BioTechnologia vol. 94(3) • pp. 285-290 • 2013
Journal of Biotechnology, Computational Biology and Bionanotechnology SHORT COMMUNICATION

Identification of miRNAs and their potential targets in halophyte plant *Thellungiella halophila*

BAHMAN PANAHI*[1], SEYED ABOLGASEM MOHAMMADI[1], ESMAIIL EBRAHIMIE[2]

[1] Department of Biotechnology and Plant Breeding, Tabriz University, Tabriz, Iran

[2] Schools of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia

*Corresponding author: panahibahman@ymail.com

Abstract

MicroRNAs (miRNAs) are a class of non-coding RNAs with important role in gene regulation in various organisms. These RNAs regulate gene expression at the post-transcriptional level. To date, several hundred plant miRNAs have been deposited in the miRBase database. Many of them are conserved during the evolution of terrestrial plants, suggesting that the well-conserved miRNAs may also retain homologous target interactions. So far, there has been no experimental or computational identification of miRNAs and their target genes in *Thellungiella halophila*. Here, using a computational homology based search approach and according to a series of filtering criteria, a total of 8 miRNAs belonging to 4 miRNA families were detected from the Expressed Sequence Tags (EST) databases. Potential target genes of these predicted miRNAs were subsequently assessed. Our findings showed that among the target genes, most of them encode transcription factors and enzymes that participate in regulation of development, growth and other physiological processes.

Keywords: *Thellungiella halophila*, Expressed Sequence Tags, microRNA, target genes

Abbreviations: ΔG – folding free energies, EST – Expressed Sequence Tag, miRNA – microRNA

Introduction

MicroRNAs (miRNAs) are single-stranded non-coding RNAs, which are approximately 21-22 nucleotides (nt) in length and play an important role in many biological processes (Zang et al., 2006). In plants, miRNA genes originate mostly from independent transcriptional units that are transcribed by RNA polymerase II into long primary transcripts (pri-miRNAs) (Chen, 2005). Subsequently, the pri-miRNAs are cut into miRNA precursors (pre-miRNAs) with stem-loop (hairpin) structure(s). The loop region of the hairpin is removed by the ribonuclease III-like enzyme Dicer (DCL1) (Kurihara and Watanabe 2004), and the remainder is exported to the cytoplasm by proteins called Hasty (Park et al., 2005). The mature miRNA is incorporated into the RNA induced silencing complex (RISC) and guides RISC to complement mRNA targets. Eventually, the RISC inhibits translation elongation, or triggers the degradation of target mRNA (Lin et al., 2005). Mature miRNAs bind to the complementary sites on target mRNAs and repress the expression of target RNAs through direct target cleavage or by translational repression (Axtell et al., 2007). In plants, miRNAs negatively regulate their target genes, which are involved in a range of developmental processes such as response to pathogen invasion (Zhang et al., 2006), hormone signaling (Guo et al., 2005), environmental stresses (Jones-Rhoades et al., 2004), and promotion of anti-viral defense (Gan et al., 2008). Identifying miRNAs and their target genes is therefore the key point in understanding their function and underlying control mechanisms in plants.

There are four approaches for identifying miRNAs: i) genetic screening (Lee et al., 1993), ii) direct cloning after isolation of small RNAs (Mead et al, 2008), iii) computational strategy, and iv) expressed sequence tags (ESTs) analysis (Mathews et al., 1999).

Compared to the experimental approaches, computational methods have been proved to be faster, affordable, and more effective, therefor contributing mostly to today's plentiful storage in miRBase (Griffiths-Jones S., 2006). Zhang et al. (2006) classified these computational approaches into five major categories: homology search-based, gene search-based, neighbor stem-loop search-based, comparative genomic algorithm-based and phylo-

genetic shadowing-based. The homology search-based approach, which is based on conserved sequences and secondary structures, searches nucleotide databases using the BLAST program, and has been used to identify new miRNAs in the genomes of model species, such as *Arabidopsis thaliana* (Adai et al., 2005), *Oryza sativa* (Li et al., 2005, Zhang et al., 2006) and *Lycopersicon esculentum* (Yin et al., 2008). Zhang et al. (2005) developed an efficient strategy for identifying plant miRNAs using EST analysis. This approach provides the opportunity of rapid analysis of miRNAs and their functions in species with limited knowledge about genome sequences (Zhang et al., 2006; Nasaruddin et al., 2007; Jin et al., 2008). Although to date, many glycophytes and medicinal plant miRNAs have been discovered (Carrington and Ambros, 2003), identification and target prediction of miRNAs in halophytic plants has not been as yet reported.

Thellungiella halophila is a halophyte which is a close relative of *A. thaliana*, and has been used as a model for studying salt tolerance in plants. *T. halophila* shares many of the advantages of a model system with *A. thaliana*. It has a small genome (less than twice the size of the *A. thaliana* genome), short life cycle, shows abundant seed production and can be easily transformed (Bressan et al., 2001).

The goal of this study is identification of new miRNAs and their targets in *T. halophila* to improve our understanding of the possible roles of miRNAs in regulating the growth and development of halophytic plants.

Materials and methods

Datasets of miRNAs and EST sequences

In order to search for potential miRNAs in *T. halophila*, previously known plant miRNAs, precursor sequences from *A. thaliana*, *Zea mays*, *O. sativa*, *Glycine max, Sorghum bicolor* and other plant species were downloaded from the miRBase (miRBase 18, August 2012) (http://www.mirbase.org) (Griffiths-Jones, 2006). These miRNAs (about 21,000 miRNAs) were defined as the reference set of miRNA sequences. To avoid the redundant or overlapping miRNAs, the repeated sequences of miRNAs within the above-mentioned species were removed and the remaining sequences were used as query sequences for BLAST search. For this purpose, *T. halophila* ESTs (about 40,000) and mRNA databases were obtained from the National Center for Biotechnology Information (NCBI) (http://ftp. ncbi.nlm.nih.gov).

Potential miRNAs and their precursors

The reference sequences were used as a query for homology search against our local nucleotide sequence database at e-value threshold < 0.01 using BLAST + 2.2.22 program (Altschul et al., 1997). The adjusted BLAST parameter settings were as follows: an expected value cutoff of 10; a low-complexity sequence filter; 1000 descriptions and alignments; and automatically adjusted parameters for short input sequences to improve the veracity of outputs. If the matched sequences were less than the previously known mature miRNA sequences, the non-aligned parts were inspected and manually compared to determine the number of matching nucleotides. All BLAST results were saved. EST sequences that have only 0-3 nt mismatches compared with the query miRNA sequences were chosen manually. The secondary structures of the selected EST sequences were predicted and generated using the web-based software Mfold 3.2 (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi) (Zuker, 2003). The following parameters were used in predicting the secondary structures: a fixed folding temperature of 37 °C; ionic conditions set at 1 M NaCl with no divalent ions; and grid lines in the energy dot plots turned on. Other parameters followed the default parameters. Generally, the lowest-energy structure corresponds to helices in optimal folding. The precursor sequences were searched at 50 nucleotides upstream or downstream from the location of mature miRNAs with an increment of 10 nucleotides. The following criteria were used to select RNA sequences as candidate miRNA precursors: i) RNA sequence can fold into an appropriate stem loop hairpin secondary structure, ii) mature miRNA sequence hairpin site in one arm of the structure, iii) miRNAs have less than four mismatches with the opposite sequence in the other arm (miRNA*), iv) no loop or break in miRNA sequences, v) predicted secondary structure has greater MFEs of −18 (≤ −18 kcal/mol).

Prediction of targets

The predicted *T. halophila* miRNAs were used as query against the *A. thaliana* DFCI gene index (AGI) release 13 using miRU (http://bioinfo3.noble.org/ psRNATarget/) following the criteria of i) maximum expectation value 3;

Table 1. Newly identified miRNAs from ESTs of *Thellungiella halophila*

New miRNA		Gene ID	NM (nt)	LM (nt)	LP (nt)	A+U (%)	ΔG (kal/mol)
tha-miR-3637	AUUUAUGUAUUGUGAAGAAGACAU	BY835382	2	24	459	75	−74.74
tha-miR-2646	AUGACAUUUAUUGAUAUUCAA	BY833381	1	21	508	81	−80.09
tha-miR-773a	AUGGACUAUGCAUUUCGAGU	BM985699	2	20	840	60	−165.03
tha-miR-773b	UCGUUCAAAAUCCCUCUUCUUU	BM986076	2	21	830	66	−176.16
tha-miR-773c	GAGUUGAUGGACGUUGUGAUC	BM986112	2	21	830	52	−183.06
tha-miR-773d	AUGGACUAUGCAUUUUGAGUUU	BY808734	1	22	635	68	−209.65
tha-miR-773e	AUGGACUAUGCAUUUCGAGUUU	BY800541	1	22	659	63	−229.1
tha-miR-2650	CAAGCUGCUUUUUGUAGAAAUUC	BY810494	1	23	582	65	−118.59

NM – number of mismatch; LM – length of mature miRNAs; LP – length of precursor; ΔG – folding free energies

Table 2. Potential targets of the identified miRNAs in *Thellungiella halophila*

MiRNAs	Target sites	Targeted proteins	E value	Gene IDs
tha-miR3637	1	Early nodulin-like protein	1.5	AT5G53870
	1	Leucine-rich repeat protein kinase family protein	1.5	AT3G42880
	1	Protein kinase superfamily protein	1.5	AT1G53165
	1	F-box and associated interaction domains-containing protein	2	AT5G56750
	1	N-MYC down regulated-like 1	2	AT5G56750
	1	VH1-interacting kinase	1.5	AT1G14000
	1	Global transcription factor group B1	2.5	AT1G65440
	1	BIG PETAL P	2.5	AT1G59640
	1	Transposable element gene	2.5	AT4G07915
	1	Zinc finger protein	2.5	AT3G18290
	1	Pseudogene of leucine-rich repeat protein	2	AT3G25495
	1	S1FA-like DNA-binding protein	3	AT3G09735
	1	Endoplasmic reticulum retention defective	3	AT3G25040
	1	Galactose oxidase/kelch repeat superfamily protein	3	AT1G19460
	1	CCT motif family protein	2.5	AT1G63820
tha-miR2646	1	Adenine nucleotide alpha hydrolases-like superfamily protein	1	AT3G01520
	1	DegP protease 13	1	AT5G40560
	1	Elicitor-activated gene 3-2	1	AT4G37990
	1	Pentatricopeptide repeat (PPR) superfamily protein	1	AT5G65820
	1	Transposable element gene	1	AT5G43800
	1	Clathrin light chain protein	1.5	AT2G20760
	1	Beta glucosidase 18	1.5	AT1G52400
	1	Vacuolar iron transporter (VIT) family protein	1	AT4G27860
	1	Transposable element gene	1.5	AT5G31702, AT4G02960, AT3G28153, AT5G33428

Table 2. Continued

MiRNAs	Target sites	Targeted proteins	E value	Gene IDs
tha-miR2646	1	Eukaryotic translation initiation factor	2	AT3G60240
	1	Meprin and TRAF (MATH) homology domain-containing protein	2	AT3G44800
tha-miR773	1	Granulin repeat cysteine protease family protein	0.5	AT5G43060, AT3G19390
	1	Cysteine proteinases superfamily protein	1.5	AT1G29090
	1	PQ-loop repeat family protein / transmembrane family protein	2.5	AT4G36850
	1	Cysteine proteinases superfamily protein	2.5	AT3G45310
	1	Senescence-associated gene 12	3	AT5G45890
	1	Defensin-like (DEFL) family protein	2.5	AT5G27495
	1	Cysteine proteinases superfamily protein	3	AT3G19400
	1	Papain family cysteine protease	2.5	AT4G11320
	1	UbiA prenyltransferase family protein	3	AT1G60600
	1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	3	AT1G62510
	1	Pyridoxamine 5'-phosphate oxidase family protein	2.5	AT2G46580
	1	Protein-tyrosine phosphatase-like, PTPLA	3	AT5G59770
	1	Proteasome subunit PAB1	3	AT1G16470
	1	NAD(P)-binding Rossmann-fold superfamily protein	3	AT2G24190
tha-miR2650	1	Hydrolases superfamily protein	0.5	AT2G42450
	1	Cold regulated 15b	1	AT2G42530
	1	Gibberellin 2-oxidase 7	1	AT1G50960
	1	Cox19 family protein (CHCH motif)	1	AT1G09794
	1	ARM repeat superfamily protein	1.5	AT3G02840
	1	BET1P/SFT1P-like protein 14A	1.5	AT3G58170
	1	Ubiquitin-conjugating enzyme 32	1.5	AT3G17000
	1	Alpha 1,4-glycosyltransferase family protein	1.5	AT5G01250
	1	NIMA-related kinase 3	1.5	AT5G28290
	1	Cullin	1	AT4G02570
	1	Poltergeist like 1	1.5	AT2G35350
	1	Transposable element gene	1.5	AT4G06523

ii) multiplicity of target sites 2; iii) range of central mismatch for translational inhibition is from nucleotide 9 to 11; iv) maximum mismatches at the complementary site ≤ 4 without any gaps.

Nomenclature of miRNAs

Names were assigned to the predicted miRNAs in a similar pattern to miRBase (Griffiths-Jones, 2006). The mature sequences were designated "miR", and the precursor hairpins were labeled as "mir" with the prefix "tha" for *T. halophila*. In the cases where distinct precursor sequences had identical miRNAs with different resources and mismatch pattern, they were named as tha-mir-1-a and tha-mir-1-b. In this research work, both tha-mir-3a and tha-mir-3b have identical miRNA sequences; however, the resources were different (Table 1, Fig. 1).

Results and discussion

Identification of potential Thellungiella halophila miRNAs

Most of mature miRNAs are evolutionarily conserved from species to species within the plant kingdom,

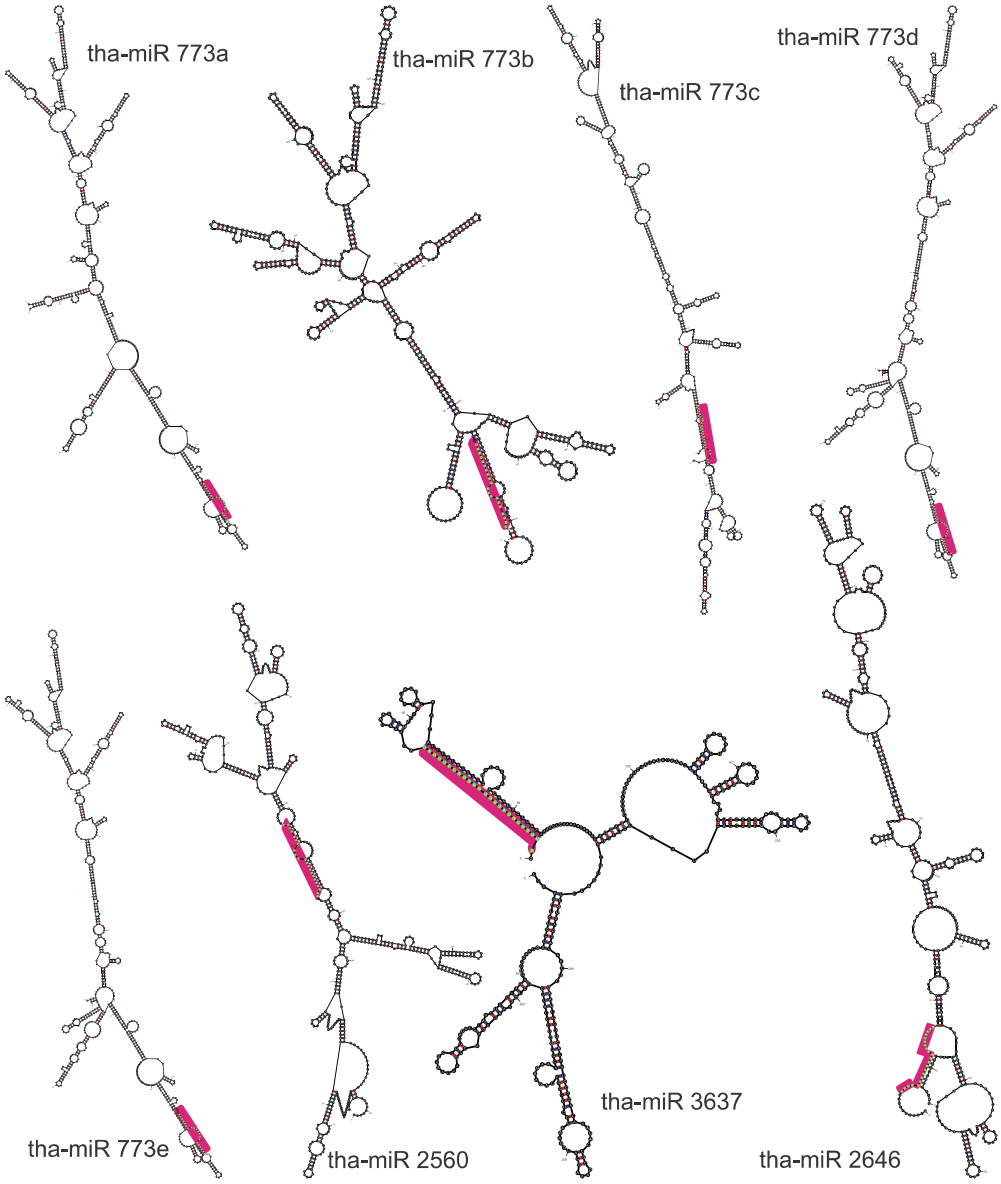

Fig. 1. Mature and precursor sequences and the predicted stem and loop structures of newly identified miRNAs in *Thellungiella halophila*

which facilitates the prediction of the existence of new miRNA homologs in other plant species (Axtell and Bartel, 2005). A total of 16,516 ESTs were obtained from the GenBank. As described in Materials and Methods section, following the BLASTN searches, all similar hits except for coding sequences, were maintained for the secondary structure analysis. Eventually, 8 potential *T. halophila* miRNAs assorted to 4 miRNA families were identified in the EST database (Table 1). During screening for the potential miRNAs, sequences of the candidate pre-miRNAs were evaluated for their A+U content, which ranged from 52 to 81% (Table 1). The average length of ESTs was 654 nt and the longest was 795 nt, while most of miRNA precursors had 80–150 nt as identified by MirEval software (Ritchie, 2008). This suggests that the EST may contain other element sequences in addition to miRNA precursor sequence (Zhang et al., 2006). The lengths of these newly identified miRNAs varied from 20 to 24 nt.

Predicted targets for identified miRNAs

Using the newly identified miRNA sequences as BLAST search queries, the target mRNAs were identified and divided into several groups (Table 2). The largest group contained targets that most probably encode regulatory factors known to be involved mainly

in plant growth and developmental patterns. Other groups contained targets encoding a range of different proteins implicated in various metabolic processes, hormone responses, stress defense and signaling, transcription factors, etc. (Table 2).

The miRNA target gene identification is an important step for understanding the role of miRNAs in a gene regulatory network. Our prediction of target genes for the *T. halophila* miRNAs is in line with the knowledge that more than one gene is usually regulated by an individual miRNA and is similar to the recent findings in other plant species (Jones-Rhoades and Bartel, 2004; Das and Mondal, 2010). These findings suggest that miRNA research should be focused on networks rather than individual connections between miRNA and strongly predicted targets.

References

Adai A. et al. (2005) *Computational prediction of miRNAs in Arabidopsis thaliana*. Genome Res. 15: 78-91.

Altschul S.F., Madden T.L., Schäffer A.A., J. Zhang, Zhang Z., Miller W., Lipman D.J. (1997) *Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs*, Nucl. Acids Res. 25: 3389-3402.

Axtell M.J., Bartel D.P. (2005) *Antiquity of microRNAs and their targets in land plants*. Plant Cell 17: 1658-1673.

Bressan R.A., Zhang C., Zhang H., Hasegawa P.M., Bohnert H.J., Zhu J.K. (2001) *Learning from the Arabidopsis experience. The next gene search paradigm.* Plant Physiol. 127: 1354-1360.

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

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

Das A., Mondal T.K. (2010) *Computational Identification of Conserved microRNAs and Their Targets in Tea(Camellia sinensis).* Amer. J. Plant Sci. 1: 77-86.

Griffiths-Jones S. (2006) *miRBase: the microRNA sequence database.* Meth. Mol. Biol. 342: 129-138.

Guo S.H., Xie Q., Fei F.J., Chua H.N. (2005) *MicroRNA Directs mRNA Cleavage of the Transcription Factor NAC1 to Downregulate Auxin Signals for Arabidopsis Lateral Root Development*. Plant Cell 17: 1376-1386.

Jin W., Li N., Zhang B., Wu F., Li W., Guo A., Deng Z. (2008) *Identification and Verification of microRNA in Wheat (Triticum aestivum).* J. Plant Res. 121: 351-355.

Jones-Rhoades M.W., Bartel P.D. (2004) *Computtional Identification of Plant microRNAs and Their Targets, Including a Stress-Induced miRNA.* Mol. Cell 14: 787-799.

Kurihara Y., Watanabe Y., (2004) *Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions.* Proc. Natl. Acad. Sci. USA 101: 12753-12758.

Larkin M., Blackshields G., Brown N.P. et al. (2007) *Clustal W and Clustal X*. Bioinformatics 23(21): 2947-2948.

Lee R.C., Einbaum R.L., Ambros V. (1993) *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell 75: 843-854.

Li Y., Li W., Jin Y.X. (2005) *Computational identification of novel family members of microRNA genes in Arabidopsis thaliana and Oryza sativa*. Acta Biochim. Biophys. Sin. (Shanghai) 37: 75-87.

Lin S.L., Chang D., Ying S.Y. (2005) *Asymmetry of intronic pre-miRNA structures in functional RISC assembly*. Gene 356: 32-38.

Lu Y., Gan Q., Chi X., Qin S. (2008) *Roles of microRNA in Plant Defense and Virus Offense Interaction*. Plan Cell Rep. 27: 1571-1579.

Mathews D.H., Sabina J., Zuker M., Turner D.H. (1999) *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure*. J. Mol. Biol. 288: 911-940.

Mead E.A., Tu Z. (2008) *Cloning, characterization, and expression of microRNAs from the Asian malaria mosquito, Anopheles stephensi*. BMC Genomics 23: 244-253.

Nasaruddin M.N., Harikrishna K., Othman Y.R., Hoon S.L., Harikrishna A.J. (2007) *Computational Prediction of microRNAs from Oil Palm (Elaeis guineensis Jacq.) Expressed Sequence Tags*. Asian Pacific J. Mol. Biol. Biotechnol. 15: 107-111.

Park M.Y., Wu G., Gonzalez-Sulser A., Vaucheret H., Poethig R.S. (2005) *Nuclear processing and export of microRNAs in Arabidopsis*. Proc. Natl. Acad. Sci. USA 102: 3691-3696.

Ritchie W. et al. (2008) *Mireval: a web tool for simple microRNA prediction in genome sequences*. Bioinformatics 24: 1394-1396.

Tamura K., Dudley J., Nei M., Kumar S. (2007) *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. Mol. Biol. Evol. 24: 1596-1599

Yin Z., Li C., Han X., Shen F. (2008) *Identification of conserved microRNAs and their target genes in tomato (Lycopersicon esculentum).* Gene 41: 460-466.

Zhang B., Pan X., Anderson A.T. (2006) *Identification of 188 Conserved Maize microRNAs and Their Targets*. FEBS Lett. 580: 153753-153762.

Zhang B., Pan X., Cobb G.P., Anderson T. (2006) *Plant microRNA: A Small Regulatory Molecule with Big Impact*. Develop. Biol. 289: 3-16.

Zujun Yin, Chunhe Li, Xiulan Han, Fafu Shen (2008) *Identification of conserved microRNAs and their target genes in tomato (Lycopersicon esculentum).* Gene 414: 60-66.

Zuker M. (2003) *Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction*. Nucl. Acids Res. 31: 3406-3415.