Putative microRNA Analysis of the Kiwifruit *Actinidia chinensis* through Genomic Data

Bihter Avsar and Danial Esmaeili Aliabadi

Sabanci University, Faculty of Enginnering and Natural Sciences, Istanbul, Turkey

E-mail: bihteravsar@sabanciuniv.edu, danialesm@sabanciuniv.edu

Abstract—**MicroRNAs are important regulators in the cells that are well defined in various roles. With the advent of new generation sequencing technologies, identification of miRNAs studies increase rapidly. In here, we identified 58 putative miRNAs through kiwifruit genome by using *in silico* methods. The computational analysis was done through genome and transcriptome data of Chinese kiwifruit cultivar 'Hongyang' which has important properties including the high content of vitamin C, carotenoids and flavonoids. Since kiwifruit shares some portion of its genes with diverse plant families, this study may contribute to the further biotechnological studies in other close relatives.**

Index Terms—**Kiwifruit, *Actinidia chinensis*, miRNA, microRNA, transcriptome, genome.**

I. INTRODUCTION

The kiwifruit or Chinese gooseberry is the edible berry member in the Actinidia genus [1]. Current research shows that Actinidia genus contains 54 species and 75 taxa [2] that they are known as perennial, deciduous and dioecious plants. Genomic analysis reveals that the kiwifruit species have often polyploidy structure with a chromosome number as x=29 is resulted from hexaploidization and two more recent whole genome duplication events [3], [4]. These duplication events have provided neofunctionalization of important genes including in vitamin C, flavonoid and carotenoid metabolic activities.

The kiwifruit is mostly grown in China and it is one of the native horticultural crops [4]. It also belongs to the order of Ericales in the Asterid lineage and it completes the divergence from Solanacea species such as tomato and potato [4]. It has also one of the well-known fleshy fruit since it is a good source of several vitamins, minerals, dietary fibres and other related health benefit dietary nutrients [5]. 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 [5]. It has also found that the kiwifruit support immune system either by up-regulating some related genes or activating 'DNA-repair' mechanism in the cells [5]. Taken together, the

researchers highly recommend to consume kiwifruit although it may be an allergen for some individuals because of actinidain content in it [6]-[8]. In here, we analyzed some putative miRNAs by using genome and transcriptome data of kiwifruit. The heterozygous kiwifruit cultivar 'HongYang' was sequenced and analyzed by Huang *et al.* [4] to elucidate agronomical properties much better and to provide a valuable resource for the evolutionary processes in the Asterid lineage [4].

II. MATERIALS AND METHODS

A. Reference miRNAs

Currently available mature miRNA sequences (8,496 sequences; 73 plant species) were downloaded from miRBase release 21 [9]. This corresponded to 4,802 unique mature miRNA sequences, those were used as a query in homology-based *in silico* miRNA identification.

B. Kiwifruit Dataset

Both genome and transcriptome datasets were used for miRNA analysis in kiwifruit. Genome sequence was retrieved from http://bioinfo.bti.cornell.edu/cgi-bin/kiwi/home.cgi web site. Transcriptome data was retrieved from NCBI databank (Hongyang transcriptome: SRR926770). Transcriptome data was used as raw reads and the reads were assembled by Trinity software (http://trinityrnaseq.github.io/).

C. In Silico miRNA Identification Based on Sequence Similarity 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 [10]-[12]. Prediction was employed using two previously developed, in-house Perl scripts: SUmirFind and SUmirFold, described in detail in the publications [10]. In the first step of homology-based miRNA prediction, SUmirFind script, which utilizes BLAST+ stand-alone toolkit, version 2.2.25 [13] was used for detection of database sequences with homology (mismatch cutoff parameter set to <3) to previously known plant mature miRNAs [12]. 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 both genomic dataset and transcriptome dataset seperately.

D. Representation of Genomic miRNAs

Repeated identical miRNAs that were resulted from the similar query miRNA stem loop sequences were eliminated to avoid over-representation. The hits below 10 was not included in the analysis.

E. Expression Analysis of Predicted miRNAs

For genomic data of *Actinidia chinensis*, the pre-miRNA sequences were retrieved and the duplicate sequences were removed to prevent over-representation. By using BLAST+ stand-alone toolkit, version 2.2.25 [13], pre-miRNA sequences were blasted to kiwifruit-related EST (Expressed Sequence Tag) sequences downloaded from NCBI database. 57,757 EST sequences were downloaded and the stricted criteria was used for the analysis as the only miRNA families who had hits above the threshold as 98% identity and 99% query coverage were retrieved.

F. Target Annotation of Predicted miRNAs

Mature sequences were collected from each species and duplicate ones were removed and were used as query sequences. The built-in-database was created for expression analysis was also used for annotation analysis. By using online web tool, psRNA (http://plantgrn.noble.org/psRNATarget/target), analysis was performed and hit sequences were retrieved. These retrieved sequences were input data and used for Blast2 Go online web tool (http://www.blast2go.com/b2ghome). Target annotation charts were created for kiwifruit based on its genomic and transcriptomic miRNAs together. For predicted mature miRNA sequences, known targets of homologous miRNAs in other plants were also detected in miRBase [8] website to confirm their experimentally validated targets.

III. RESULTS AND DISCUSSIONS

A. Putative Predicted miRNAs through Transcriptome and Genome Data of Actinidia Chinensis

We predicted 58 putative miRNAs totally. Out of 58 miRNAs, 12 miRNAs were found in Hongyang transcriptome data whereas 52 miRNAs were predicted in Hongyang genome *in silico* (Fig. 1). Mature miRNA start and end points are showed designated by arrows. Structures are predicted using UNAFold program - an implementation of *Zuker algorithm*. 6 miRNA families (miR159, miR164, miR319, miR403, miR414, miR482) were predicted as common miRNAs between transcriptome and genome data.
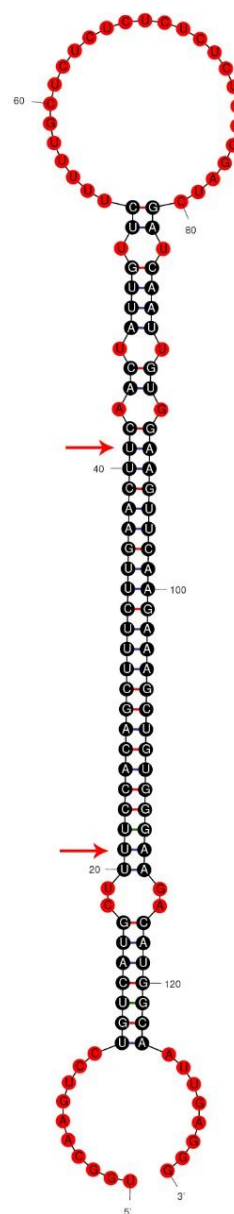

Figure 1.   Identified pre-miRNA stem loop structure of selected miRNA on kiwifruit genome.

B. Representation miRNAs in Actinidia Chinensis Genome

The representation analysis was just conducted for genomic data in order to show the miRNA genes clustering through the genome. Below 10 hits were not included in the analysis because they might be contamination or 'young-miRNAs'. On the other hand, the highest number of hits might be repetitive elements. because most of the transposable elements were domesticated into microRNA genes [14]. In total, 31 miRNAs representation is shown for kiwifruit organism (Fig. 2). According to the results, miR156 families was highly represented in the genome. The recent studies show that the miR156 families are also important regulators in flowering timing and developmental processes.

C. *Expression Analysis of miRNAs through Kiwifruit Genomic Data*

Expression analysis was performed by using EST sequences from NCBI databank and we created a database. The mature miRNAs sequences were blasted against to this database and the only miRNA families who had hits above the threshold as 98% identity and 99% query coverage were retrieved. According to the results, miR535 families was the only one that had the criteria mentioned above and it is also represented in the genome. miR162 families had expression profile with 75% coverage and is also represented in the genome since it has enough hits above the threshold.

D. *Target Prediction and Gene Ontology Analysis*

Potential targets were predicted from miRBase [9]. All putative miRNAs experimentally validated targets in genomes and transcriptomes were searched through the database. According to the results, miR156, miR157, miR159, miR160, miR162, miR164, miR165, miR166, miR167, miR168, miR169, miR170, miR171, miR172, miR319, miR394, miR395, miR396, miR397, miR398 and miR399 had experimentally validated targets (Table I). Most of these targets were transcription factors, promoter-binding proteins, F-box proteins, ATP sulphurylases, copper-superoxide dismutases and phosphatase transporters. For GO (Gene ontology) analysis Blast2Go online web tool (http://www.blast2go.com/b2ghome) was used. Both genomic and transcriptomic miRNAs' targets sequences were uploaded to the web site (Fig. 3). According to the charts, the predicted miRNAs target mostly metabolic and cellular processes. Catalytic activities and binding processes have higher proportions as molecular functionality of predicted putative miRNAs. They mostly act in the cells and organelles as cellular compartments.
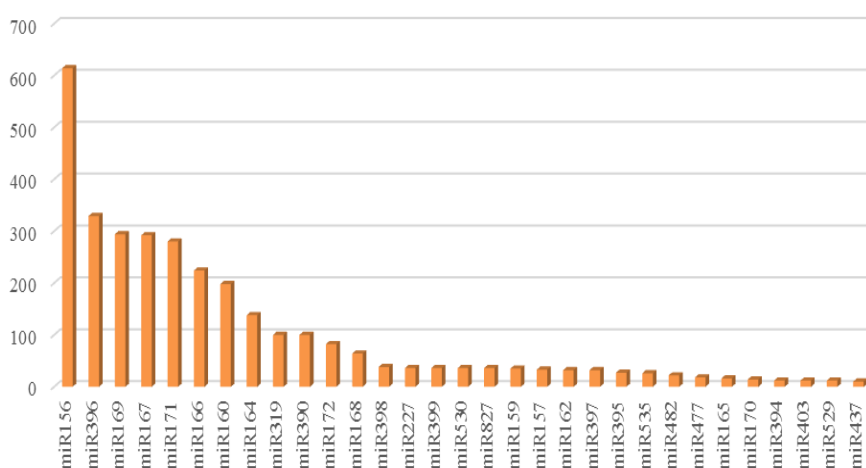

Figure 2.   Representation of potential miRNAs on kiwifruit genome (The miRNAs gave above 10 hits are shown in the graph).
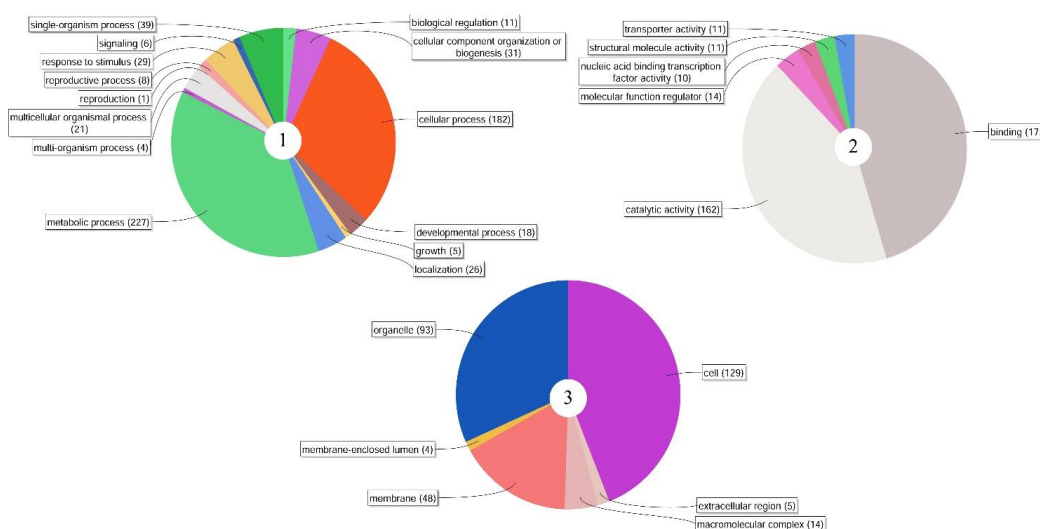

Figure 3.   Target annotation charts of kiwifruit based on GO analysis are numbered as biological processes, molecular function and cellular component respectively.

TABLE I.   EXPERIMENTALLY VERIFIED TARGET PROTEINS OF MICRORNAS

miRNA name	Targeted Protein
miR156	Squamosa-promoter Binding Protein (SBP) box
miR157	Squamosa-promoter Binding Protein (SBP) box
miR159	Binding to the promoter of the floral meristem identity gene LEAFY
miR160	Auxin response factor proteins
miR162	DICER-LIKE 1 (DL1) proteins
miR164	NAC domain transcription factors
miR165	HD-Zip transcription factors including Phabulosa (PHB) and Phavoluta (PHV)
miR166	HD-Zip transcription factors including Phabulosa (PHB) and Phavoluta (PHV)
miR167	Auxin Response Factors (ARF transcription factors)
miR168	ARGONAUTE protein.
miR169	CCAAT binding factor (CBF)-HAP2-like proteins
miR170	SCARECROW-like proteins
miR171	SCARECROW-like proteins
miR172	APETALA2-like transcription factors
miR319	TCP genes for cleavage
miR394	F-box proteins
miR395	ATP sulphurylases
miR396	Growth Regulating Factor (GRF) transcription factors, rhodanese-like proteins, and kinesin-like protein B
miR397	Laccases and beta-6 tubulin
miR398	Copper superoxide dismutases and cytochrome C oxidase subunit V
miR399	Phosphatase transporter

IV.   CONCLUSIONS

MicroRNAs(miRNAs) are class of short(21-24nt) non-coding, single stranded small RNAs that are highly conserved across plant species. Plant miRNAs have enormous roles in diverse biological processes including growth, development, abiotic and biotic stress tolerance. They act on the gene expression either by cleaving the specific target or translational repression. On the other hand, the advent of sequencing technologies provide us better understanding of the genes, genome, metabolism and plant mechanisms. With this new technology, computational tools and *in silico* analysis approach help researchers to eliminate redundant data in the further studies. Since small RNA regulators, miRNAs, are critical players in plant mechanisms, we wanted to identify them by homology-conservation method. So in the study, we predicted 58 putative miRNAs and their known targets by using computational methods via genome and transcriptome sequence information of kiwifruit.

These findings are important to enlighten further studies on *Actinidia chinensis* since the kiwifruit is becoming one of the important fleshy fruit in terms of its high vitamin C and mineral content, high antioxidant capacity, its ploidy structure, its evolution and sex determination. It is also one of the first sequenced organism in the *Actinidiaceae* family so our findings may be useful for better understanding of the other plants in the same family or even in the the the order of *Ericales*.

REFERENCES

[1]   B. C. Strik and H. Cahn, *Growing Kiwifruit*, [Covallis, Or.]: Oregon State University Extension Service, 1998.

[2]   L. Zhang, Z. Li, Y. Wang, Z. Jiang, S. Wang, and H. Huang, "Vitamin C, flower color and ploidy variation of hybrids from a ploidy-unbalanced Actinidia interspecific cross and SSR characterization," *Euphytica.*, vol. 175, no. 1, pp. 133-143, 2010.

[3]   M. A. McNeilage and J. A. Considine, "Chromosome studies in some Actinidia taxa and implications for breeding," *New Zealand Journal of Botany*, vol. 27, no. 1, pp. 71-81, 1989.

[4]   S. Huang, J. Ding, D. Deng, W. Tang, H. Sun, D. Liu, *et al.*, "Draft genome of the kiwifruit Actinidia chinensis," *Nature communications*, vol. 4, 2013.

[5]   M. A. S kinner, J. Loh, D. C. Hunter, and J. Zhang, "Gold kiwifruit (*Actinidia chinensis* 'Hort16A') for immune support," *Proceedings of the Nutrition Society*, vol. 70, no. 2, pp. 276-280, 2011.

[6]   J. S. Lucas, S. A. Lewis, and J. O. B. Hourihane, "Kiwi fruit allergy: A review," *Pediatric allergy and immunology*, vol. 14, no. 6, pp.420-428, 2003.

[7]   A. Alemán, J. Sastre, S. Quirce, M. De las Heras, J. Carnés, E. Fernández-Caldas, *et al.*, "Allergy to kiwi: A double-blind, placebo-controlled food challenge study in patients from a birch-free area," *Journal of Allergy and Clinical Immunology*, vol. 113, no. 3, pp. 543-550, 2004.

[8]   T. M. Le, M. Bublin, H. Breiteneder, M. Fernández-Rivas, R. Asero, B. Ballmer-Weber, *et al.*, "Kiwifruit allergy across Europe: Clinical manifestation and Ig E recognition patterns to kiwifruit allergens," *Journal of Allergy and Clinical Immunology*, vol. 131, no. 1, pp. 164-171, 2013.

[9]   A. Kozomara and S. Griffiths-Jones, "MiRBase: Annotating high confidence microRNAs using deep sequencing data," *Nucleic acids research*, vol. 10, pp. 1093, 2013.

[10]   M. Kantar, T. Unver, and H. Budak, "Regulation of barley miRNAs upon dehydration stress correlated with target gene expression," *Functional & Integrative Genomics*, vol. 10, no. 4, pp. 493-507, 2010.

[11]   M. Kantar, S. J. Lucas, and H. Budak, "MiRNA: Expression patterns of Triticum dicoccoides in response to shock drought stress," *Planta*, vol. 233, no. 3, pp. 471-484, 2011.

[12]   B. Zhang, X. Pan, C. H. Cannon, G. P. Cobb, and T. A. Anderson, "Conservation and divergence of plant microRNA genes," *The Plant Journal*, vol. 46, no. 2, pp. 243-259, 2006.

[13]   C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, *et al.*, "BLAST+: Architecture and applications," *BMC Bioinformatics*, vol. 10, no. 1, pp. 421, 2009.

[14]   Y. Li, C. Li, J. Xia, and Y. Jin, "Domestication of transposable elements into microRNA genes in plants," *Plos. one*, vol. 6, no. 5, pp. 192-212, 2011.