Original scientific paper
10.7251/AGRENG1803122A
UDC633:34+582.736.3]:577.2

IDENTIFICATION OF IN SILICO MIRNAS IN FOUR PLANT SPECIES FROM FABACEAE FAMILY

Bihter AVSAR[1*], Danial ESMAEILI ALIABADI[2]

[1]Sabanci University, Nanotechnology Research and Application Centre, Istanbul, Turkey
[2]Sabanci University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
*Corresponding author: bihteravsar@sabanciuniv.edu

ABSTRACT

Plant microRNAs (miRNAs) are small non-coding RNAs, about 21-24 nucleotides, which have critical regulatory roles on growth, development, metabolic and defense processes. Their identification, together with their targets, have gained importance in exploring their parts on functional context, providing a better understanding of their regulatory roles in critical biological processes. With the advent of next-generation sequencing technologies and newly developed bioinformatics tools, the identification of microRNA studies by computational methods has been increasing. In the presented study, we identified some putative miRNAs for *Cicer arietinum*, *Glycine max*, *Medicago truncatula* and *Phaseolus vulgaris* genomes. We also provided the similarity between those organisms regarding common/different miRNAs availability throughout their genomes. According to the data, the highest similarity was found between *Glycine max* and *Phaseolus vulgaris*. We also investigated the potential targets of putatively identified miRNAs for each organism. We analyzed which miRNA families were expressed *in silico*. We also showed the representation (copy number of genes) profile of predicted putative miRNAs for each organism. Since most of the food products and animal feeds consist of Fabaceae family members as it is mentioned above, these findings might help to elucidate their metabolic and regulatory pathways to use them efficiently in biotechnological applications and breeding programs.

Keywords: microRNA, *Cicer arietinum*, *Medicago truncatula*, *Glycine max*, *Phaseolus vulgaris*.

INTRODUCTION

Recently, the sufficiency of food demands becomes a critical issue since the increasing world population, drastic changes in climate and the a/biotic stress factors has threated the sustainability of agricultural production. Therefore, there is an immediate need to develop new farming technologies and biotechnological applications (Akpinar et al., 2012).

As one of the most critical and useful development, next-generation technologies help us to unravel the complex genomes of organisms in addition to having a significant impact on reducing the cost, time and required effort compare to the previous methods such as Sanger sequencing. Based on different sequencing technologies, various computational tools and analysis methods were developed. Computational microRNA identification studies on plant genomes have been increased and contributed to the recent literature efficiently. MicroRNAs (miRNAs) are small, about 21-24 nucleotides, endogenous non-coding RNAs that play various roles in plants. They are derived from the stem-loop structure, and some specific enzymes modify them. Plant microRNAs control the expression of genes encoding multiple transcription factors, stress-responsive elements, and the other proteins have roles in growth, development and physiological properties (Rogers and Chen, 2013). Computationally identified miRNAs has reached to the successful means, and some new miRNAs were identified experimental methods. These experimentally identified miRNAs had roles on abiotic stresses due to drought, salinity, heat, cold or phosphorous deficiency or biotic stresses. Currently, computational miRNA prediction is based on two approaches: 1.) Homology-based for conserved miRNA identification 2.) Some other algorithms which use support vector machine by setting some characteristics for pre-miRNA structure (Zhang et al., 2006).In our study, we used the 'homology-conserved' method to predict some putative miRNAs via using in-house Perl scripts (Avsar and Aliabadi, 2017a; Avsar and Aliabadi 2018). Legumes belong to the Fabaceae family are essential nutritional sources for foodstuffs and animal feeds. Their rich protein, starch content, oil, fiber content and the high efficiency of nitrogen fixation properties make Legumes highly valuable in the cropping cycle, and therefore they account for one-third of global primary crop production (Mantri et al., 2013). In this study, four different legume genomes were studied due to their economic importance and/or their suitable model features: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Medicago truncatula* and *Phaseolus vulgaris* (common bean). The genomes of these species have been completely sequenced, and they are available in NCBI. We putatively identified miRNAs for each species, and we compared their microRNA atlas to each other as well as the model organism "*Medicago truncatula*." These findings may help us to have a better understanding of the roles of miRNAs in abiotic stress, the miRNAs involved in symbiosis and nutrition homeostasis.

MATERIAL AND METHODS

Reference miRNAs and Datasets: Currently available mature miRNA sequences belong to Viridiplantae (8,496 sequences and 73 plant species) were downloaded from miRBase release 21 (Kozomara and Griffiths-Jones, 2013). miRBase corresponds to 4,802 unique mature miRNA sequences, and these mature miRNAs were used as a query in homology-based *in silico* miRNA identification. Legumes genomes were retrieved from NCBI. All plant assemblies were downloaded from

NCBI (GenBank accessions: GCA_000004515.3, GCA_000499845.1, GCA_000331145.1, GCA_000219495.2*).*

Homology conservation approach for miRNA identification: The prediction was employed using two previously developed, in-house Perl scripts: SUmirFind and SUmirFold[1]. In the first step of homology-based miRNA prediction, BLAST+ stand-alone toolkit, version 2.2.25 (Camacho, 2009) was used for detection of database sequences with homology (mismatch cutoff parameter set to <=3) to previously known plant mature miRNAs (Avsar and Aliabadi, 2015). In the second step, UNAFold version 3.8 was used with parameters optimized to include all possible stem-loops generated for each miRNA query to obtain secondary structures of predicted miRNAs. Perl scripts eliminated hairpins with multi-branched loops, with inappropriate DICER cut sites at the ends of the miRNA-miRNA* duplex, or with mature miRNA sequence portions at the head of the pre-miRNA stem-loop.

Representative miRNAs (gene copy number) on target genomes: The miRNA gene copy numbers were identified based on the output data from SUmirFold process mentioned in section Homology conservation approach for miRNA identification. Identical miRNA families that were resulted from the similar miRNA stem-loop sequences were eliminated to avoid over-representation.

Expressed Sequence Tag (EST) analysis, miRNA targets and target annotations of predicted genomic miRNAs: For EST analysis, the pre-miRNA sequences were retrieved, and the duplicate sequences were removed to prevent over-representation. By using the BLAST+ stand-alone toolkit, version 2.2.25, pre-miRNA sequences were blasted to EST sequences specific to each organism obtained from NCBI (Avsar and Aliabadi 2017b). The strict criteria (above the threshold as 98% identity and 99% query coverage) were used for the identification of expressed miRNA families. Mature sequences were identified, and duplicates were removed. By using online web tool, psRNA, the mature query sequences were blasted against to EST sequences. The resulting file was used for gene ontology analysis by using Blast2Go software (Conesa and Götz, 2008). The predicted mature miRNA sequences were also searched in miRBase database website to confirm their experimentally validated targets.

RESULTS AND DISCUSSIONS

Putative miRNAs in Fabacea family members: We predicted as a total of 198 putative miRNA families. Out of 198 putative miRNA families 42, 150, 44, 41 putative miRNA families in *Cicer arietinum*, *Glycine max*, *Medicago truncatula* and *Phaseolus vulgaris* genomes, respectively and 42 common miRNAs were found between all organisms (Table 1).

[1]http://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0040859.s003

Table 1. Putative miRNA families identified for each organism. Ca: *Cicer arietinum*, Gm: *Glycine max*, Mt: *Medicago truncatula*, Pv: *Phaseolus vulgaris*

Ca	Gm				Mt	Pv	Common
miR1130	miR160	miR2606	miR4406	miR9765	miR172	miR160	miR160
miR1511	miR1507	miR403	miR4410	miR1526	miR1030	miR1510	miR1510
miR1514	miR1508	miR4340	miR482	miR2089	miR1120	miR1512	miR1512
miR156	miR1509	miR4342	miR4996	miR2218	miR1128	miR1514	miR1514
miR157	miR1510	miR4343	miR5030	miR3522	miR1439	miR1515	miR1527
miR159	miR1512	miR4344	miR5034	miR4355	miR1525	miR1527	miR156
miR160	miR1513	miR4345	miR5035	miR4394	miR159	miR156	miR157
miR162	miR1514	miR4346	miR5037	miR4413	miR2118	miR159	miR159
miR164	miR1516	miR4347	miR5038	miR477	miR2218	miR162	miR162
miR165	miR1517	miR4348	miR5041	miR5205	miR2592	miR164	miR164
miR166	miR1520	miR4349	miR5042	miR5370	miR2593	miR165	miR165
miR167	miR1521	miR4350	miR5043	miR5763	miR2599	miR166	miR166
miR168	miR1527	miR4352	miR530	miR5773	miR2600	miR167	miR167
miR169	miR1531	miR4356	miR5372	miR5774	miR2601	miR168	miR168
miR170	miR1535	miR4359	miR5376	miR9742	miR2602	miR169	miR169
miR171	miR156	miR4360	miR5377	miR9743	miR2603	miR170	miR170
miR172	miR157	miR4361	miR5378	miR9766	miR2605	miR171	miR171
miR2099	miR159	miR4363	miR5380	miR9767	miR2606	miR172	miR172
miR2111	miR162	miR4364	miR5667		miR2607	miR2111	miR2111
miR2118	miR164	miR4365	miR5670		miR2608	miR2118	miR2118
miR2218	miR166	miR4366	miR5775		miR2619	miR2119	miR2119
miR2618	miR167	miR4367	miR5780		miR2627	miR2218	miR2218
miR2630	miR168	miR4368	miR5784		miR2629	miR319	miR2606
miR319	miR169	miR4369	miR862		miR2630	miR390	miR2630
miR390	miR171	miR4371	miR9723		miR2636	miR391	miR319
miR393	miR172	miR4372	miR9730		miR2652	miR393	miR390
miR394	miR1863	miR4373	miR9732		miR2655	miR394	miR393
miR395	miR2107	miR4374	miR9734		miR2670	miR395	miR394
miR396	miR2109	miR4376	miR9735		miR2671	miR396	miR395
miR397	miR2111	miR4380	miR9736		miR319	miR397	miR396
miR398	miR2118	miR4382	miR9739		miR399	miR398	miR397
miR399	miR2119	miR4384	miR9745		miR482	miR399	miR398
miR5037	miR319	miR4387	miR9746		miR5161	miR403	miR399
miR5205	miR390	miR4388	miR9749		miR5205	miR4376	miR403
miR5213	miR393	miR4390	miR9752		miR5249	miR4407	miR4376

miR5281	miR394	miR4391	miR9753	miR5281	miR4416	miR482
miR5287	miR395	miR4392	miR9754	miR5282	miR482	miR5037
miR529	miR396	miR4393	miR9755	miR5287	miR5037	miR5205
miR530	miR397	miR4395	miR9756	miR530	miR529	miR5281
miR5741	miR398	miR4399	miR9757	miR5554	miR530	miR5287
miR6275	miR399	miR4401	miR9761	miR5561	miR829	miR529
miR6440	miR408	miR4402	miR9762	miR5745		miR530
	miR5281	miR4404	miR9763	miR7696		
	miR529	miR4405	miR9764	miR7701		

According to the results, *G.max-P.vulgaris* had more common miRNAs (34) whereas *M.truncatula-P.vulgaris* (8) shared the least amount of common miRNA families. The miRNA repertoire depends on genome size so *G.max* (about 980 MB) may have more miRNA families on its genome than the other organisms: *P.vulgaris* (about 521 MB), *C.arietinum* (about 530 MB), *M.truncatula* (about 412 MB). For each organism, putative miRNA families gave detailed information including conserved miRNA ID, miRNA* sequence, pre-miRNA stem sequences, calculations related to MFE, MFEI and GC%. Lower MFE values show the high stability of predicted miRNAs. Minimal folding free-energy index (MFEI) values which were calculated using MFE and GC% values differentiate miRNAs with typically higher MFEIs (>0.67) from other types of cellular ssRNAs for which MFEIs were previously characterized; transfer RNAs (0.64), ribosomal RNAs (0.59), and mRNAs (0.62–0.66) (Schwab et al., 2005).

Representation of putative miRNAs on genomes: In here, we used unmasked data to find representatives of miRNA families on genomes. According to this analysis, for *P.vulgaris* and *C. arietinum*, highly representative miRNA families, miR171, was similar. However, for *G.max* and *M.truncatula*, miR1520 and miR5281 families were profoundly found, respectively (Figure 1). Low representations of miRNA families (less than ten copy number) were calculated, but they are not included in the graphs since they might be contamination or 'young-miRNAs.' On the other hand, the highest number of hits might be caused by repetitive elements because most of the transposable elements were domesticated into microRNA genes (Li et al., 2011).

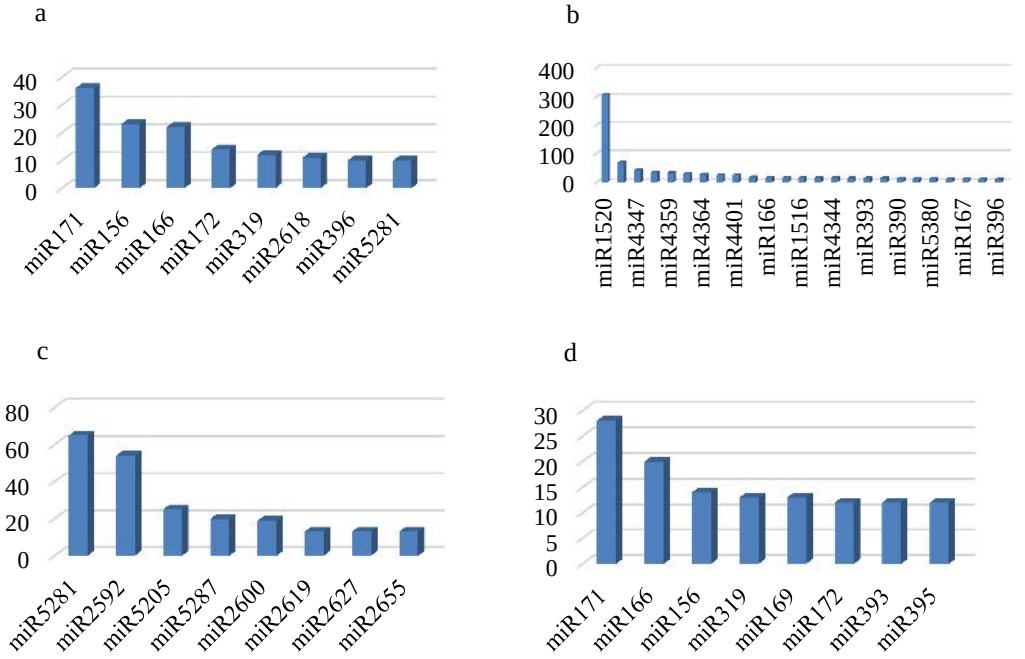

Figure 1. Representative miRNA families on genomes. a: *C. arietinum*, b: *G.max*, c: *M.truncatula*, d: *P.vulgaris*

Target prediction, gene ontology and expression analysis of identified miRNAs: We identified targets of putative miRNAs and their possible functions in the cell. As biological processes mechanisms, putative miRNA targets were mostly found in metabolic and cellular processes. Only *G.max* putative miRNAs targeted the genes found in the cellular component organization or biogenesis processes (Figure 2a). Putative miRNA targets were identified in almost all cellular components, however, for the macromolecular complex part, only *C.arietinum* and *M.truncatula* had low percent of target sequences (Figure 2b). Molecular functions of putative miRNA targets were also detected for all organisms. Catalytic activity and binding functions had the highest percentage whereas structural molecule activities of targets were only identified for *C.arietinum* putative miRNAs (Figure 2c).

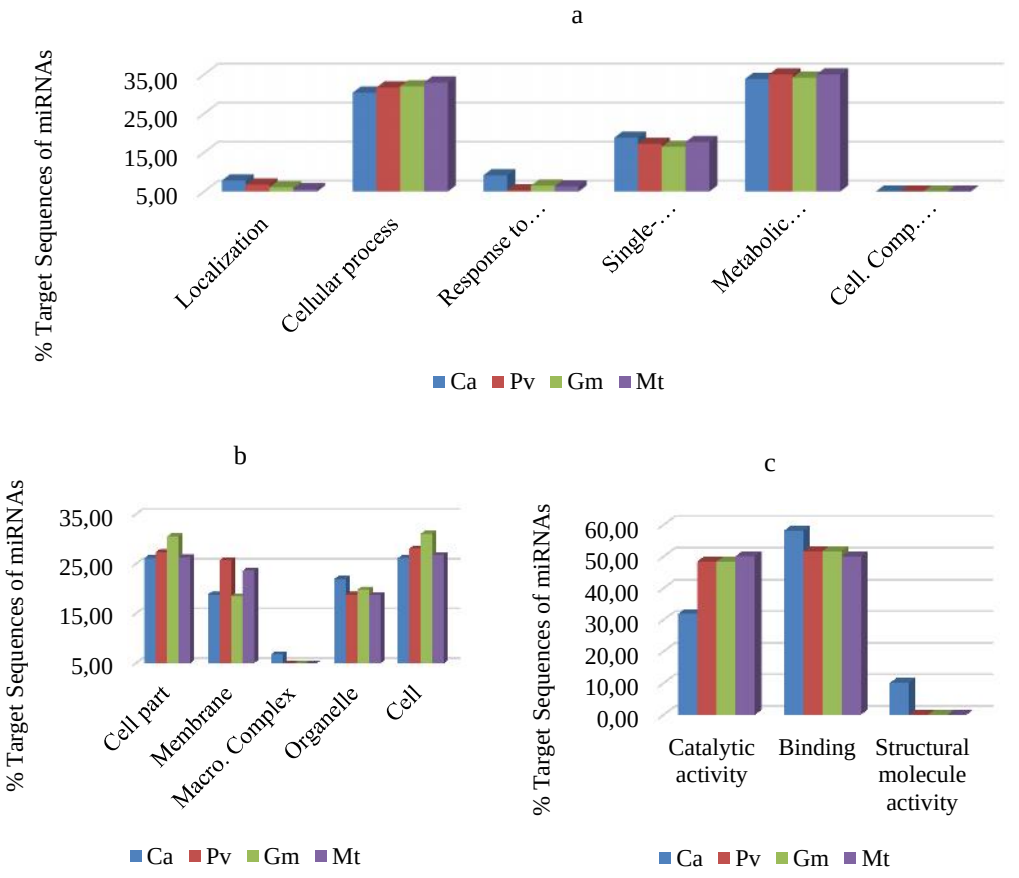

Figure 2. a: Biological processes of miRNA targets, b: Cellular component of miRNA targets, c: Molecular functions of miRNA targets. Ca: *Cicer arietinum*, Gm: *Glycine max*, Mt: *Medicago truncatula*, Pv: *Phaseolus vulgaris*

We also analyzed the expression of the predicted miRNAs *in silico*. For this purpose, the pre-miRNA sequences from each miRNA families were selected and blasted against to EST databases of each organism. In *C.arietinum*, only miR156 families had high homology to different EST sequences in GenBank. In *G.max*, we found 34 different miRNA families (miR1507, miR1508, miR1509, miR1510, miR1514, miR1520, miR156, miR160, miR162, miR166, miR167, miR168, miR171, miR172, miR2089, miR210, miR2109, miR211, miR2218, miR319, miR3522, miR394, miR395, miR396, miR398,miR399, miR403, miR408, miR482, miR4996, miR5038, miR529, miR5372, miR5667) showed a high homology to EST sequences. In *M.truncatula*, eight putative miRNAs were identified as miR159, miR2118, miR2218, miR319, miR399, miR482, miR5281, miR7696. For *P.vulgaris*, miR151, miR167, miR168, miR171, miR211, miR2118, miR221 and

miR399 families were given positive results according to the threshold mentioned in Materials and Methods section. For EST databases retrieved from NCBI, *C.arietinum* had the least amount of EST sequences whereas *G.max* had the most amount of EST sequences. Therefore, this may affect the identified *in silico* expressed miRNA families that show variation between the organisms.

CONCLUSIONS

MicroRNA discoveries provide us an opportunity to understand better complex regulatory systems in plants and in particular those involved in a/biotic stress conditions. This study helps research community to develop stress-tolerant crops by breeding programs. Additionally, unraveling the roles of miRNAs in the symbiotic relationships of legumes in overcoming several important agriculturally limiting environmental stresses is of high priority. Our findings may also help researchers to understand the regulatory roles of putative miRNAs in Fabaceae species which show genetic diversities and those which was analyzed by some molecular markers (Avsar, 2011). For the future studies, widely distributed and highly conserved miRNA families should be experimentally validated. These miRNAs are known as essential elements in different mechanisms ranging from abiotic stress tolerance to seed development. Furthermore, performing evolutionary studies for close relatives to understand their similarities/differences based on the miRNA repertoires and the functions of these putative miRNAs inside the organisms are valuable.

REFERENCES

Akpinar, B. A., Avsar, B., Lucas, S. J., Budak, H. (2012). Plant abiotic stress signaling. *Plant signaling & behavior*, *7*(11): 1450-1455.

Avsar, B. (2011). Genetic diversity of Turkish spinach cultivars (Spinacia oleracea L.). *A master dissertation, graduate school of engineering and sciences, Izmir Institute of Technology, Turkey*.

Avsar, B., Esmaeili Aliabadi, D. (2015). Putative microRNA analysis of the kiwifruit Actinidia chinensis through genomic data. *International Journal of Life Sciences Biotechnology and Pharma Research*, *4*(2): 96-99.

Avsar, B., Aliabadi, D. E. (2017). In silico analysis of microRNAs in Spinacia oleracea genome and transcriptome. *International Journal of Bioscience, Biochemistry and Bioinformatics*, *7*(2): 84.

Avsar, B., Esmaeilialiabadi, D. (2017). Identification of microRNA elements from genomic data of European hazelnut (Corylus avellana L.) and its close relatives. *Plant Omics*, *10*(4):190-196.

Avsar, B., Aliabadi, D.E. (2018). In silico identification of microRNAs in 13 medicinal plants. *Turkish Journal of Biochemistry*.42(s1): 57.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, *10*(1): 421.

Conesa, A., Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics*, *2008*.

Kozomara, A., Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, *42*(D1): D68-D73.

Li, Y., Li, C., Xia, J., Jin, Y. (2011). Domestication of transposable elements into microRNA genes in plants. *Plos one*, *6*(5): e19212.

Mantri, N. , Ford, R. , Pang, E. , Pardeshi, V., Basker, N. (2013). The role of miRNAs in legumes with a focus on abiotic stress response. *The Plant Genome*, 1-43.

Rogers, K., Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. *The Plant Cell*, *25*(7): 2383-2399.

Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. *Developmental cell*, *8*(4): 517-527.

Zhang, B., Pan, X., Wang, Q., Cobb, G. P., Anderson, T. A. (2006). Computational identification of microRNAs and their targets. *Computational biology and chemistry*, *30*(6): 395-407.