*, *Oryza sativa* and *Populus trichocarpa* miRNAs.

Fig. 3. The Flax miRNA phylogenetic analysis.
The Phylogenetic analysis of the pre-miRNA (156) of Flax (lus) showed that on the basis of pre-miRNA sequences, it is more close to *Populus trichocarpa* (ptc) than the *Arabidopsis* (ath & aly) and *Oryza sativa* (osa).

The targets of flax miRNAs: The miRNAs targets identification is an interesting and demanding step for the novel annotated miRNAs. This paper summarizes a total of 48 target sequences (Table 2) for the novel 26 Flax miRNAs. The most of the targets are already reported as miRNA targets in other plant and animal species (Barozai *et al.*, 2008; Fuilang *et al.*, 2010).

One important protein family targeted by miRNA is the transcription factors (Fuilang *et al.*, 2010; Oswaldo *et al.*, 2010). The Flax miRNAs; miR-156 and miR-408-3p target MYB-related and Zinc finger transcription factors respectively. Other Flax miRNAs target rust resistance proteins, retrotransposons, development & growth related proteins and hypothetical proteins. Similar findings were reported for various plant species (Barozai *et al.*, 2008; Fuilang *et al.*, 2010; Oswaldo *et al.*, 2010).

Conclusion

For the first time in any Flax (*Linum*) species, a sum of 26 miRNAs from 19 families is reported. The gene regulation mechanism will be elucidated through these findings in the future. The proteins targeted by these miRNAs will also help in the engineering of such Flax that can combat and resist the biotic and abiotic stresses. Furthermore, the ESTs based identification confirmed the miRNAs expression.

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.

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*., doi:10.1007/s11033-012-1753-2.

Barozai, M.Y.K. and H.A.Wahid. 2012. Insilico identification and characterization of cumulative abiotic stress responding Genes in Potato (*Solanum tuberosum* L.) (Press).

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., 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. of 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.

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

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

Fuliang, X., P.F. Taylor and B. Zhang. 2010. Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*) *Planta.*, 232: 417-434.

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

Hammond, S.C., E. Bernstein, D. Beach and G.J. Hannon. 2000. An RNA-directed nuclease mediates posttranscriptional gene silencing in *Drosophila* cells. *Nat.*, 404: 293-296.

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

Kurt, O., H. Uysal and S. Uzun. 2005. Non-destructive leaf area estimation of Flax (*Linun usitatissimum* L.). *Pak. J. Bot.*, 37(4): 837-841.

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. Expl. Bot.*, 57 (11): 2601-2612.

Muhammad, Z. and F. Hussain. 2010. Vegetative growth performance of five medicinal plants under NaCl salt stress. *Pak. J. Bot.*, 42(1): 303-316.

Oswaldo, V., S.S. Yang, A. Rosaura, H.G. Peter, L.R. Jose, P.V. Carroll and H. Georgina. 2010. MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity *New Phyt.*, 187: 805-818.

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.

Weber, M.J. 2005. New human and mouse microRNA genes found by homology search. *FEBS J.*, 272: 59-73.

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

(Received for publication 15 April 2011)