EVOLUTIONARY INSIGHTS INTO MICRORNAS OF KIWIFRUIT *ACTINIDIA CHINENSIS* AND ITS CLOSE RELATIVES

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

[1]Sabanci University, Nanotechnology Research and Application Centre, Tuzla, 34956, Istanbul, Turkey
[2]Sabanci University, Faculty of Engineering and Natural Sciences, Tuzla, 34956, Istanbul, Turkey
*Corresponding author: bihteravsar@sabanciuniv.edu

Abstract

The advent of next-generation sequencing technologies and newly developed bioinformatics tools have provided us complete sequence information of organisms. Plant microRNAs (miRNAs), small non-coding RNAs about 21-24 nucleotides, and their regulatory roles in biological processes have been uncovered since the identification of the first miRNA. MicroRNA biogenesis and modes of actions have also been elucidated in previous studies.
In the presented study, we identified putative microRNAs from *Actinidia chinensis*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Solanum tuberosum* and *Vitis vinifera* to compare their miRNA repertoire. According to the results, the highest synteny was found between *V.vinifera-A.chinensis* and the least synteny was found between *A.thaliana-A.chinensis*. The highest number of putative miRNAs were identified from *A.thaliana* whereas the least amount of putative miRNAs were identified from *V.vinifera*. This may be depended on the size of the genomes. We also analyzed the targets of putatively identified miRNAs for each organism. Expectedly, the target pathways of the predicted putative miRNAs were similar between the closest organisms. Expressed miRNA families and copy number of miRNA genes were compared between all organisms. In *A.thaliana*, the number of expressed putative miRNAs are more than the other organisms. For all the organisms, different miRNA families had the high copy number of genes. Therefore, highly represented miRNA families on each genome may have specific functional roles. The findings in this study will help the research community to identify the roles of miRNA players on critical biological pathways.

Keywords: MicroRNAs*, Actinidia chinensis, Solanum tuberosum, Solanum lycopersicum, Vitis vinifera*

Introduction

Increasing world population, the drastic climate changes, the threat of biotic stress and abiotic stress factors on the plants and the scarcity of arable lands have created a big concern about the future sufficiency of food demands. To prevent this unwanted scenario in the next centuries, some new agricultural technologies should have been developed to ensure food security (Liu and Chen, 2010; Akpinar et al., 2012; Avsar, 2011). Recent studies show that revolutionary advances in next-generation sequencing techniques became popular in plant biology area. The advent of this new technologies has increased our knowledge from genomics to plant breeding and evolutionary studies. An increasing number of sequenced plant genomes will provide us the greater understanding of their genome, growth and developmental mechanism and evolutionary processes (Egan et al., 2012).
Non-coding RNAs or ncRNAs, such as transfer RNA (tRNA), ribosomal RNA (rRNA) and small nuclear RNA (snRNA) have been identified for a long time ago. Among ncRNAs, microRNAs (miRNAs) and small interfering RNA (siRNA) have gained more attention since they have been found as essential regulators of gene expression (Choudhuri, 2009). MicroRNAs have short lengths about 21-24 nucleotides emerged from their longer precursor sequences that are variable sizes between the plant species contrary to the animal

counterparts. After the discovery of plant miRNAs and their great regulatory roles in the cell, both computational and experimental strategies have been studied for identification of the related genes and microRNAs' targets. Experimental methods are based on direct cloning and genetic screening whereas computational methods are dependent on new generation sequencing technologies because this powerful technology is time-saving and cost-efficient. With the help of the bioinformatics and computational tools, miRNA studies have also been increased enormously for plant genomes. (Zhang and Wang, 2015; Avsar and Aliabadi, 2017a; Avsar and Aliabadi, 2017b; Avsar and Aliabadi, 2018).

The kiwifruit or Chinese gooseberry is the edible berry member in the Actinidia genus and recent genomic studies reveals that the kiwifruit species often have polyploidy structure with a chromosome number as x=29 results from hexaploidization and two more recent whole genome duplication events (McNeilage and Considine, 1989; Huang et al., 2013) and it completes the divergence from Solanaceae species such as tomato and potato. These duplication events have provided neo-functionalization of important genes including in vitamin C, flavonoid, and carotenoid metabolic activities. It also has one of the well-known fleshy fruit since it is an excellent source of several vitamins, minerals, dietary fibers and other related health benefit dietary nutrients (Skinner et al., 2011). In addition to this, recent studies show that the consumption of kiwifruit has positive effects on cardiovascular health through antioxidant activity and by promoting gut microflora. It has also found that the kiwifruit support immune system either by up-regulating some defense-related genes or activating 'DNA-repair' mechanism in the cells (Skinner et al., 2011).

Taken together, we predicted some putative miRNAs in silico between the kiwifruit *A. chinensis* and its close relatives by using genomic sequence data and computational tools to gain a better understanding of their relations and to provide a valuable resource for the evolutionary processes in the Asterid lineage. We also showed that clustering of miRNAs repertoire between chosen organisms does not show the same pattern as the relationship between their whole genome.

Material and Methods

Reference miRNAs and Datasets: 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. For *A.chinensis*, genome sequence was retrieved from the website. *A.thaliana*, S. lycopersicum, *S.tuberosum* and *V.vinifera* masked Ensemble Plants website provided genomic sequences.

In silico miRNA Identification based on Sequence Homology and Secondary Structure Conservation: A two-step strategy was adopted based on the preliminary selection of database sequences with homology to a previously known plant mature miRNA and their subsequent retention assessing the consistency of their secondary structure with pre-established pre-miRNA features (Zhang et al., 2006). The prediction was employed using two previously developed, in-house Perl scripts: SUmirFind and SUmirFold, described in detail in the publications. In the first step of homology-based miRNA prediction, SUmirFind script, which utilizes BLAST+ stand-alone toolkit, version 2.2.31 (Camacho et al., 2009) was used for detection of database sequences with homology (mismatch cutoff parameter set to <=3) to previously known plant mature miRNAs (Zhang et al., 2006).In the second step, SUmirFold, a script that generates secondary structures through UNAFold version 3.8 was used with parameters optimized to include all possible stem-loops generated for each miRNA query. SUmirFold output was further processed to eliminate redundant hits, resulting from cases where identical miRNAs were predicted from two similar query mature miRNA sequences. Moreover, 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 were also manually removed. This process was done for all genomic datasets of plant species.

Clustering of plant species based on the variety of their cumulative miRNAs: Mature miRNA sequences from each plant species were separately listed and the binary matrix showing cumulative miRNA datasets in all five plant species was formed. For hierarchical clustering, euclidian distance based centroid clustering was adopted. Distance matrix construction, clustering, and dendrogram generation were performed in the MINITAB program.

Target annotation of predicted miRNAs: Mature sequences were collected from each species and duplicates were removed. By using an online web tool, psRNA, an analysis was performed, and hit sequences were retrieved. These retrieved sequences were used as input data for the Blast2Go online web tool (Conesa and Götz, 2009). Target annotation charts were created for all plant species. Experimentally validated targets of predicted putative miRNAs were also detected in miRBase (Kozamara and Jones, 2013).

The copy number of miRNA genes and expression analysis of predicted miRNAs: Repeated same miRNAs that resulted from the similar query miRNA stem-loop sequences were eliminated to avoid over-representation. The EST-database was formed separately as a specific to each plant species, and the restricted criteria were used for the analysis as the only miRNA families who had hits above the threshold as 98% identity, and 99% query coverage were retrieved.

Results and Discussions

Putative predicted miRNAs and their distribution across plant species: We identified putative miRNAs from each organism (Figure 1).

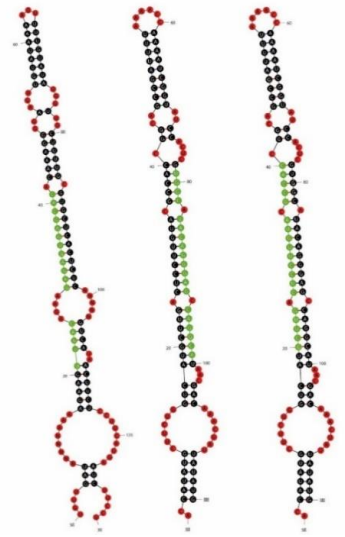


Figure 1. Predicted pre-miRNA stem-loop structures of selected miRNAs from *A.thaliana*. Mature miRNAs start and end points are shown by green color.

For each organism, average sequence length with a median for identified pre-miRNAs, mature miRNAs, average GC % content for pre-miRNAs with median and min-max values were calculated, and they are shown in Figure 2. Minimal folding free energy index (MFEI), which is calculated from the minimum folding free energy (MFE), sequence length and %GC content of the pre-miRNA, differentiates 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). The low negative MFE values show the higher stability of the predicted miRNA (Zhang et al., 2007; Jin et al., 2008).

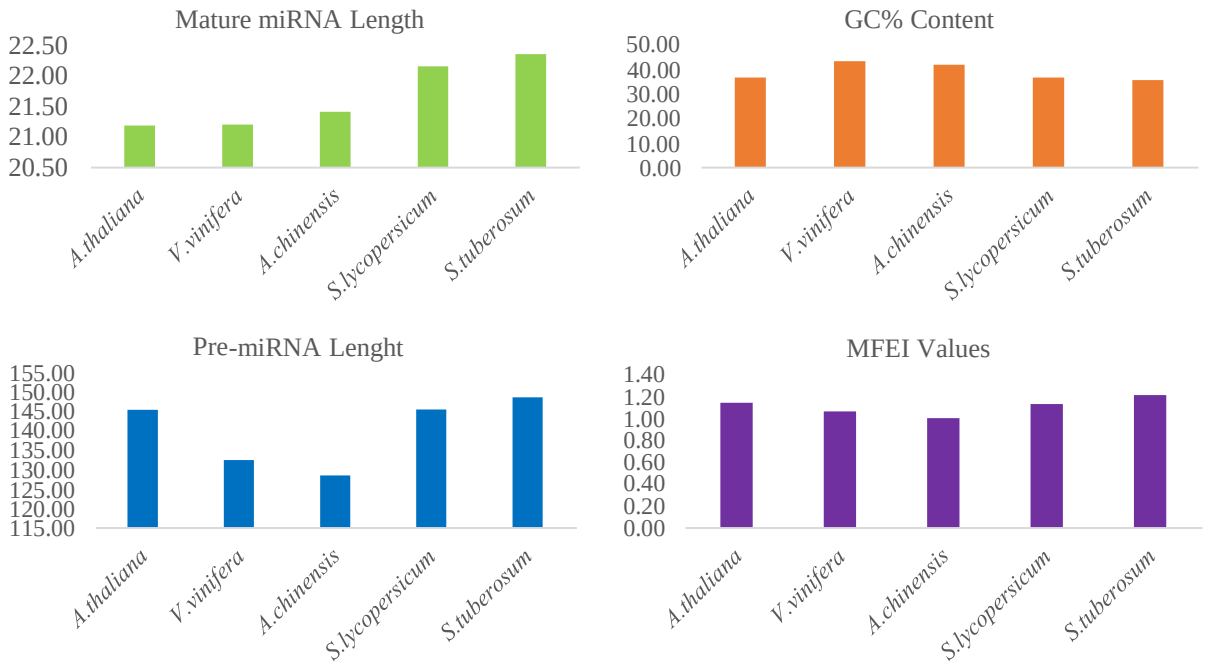


Figure 2. MFE and MFEI tables with minimum, maximum and median values for unique pre-miRNA sequences.

Prediction of putative miRNAs from genomes of different plant species enabled us to compare the variety of miRNAs across species. In the *A.thaliana* genome, we predicted as 93 putative miRNAs; in *S.lycopersicum* genome 57 putative miRNAs were predicted and analyzed. *S.tuberosum* had 61 putative miRNAs to be analyzed whereas, in *V.vinifera* genome, we predicted 43 putative miRNAs (Figure 3). In our previous study, we found 52 putative miRNAs in *A.chinensis* genome during the last study (Avsar and Esmaeili, 2015). According to our results, *A.thaliana* had the highest number of putative miRNAs in its genome whereas *V.vinifera* had the lowest number of miRNAs. Although the genome size of V. vinifera (~500Mbp) is larger than the *A.thaliana* (~135Mbp), the number of predicted putative miRNAs did not show the similar correlation. This may be caused by different evolutionary biogenesis characteristics of miRNAs across plant species.

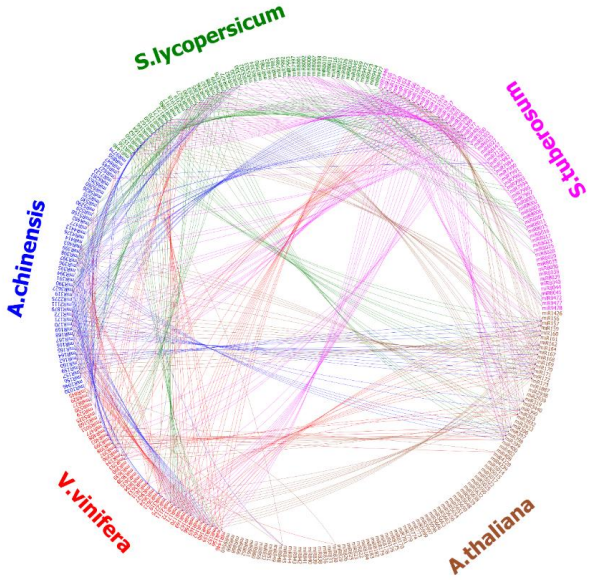


Figure 3. All miRNAs predicted from five organisms (Avsar et. al., 2016).
Link: http://APresenter.com/view.faces?id=2055757656

Between all those five plant species, some miRNA families were found commonly: miR156, miR157, miR160, miR162, miR167, miR169, miR170, miR171, miR172, miR319, miR390, miR396, miR398, miR399. Those common miRNAs are probably not "species-specific" type of miRNAs but have some critical regulatory roles including development, growth, modulation of auxin-response. All common miRNAs between organisms are shown in Figure 4.

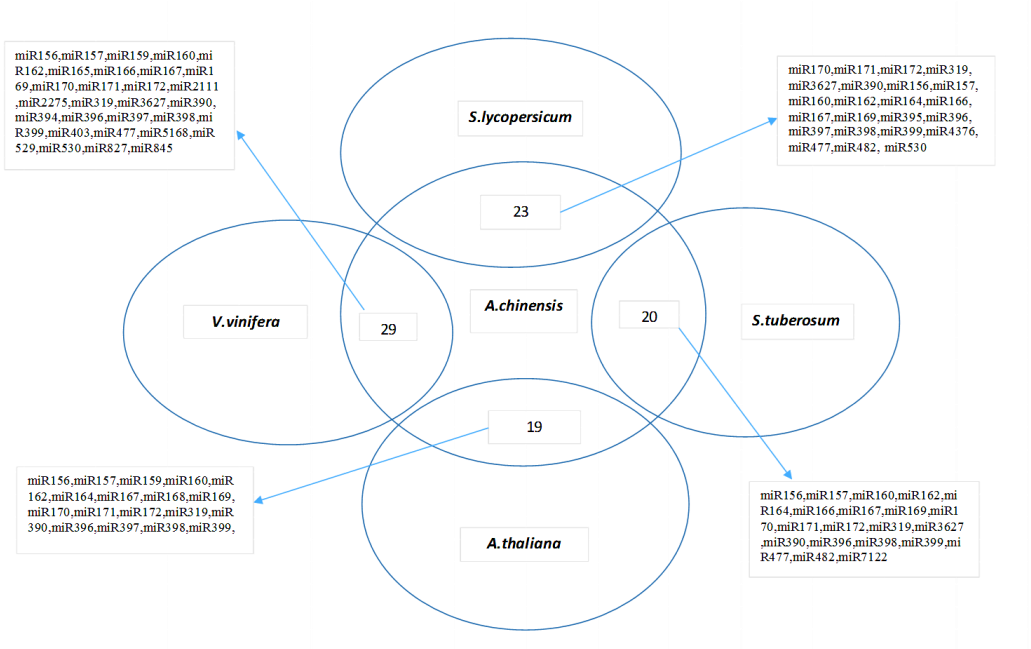


Figure 4. Common miRNAs between all organisms.

Clustering of putative miRNAs: The clustering of miRNA families of plant species showed that the higher similarity between *S.lycopersicum-S.tuberosum* than *V.vinifera-A.chinensis* (Figure 5). Huang et al. analyzed some other sequenced plant genomes to compare with the kiwifruit and they showed that the tomato had the closest evolutionary relationship (Huang et al., 2013) however regarding miRNA families clustering, grapevine showed more similarity to kiwifruit genome. These findings might be used for further evolutionary studies.

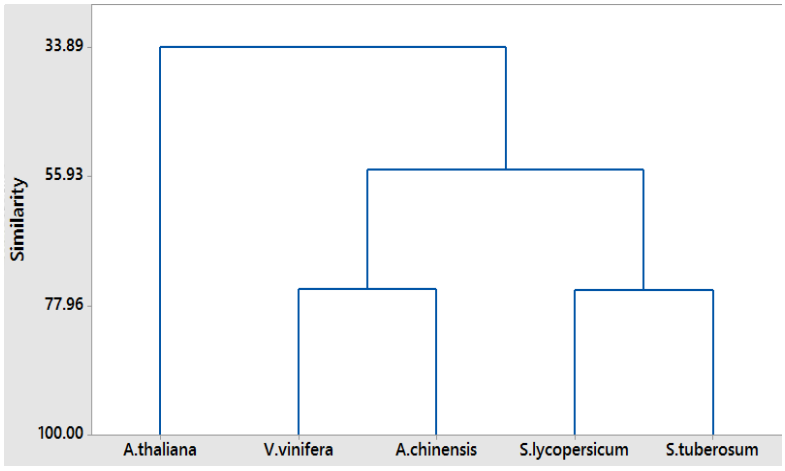


Figure 5. Clustering analysis for all species.

Targets of putative miRNAs and their annotations: Duplicate mature miRNA sequences were removed and the remaining mature miRNA sequenced were used for target annotations. These annotations were conducted to identify the miRNA targets in biological processes, molecular pathways, and cellular components.Based on the biological processes results, targets of *S.tuberosum* and *S.lycopersicum* putative miRNAs showed similarity as they have roles in metabolic processes, single-organism processes, response to stimulus and cellular processes. However, *A.chinensis* and *V.vinifera* had the similar functions mentioned above in addition to the cellular component organization and localization. *A.thaliana* miRNA targets were not identical to any organism since they had roles in different pathways (Figure 6).

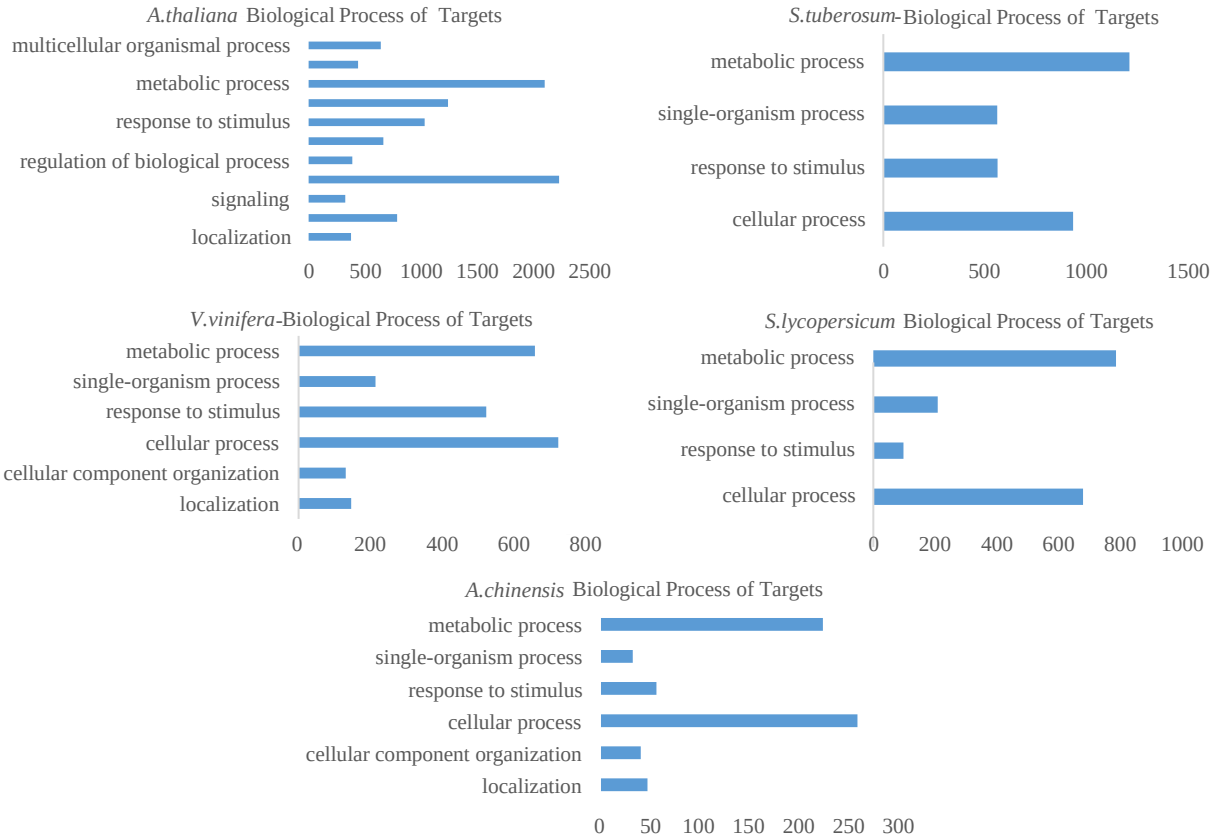
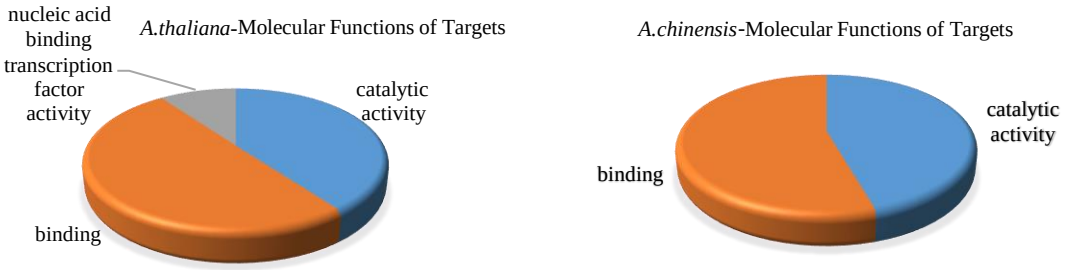


Figure 6. Blast2Go target annotations chart based on biological processes.

For the molecular functions of targets, almost all the organisms showed the similar results except for *A.thalina*. Putative miRNA targets are mostly found in catalytic activity and binding processes. In *A.thaliana*, additionally, nucleic acid binding transcription factor was found. In *A.chinensis*, putative miRNA targets were primarily found in binding activities (Figure 7).



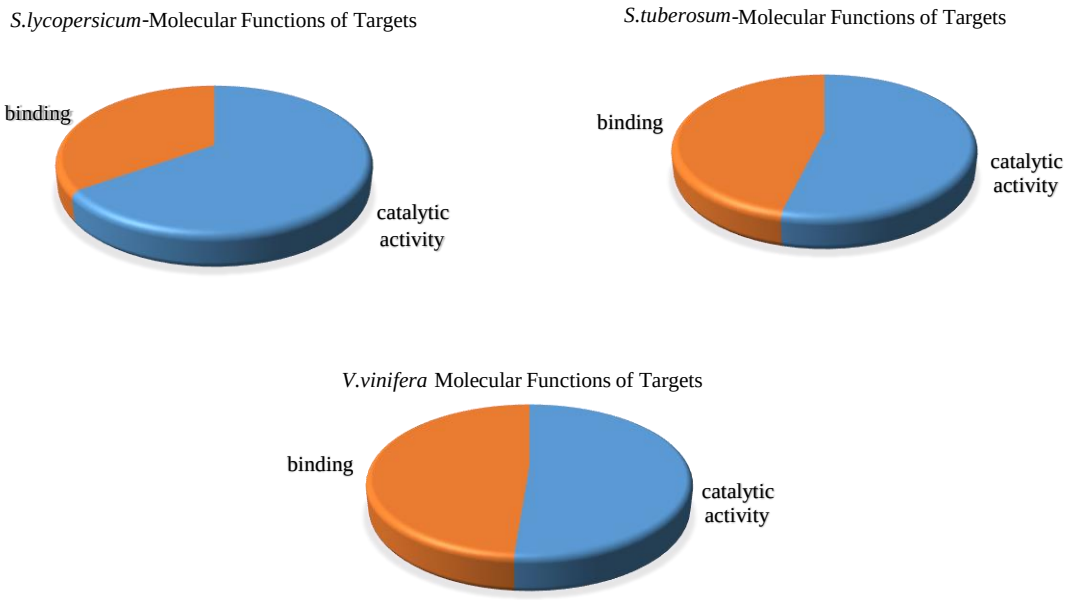


Figure 7. Blast2Go target annotations chart based on their molecular functions.

Representative miRNA families and the expression analysis of predicted miRNAs: We cannot determine the absolute copy number of each miRNA families by indeed since some miRNA genes may be covered by more than one while others may not be covered at all. However, the representation of each miRNA within the dataset provides us to estimate its prevalence on the genome. According to those results, miR169 and miR156 families were highly represented in the *A.thaliana* genome. In *S.lycopersicum*, miR5303 families had the highest representative whereas in *S.tuberosum* miR7984 families were represented mostly. miR3631 families were the most represented in *V.vinifera* genome however in *A.chinensis*, miR156 families and miR171 families were highly represented.

1,527,298 and 300,445 EST sequences were retrieved from NCBI database for *A.thaliana* and *S.lycopersicum*, respectively. 250,136 and 460,107 EST sequences were retrieved for *S.tuberosum* and *V.vinifera*, sequentially. Finally, 132,583 *A.chinensis* EST sequences were obtained from NCBI. All EST databases were built separately and analyzed in silico. According to those results, the highest number of expressed miRNA families (miR160, miR167, miR171, miR398, miR5654, miR5656, miR8170, miR826, miR830, miR840, miR860, miR869) was found in *A.thaliana* genome whereas the least number of expressed miRNA families was found in *A.chinensis* (miR535). This may result from the number of EST sequences in the database that we used since *A.thaliana* had the highest EST sequences and *A.chinensis* had the lowest EST sequences in their databases. All other expressed miRNA families for all organisms are shown in Table 1.

Table 1. *In silico* expression analysis on all plant species.

A.thaliana	miR160, miR167, miR171, miR398, miR5654, miR5656, miR8170, miR826, miR830, miR840, miR860, miR869
A.chinensis	miR535
S.lycopersicum	miR157, miR156, miR172, miR482, miR5303, miR5304, miR7983, miR7997
S.tuberosum	miR6027, miR156, miR172, miR396
V.vinifera	miR3623, miR172, miR3633, miR3634, miR396, miR397, miR827

Conclusions

In this study, *A.thalinana*, *V.vinifera*, *S.lycopersicum* and *S.tuberosum* were used for the comparison of miRNA repertoire since they were also performed for better understanding the genome of *A.chinensis* in draft genome sequencing article (Huang et al., 2013). *A.chinensis* is the first sequenced genome in the order of Ericales, so our study will be a good source for comparative genomics and evolutionary studies in the asterid lineage. It has also substantial duplication events after the divergence from tomato and potato and the genome structure might be affected by gaining some additional properties. In our study, we predicted that *A.chinensis* harbored some putative miRNAs that were not found in *S.tuberosum* or *S.lycopersicum* and these different putative miRNAs might result from the divergence events. Kiwifruit also has a rich source of vitamin C and other nutritional compounds so those identified miRNAs might have some regulatory roles on the pathways. In other words, kiwifruit genome will be an invaluable source of better understanding genome evolution and improvement of some agronomical properties in other plants including nutrient metabolism, disease resistant, sex determination and polyploidy events and the findings in the presented study will help to research community to elucidate the roles of miRNAs in these critical pathways.

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., Aliabadi E.E., Yousefnezhad R. (2016). Academic presenter: a new storytelling presentation software for academic purposes. *arXiv preprint*. 1-9.

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.

Choudhuri, S. (2009). Lesser known relatives of miRNA. *Biochemical and biophysical research communications*, 388(2): 177-180.

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

Egan, A. N., Schlueter, J., Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. *American journal of botany*, 99(2): 175-185.

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., Liu, Y. (2013). Draft genome of the kiwifruit Actinidia chinensis. *Nature communications*, 4.

Jin, W., Li, N., Zhang, B., Wu, F., Li, W., Guo, A., Deng, Z. (2008). Identification and verification of microRNA in wheat (Triticum aestivum).*Journal of plant research*, 121(3): 351-355.

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

Liu, Q., Chen, Y. Q. (2010). A new mechanism in plant engineering: the potential roles of microRNAs in molecular breeding for crop improvement.*Biotechnology advances*, 28(3): 301-307.

McNeilage, M. A., Considine, J. A. (1989). Chromosome studies in some Actinidia taxa and implications for breeding. *New Zealand journal of botany*,27(1): 71-81.

Skinner, M. A., Loh, J., Hunter, D. C., Zhang, J. (2011). Gold kiwifruit (Actinidia chinensis 'Hort16A') for immune support. *Proceedings of the nutrition society*, 70(02): 276-280.

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.

Zhang, B., Wang, Q., Pan, X. (2007). MicroRNAs and their regulatory roles in animals and plants. *Journal of cellular physiology*, 210(2): 279-289.

Zhang, B., Wang, Q. (2015). MicroRNA-Based Biotechnology for Plant Improvement. *Journal of cellular physiology*, 230(1): 1-15.