Plant Omics Journal POJ

POJ 10(04):190-196 (2017) ISSN:1836-3644
doi: 10.21475/poj.10.04.17.pne707

Identification of microRNA elements from genomic data of European hazelnut (*Corylus avellana* L.) and its close relatives

Bihter Avsar[1*] and Danial Esmaeili Aliabadi[1]

Sabanci University, Faculty of Engineering and Natural Sciences, Istanbul, TURKEY

***Corresponding author: bihteravsar@sabanciuniv.edu**

Abstract

Plant microRNAs (miRNAs) are small and non-coding endogenous RNAs which have numerous regulatory roles in cells. These critical players regulate pathways either by inducing translational repression or messenger RNA (mRNA) decay. Newly developed bioinformatics tools and computational methods have been increased to identify miRNAs with their targets inside the genome. In this study, we predicted and identified 57 putative miRNAs through *Corylus avellana* (*C. avellana*) genomic data *in silico*. We also predicted some other putative miRNAs from *Arabidopsis thaliana* (*A. thaliana*), *Ricinus communis* (*R. communis*), *Populus trichocarpa* (*P. trichocarpa*) and *Vitis vinifera* (*V. vinifera*) to compare with the *C. avellana* organism since previous studies have indicated high similarities between these genomes and proteome atlases. The miRBase 21 was used as a reference dataset and the putative miRNAs were identified for the genome of each organism. We used homology conserved method to identify putative miRNAs. Based on our findings, *C. avellana* miRNA content was found to be highly similar to *V. vinifera, R. communis* and *P. trichocarpa*. Also, we found the targets of these hazelnut putative miRNAs and their possible functions inside the cell. One of our major discoveries is that miR171 families are highly represented (the copy number of miRNA) in the hazelnut genome to provide clues for microRNA domestication. The miR396, miR482, and miR2118 families were found as *in silico* expressed putative miRNAs by using computational methods. All these findings may help us better understanding the miRNA repertoire of the hazelnut organism and provide valuable insight about the regulatory roles of predicted putative miRNAs which are shared with other organisms (*A. thaliana*, *R. communis*, *P. trichocarpa*, *V. vinifera*) for further studies.

Keywords: *Arabidopsis thaliana*; *Corylus avellana*; microRNA; *Ricinus communis*; *Populus trichocarpa*; *Vitis vinifera.*
Abbreviations: *A. thaliana_Arabidopsis thaliana*; bp_base pair_*C. avellana-Corylus avellana*; EFB_Eastern filbert blight; EST_Expressed sequence tag; MFEI_Minimal folding free-energy index; MFE_Minimal folding-free energy; miRNAs_microRNAs; mRNA_messenger RNA; NGS_Next generation sequencing; nt_nucleotide; *R. communis_Ricinus communis*; *V. vinifera_Vitis vinifera*.

Introduction

Although the black sea region of Turkey produces around 70% of the global hazelnut supply, the USA is the largest producer of European hazelnut (i.e., *C. avellana*) (Gokirmak et al. 2009; Rowley, 2016). The consumer demand of hazelnuts, more specifically *C. avellana*, has been increased due to important agronomic properties and being known as the main content of butter, chocolate products, and various pastes (Rowley, 2016). In addition to these, their kernels and fibers are used in some foods and their shells are used for landscaping or groundcover (Rowley, 2016). *C. avellana* is monoecious, and it is diploid with 11 chromosomes ($2n = 2x = 22$). Its genome size is around 378Mbp, and it has a short life cycle. In many years, hazelnut producers and breeders have to deal with a fungal disease known as Eastern Filbert Blight (EFB) which causes a severe crop loss in susceptible cultivars. To address this problem, researchers found the resistance allele, Gasaway gene, to EFB disease and developed the cultivar called as "Jefferson (OSU703.007)"(Mehlenbacher, 2011). It was mentioned that diverse gene alleles coming from the wild relatives and progenitors of plants contribute the adaptive processes such as abiotic and biotic stress environments (Akpinar et al. 2012). The genomic DNA of Jefferson cultivar were sequenced using high-throughput next generation sequencing (NGS) technologies and analyzed by bioinformatics tools, which is led to develop novel genomic tools in recent years. Thanks to the sequenced genome of Jefferson, newly developed genomics tools will become available; therefore, the variants can be identified among other cultivars, and molecular markers related to important agronomical properties can be developed (Rowley, 2016). Sequence comparisons and gene ontology analysis showed that hazelnut proteins have a high similarity with grape (i.e., *V. vinifera*), poplar (i.e., *P. trichocarpa*) and castor bean (i.e., *R. communis*) proteins (Rowley, 2012). This data will be of importance to breeders in their future breeding efforts.

MicroRNAs (miRNAs) are non-coding small RNAs (18-24nt) with important roles ranging from development to disease resistance. These small but effective regulators are important elements in elucidating gene regulation pathways in the cell (Tang, 2010; Rogers and Chen, 2013). To date, many studies have focused on the identification of miRNAs methods including cloning, genetic screening, microarray profiling and computational mining approaches (Zhang, 2006). While those experimental methods are expensive and time-consuming, miRNA data mining methods are rapid and cost effective. Thus, with the help of NGS technologies, computational miRNA identification methods have been improved and used practically. Although these tools enable us to unravel the whole miRNA repertoire of an organism,

the expression profile has yet to be shown either through experimental methods or by sequencing the whole transcriptome of that organism. In summary, *in silico* miRNA prediction methods help us to understand the mechanism of miRNA regulators inside the cell, miRNA-target interaction, and miRNA-dependent phenotypical differences between the organisms and their evolutionary pathways. In this study, we use computational methods to analyze the miRNA repertoire of *C. avellana,* and herein, the contributions of this study can be enumerated as follows:

- MicroRNA profiling for the European hazelnut (*C. avellana*) is important since it has important agronomical properties.
- By using the homology-based conservation method, we found putative miRNA families both from *C. avellana* genomic data and its close relatives including *V. vinifera*, *P. trichocarpa*, *A. thaliana* and *R. communis* genomic data. These findings provide us to gain a better understanding of their relationship and to compare them with each other. Similarly, we also used transcriptome assembly to identify expressed miRNAs *in silico*.
- We discovered that the most common putative miRNAs are between *V. vinifera* and *C. avellana*, whereas the least common miRNAs were found between *A. thaliana* and *C. avellana.*
- Our findings may provide insight into understanding the role(s) of miRNAs and their targets on some regulatory mechanisms. Conserved and non-conserved miRNAs between the *C. avellana* and the other organisms mentioned above provide us invaluable clues about their relationship. These results may help the research community in further studies.

Results and Discussions

Putative miRNAs from C.avellana genomic assembly and its close relatives

We identified the putative miRNA families from *C. avellana*, *V. vinifera*, *A. thaliana*, *R. communis*, and *P. trichocarpa* by using homology-based conservation approach (Avsar and Aliabadi, 2015; Avsar and Aliabadi, 2017). After using UNAFold, an implementation of the Zuker folding algorithm, we predicted putative miRNA families in organisms from their pre-miRNA stem-loop structures (see Fig 1; Data S1,S2,S3,S4, and S5). 57 and 93 putative miRNA families were identified in *C. avellana* and *A. thaliana,* respectively, whereas 40, 89, 43 putative miRNA families were predicted in *R. communis*, *P. trichocarpa,* and *V. vinifera* genomes, respectively (see Fig 2). All those four organisms (*A. thaliana*, *P. trichocarpa*, *R. communis* and *V. vinifera*) had some common miRNAs between *C. avellana* including miR156, miR157, miR159, miR160, miR167, miR169, miR170, miR171, miR172, miR319, miR393, miR396, miR398, and miR399 families. Other common putative miRNA families are also shown in Fig 3, separately. Based on these results, *V. vinifera* has more putative miRNAs in common with *C. avellana,* and the genomes of *P. trichocarpa* and *R. communis* also contain putative miRNAs highly similar to those of *C. avellana*. *A. thaliana* miRNA content is found less similar to *C. avellana* although its number of predicted putative miRNAs are higher than the rest of the organism. However, this may be related to the miRBase database used in homology-based approach which has more available miRNAs belong to *A. thaliana*.

Our findings and comparative analyses are consistent with the study of Rowley, 2016. According to this study, 82.5% percent of the annotated genes from *C.avellana* are presented in the closest genera as *V. vinifera*, *P. trichocarpa* and *R. communis*. We found the similar results for microRNA repertoire of *C. avellana* organism and its close relatives. Thus, the domesticated microRNA genes of European hazelnut have also high similarities with *V. vinifera*, *P. trichocarpa* and *R. communis* miRNA genes.

Whole miRNA repertoire of every organism and their common miRNA families were illustrated in Fig 4. *C.avellana* has more common putative miRNA families with *V. vinifera, P. trichocarpa*, and *R. communis* other than *A. thaliana*. Data in Figure 4, visualizes not only common miRNAs between *C. avellana* and the others but it also reveals the common/different miRNA families between *A. thaliana*, *V. vinifera*, *P. trichocarpa*, and *R. communis*. Based on this visualization, the most common putative miRNAs are shown between *R. communis–P. trichocarpa* (32 putative miRNA families) whereas the least common putative miRNA families are shown between *A. thaliana –R. communis* (17 putative miRNA families). 19, 20, 27, and 27 putative miRNA families were identified as common between *A. thaliana–P. trichocarpa*, *A. thaliana–V. vinifera*, *R. communis–V. vinifera*, and *P. trichocarpa–V. vinifera*, respectively**.** The Figure 4 can be magnified via Academic Presenter software[1] (Avsar et al. 2016) in detail.

The Jefferson genome has some resistance gene(s) to fungal disease. Although no direct relationship has been found yet between microRNA and Jefferson's EFB resistance, it is still a question whether those putative microRNAs predicted in our study have key regulatory roles on this mechanism. Moreover, recent studies have pointed out that miRNA machinery is involved in regulating the plant defense system against some fungal pathogenesis (Gupta et al. 2014; Zhao et al. 2012). In one of the study, regulation of some miRNA families from *P. trichocarpa* including miR1448 , miR1450, miR156, miR159, miR160, miR164, miR166, miR168, miR172, miR319, miR398, miR408, and their targets in response to fungal infection were reported (Zhao et al. 2012). Zhao et al. studied miRNA-mediated regulation in the development of stem canker disease caused by the *Botryosphaeria dothidea* fungus in *P. trichocarpa* and 12 miRNA families mentioned above were found to be fungi-responsive miRNA families by using expression analysis. Since we also identified the same miRNA families in our study (miR156, miR159, miR160, miR164, miR166, miR172, miR319, miR398, and miR408), our findings may provide significant information about miRNA profiling and their targets in plants under different fungal attack. Thus, the up/down regulation of these potential candidate miRNAs should be checked by experimental methods, and defense mechanisms should be elucidated.

We also calculated mature miRNA lengths, pre-miRNA lengths, GC% content and MFEI values for each predicted organism in our study (see Data S1, S2, S3, S4, and S5). Between all these organisms, we found the maximum and minimum mature miRNA lengths as 24 bp and 19 bp, respectively. For pre-miRNA lengths, we calculated 420 bp and 88 bp as maximum and minimum values, respectively. GC% content was found 66.9 as maximum and 24.2 as minimum values. Minimal folding free-energy index (MFEI) is a measure that aids distinguishing miRNAs, with typically higher MFEIs (> 0.67), from other types of cellular ssRNAs for which MFEIs were previously characterized such as transfer RNAs (MFEI = 0.64), ribosomal RNAs (MFEI = 0.59), and mRNAs (MFEI $\in$[0.62 – 0.66]) (Schwab et al. 2005). In our study, MFEI values were found 2.2 and 0.72 as

[1] www.APresenter.com/view.faces?id=1486485760

Table 1. Experimentally validated target proteins of predicted miRNAs in miRBase.

miRNA Name	miRBase Target
miR156	Squamosa-promoter Binding Protein (SBP) box.
miR157	Squamosa-promoter Binding Protein (SBP) box
miR159	MYB and TCP transcription factors.
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) that regulate axillary meristem initiation and leaf development
miR166	HD-Zip transcription factors including Phabulosa (PHB) and Phavoluta (PHV) that regulate axillary meristem initiation and leaf development
miR167	Auxin Response Factors (ARF transcription factors)
miR169	CCAAT Binding Factor (CBF) and HAP2-like transcription factors.
miR170	GRAS domain or SCARECROW-like proteins, a family of transcription factors whose members have been implicated in radial patterning in roots, signaling by the phytohormone gibberellin, and light signaling
miR171	SCARECROW-like transcription factors.
miR172	APETALA2-like transcription factors.
miR319	TCP genes for cleavage
miR393	F-box proteins and bHLH transcription factors
miR394	F-box proteins and bHLH transcription factors
miR395	F-box proteins
miR396	Growth Regulating Factor (GRF) transcription factors, rhodenase-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

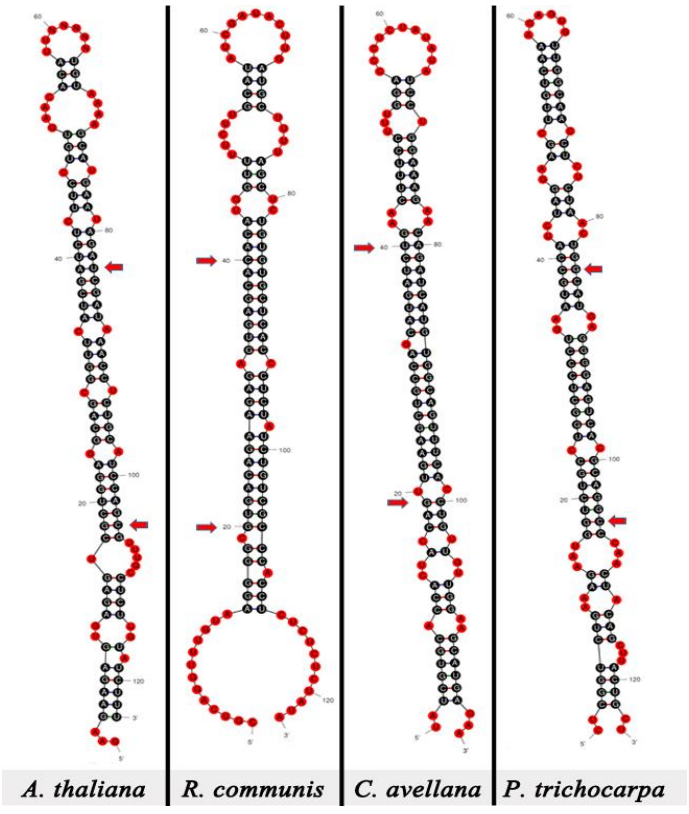

Fig 1. Predicted pre-miRNA stem-loop structures of selected miRNAs from *A. thaliana*, *R. communis*, *C. avellana* and *P. trichocarpa* by using UNAFold (an implementation of Zuker algorithm). Mature miRNAs start and end points are shown by red arrows.

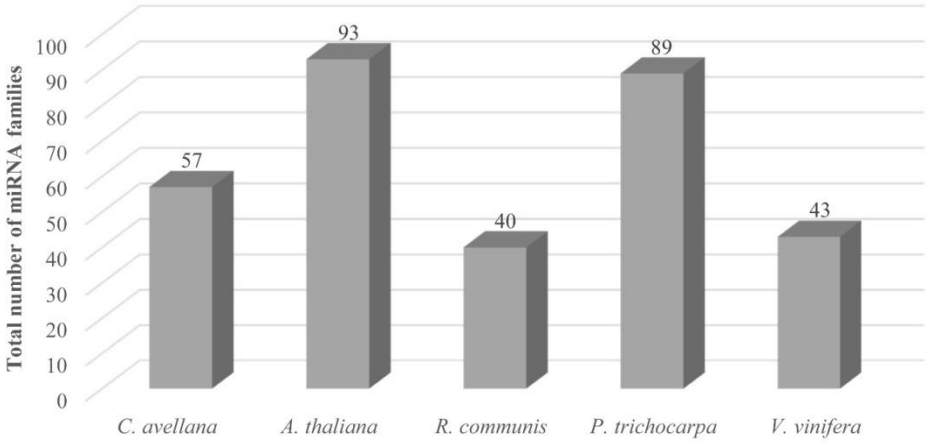

Fig 2. Total number of putative miRNA families distributed to each organism.

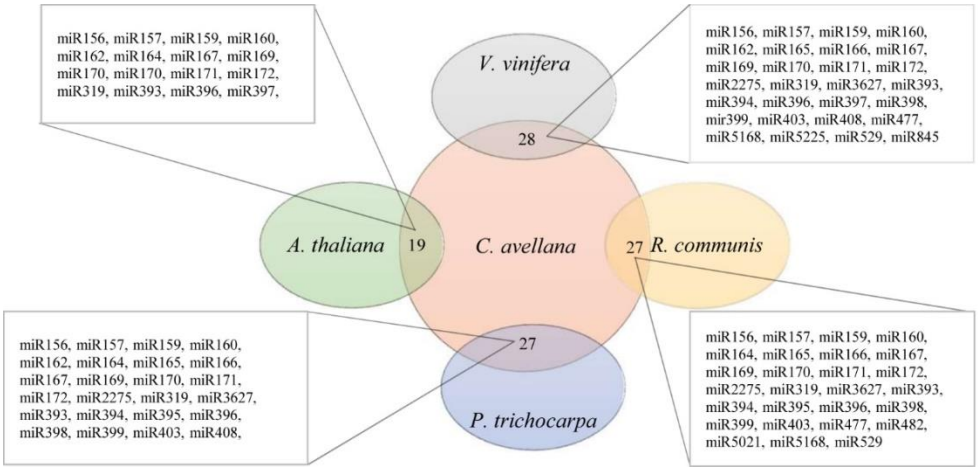

Fig 3. The number of conserved miRNA families between *C. avellana* and other organisms is shown. Those common miRNA families are represented in different colors.

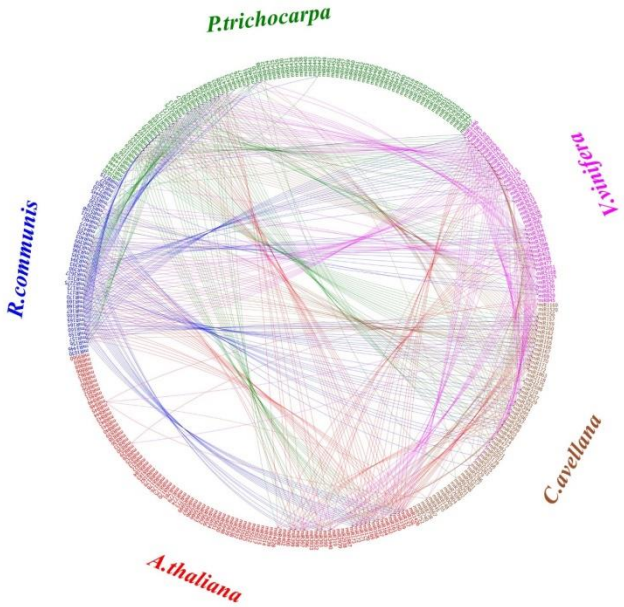

Fig 4. MicroRNA repertoire from all organisms labeled as different colors and are shown in the chart separately. All common miRNA families between each organism are also represented with the same-coloured lines inside the chart.

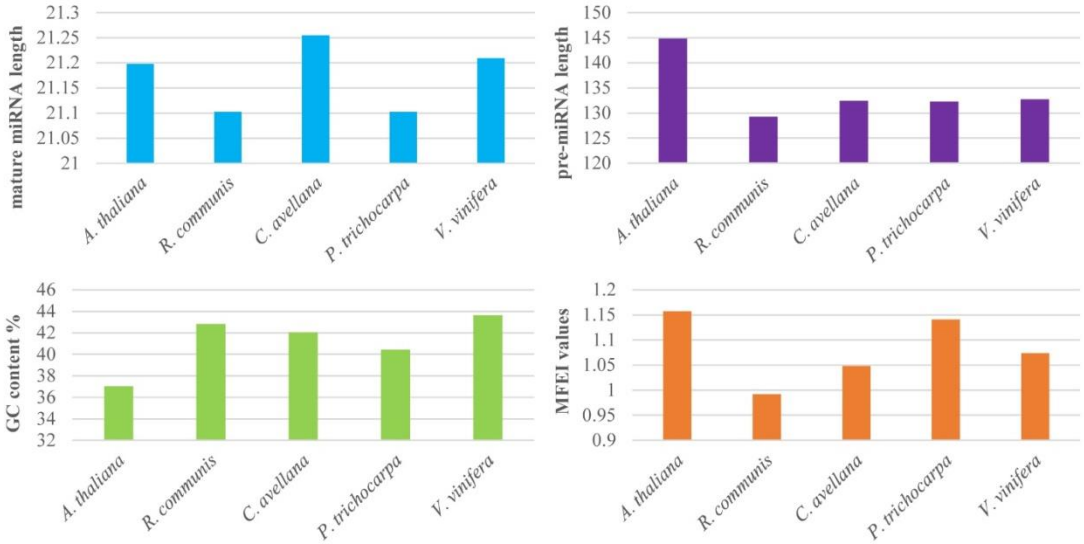

Fig 5. Brief representation of regarding the predicted miRNAs of all organisms in this study. For each organism, the average values are counted and used in the graphs.

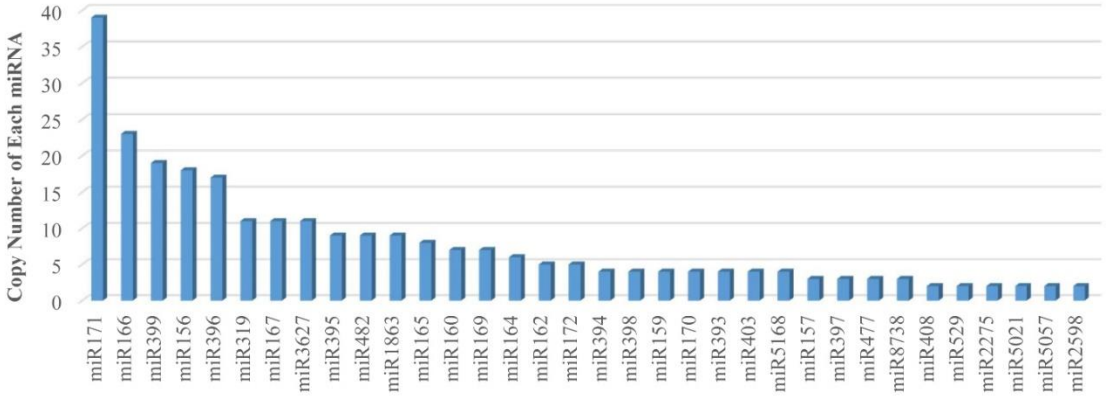

Fig 6. Representation of predicted putative miRNAs on hazelnut genome. The lowest representative miRNA families, with only one representation, are excluded here.

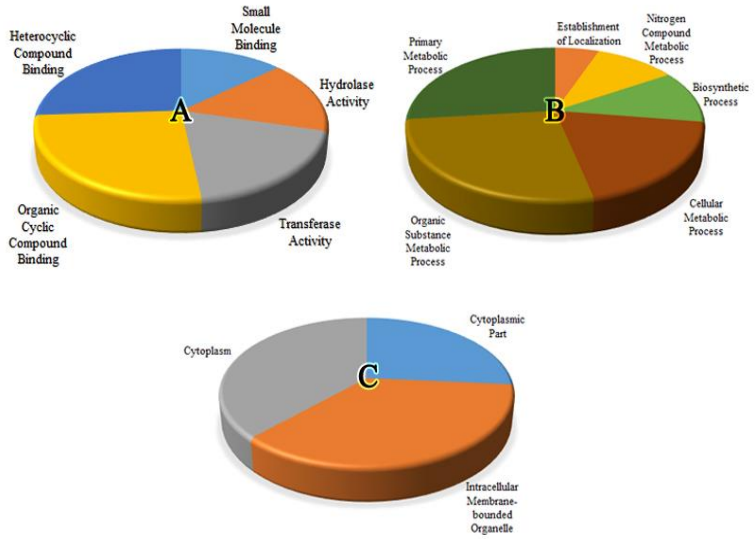

Fig 7. Functional annotation charts of hazelnut putative targets of predicted miRNAs based on Blast2Go analysis. Letters are shown as A.) Molecular Function, B.) Biological Processes and C.) Cellular Component.

maximum and minimum values. MFEI is calculated from the minimal folding-free energy (MFE), sequence length and GC% content of the pre-miRNA. Average values of all critical calculations mentioned above are shown in detail for each organism (see Fig 5). This analysis shows that our *in silico* predicted putative miRNAs were successfully differentiated from the other types of RNAs.

Representation of putative miRNA families in C. avellana genome

MiR171 families were highly represented in the hazelnut genome whereas miR408, miR529, miR2275, miR5021, miR5057, and miR2598 families had lower representation (see Fig 6). These lower representations might occurred because the corresponding miRNAs might be 'young-miRNAs' (Cuperus et al. 2011). Unlike highly conserved, ancient miRNAs, young miRNAs are often weakly expressed, lack targets, and their gene loci tend to evolve neutrally (Cuperus et al. 2011).

Those young miRNAs may have some species-specific regulatory roles inside the organisms (Sun et al. 2012; Fahlgren et al. 2007). On the other hand, the highest number of hits might be also TE-derived microRNAs because most of the transposable elements were domesticated into microRNA genes (Li et al. 2011).

Although it is really difficult to determine the certain copy number of each miRNA families (as some genomic miRNAs may be covered by more than one sequence read during the experiment whereas the others may not be covered at all), the representation of miRNA dataset provides a useful estimation about their presence on the chromosome. Those highly represented miRNAs may have a great effect on its targets (Kurtoglu et al. 2013).

Target prediction and functional annotation analysis of C. avellana putative miRNAs

We search all the targets of predicted putative miRNAs in the genome in the miRBase database. According to these results, miR156, miR157, miR159, miR160, miR162, miR164, miR165, miR166, miR167, miR169, miR170, miR171, miR172, miR319, miR393, miR394, miR395, miR396, miR397, miR398, and miR399 have experimentally validated targets (see Table 1). Most of these targets are transcription factors, promoter-binding proteins, and F-box proteins.

The target annotations are put into three main categories by the Blast2Go online web tool: Molecular functions of related targets, biological processes of related targets, and cellular components of related targets. Based on the molecular function chart, hazelnut miRNA targets mostly have functions on the organic cyclic and heterocyclic compound binding pathways (see Fig 7A).

In the biological process chart, targets have roles on metabolic events (see Fig 7B). The cellular component chart reveals that those putative miRNA-targets are mostly found in cytoplasm inside the cell (see Fig 7C). *In silico* EST analysis results show that miR396, miR482, and miR2118 families are putatively expressed in hazelnut genome. The remaining predicted miRNAs may also be transcribed, but they are not found in the current transcriptome file suggesting that they may be expressed under different conditions (Cao et al. 2014).

Materials and Methods

Reference miRNAs and sequences retrieval

The available mature miRNA sequences (8,496 sequences and 73 plant species) were downloaded from miRBase release 21 (Kozomara and Griffiths-Jones, 2013). Current miRBase corresponds to 4,802 unique mature miRNA sequences, and they were used as a query in homology-based *in silico* miRNA identification. *C. avellana* (OSU 703.007) genome was retrieved from a publicly available website[2]. *V. vinifera* (GenBank Assembly Accession Number: GCA_000003745.2) and *R. communis* (GenBank Assembly Accession Number: GCA_000151685.2) genomes were downloaded from PlantGDB[3,4]. *A. thaliana* genome (GenBank Assembly Accession Number: GCA_000001735.1) was retrieved from EnsemblPlants[5]. *P. trichocarpa* (GenBank Assembly Accession Number: GCA_000002775.2 Poptr2.0) genome sequences were also retrieved from TreeGenome database, a publicly available web site[6] (Neale et al. 2013).

In silico miRNA identification based on homology conserved method

The prediction was employed using two previously developed, in-house perl scripts: SUmirFind and SUmirFold, described in detail in the publications (Lucas and Budak, 2012). In the first step of homology-based miRNA prediction, BLAST+ stand-alone toolkit version 2.2.25 (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, 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 (Markham and Zuker, 2008). 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 eliminated by in-house script (provided on request).

Representative miRNAs (the copy number of each miRNA families) in hazelnut genome

Repeated identical pre-miRNAs (or stem-loop sequences) have different sequence read ID names but same pre-miRNA sequences. To avoid over-representation, these stem-loop sequences that were resulted from the similar query miRNA were eliminated. The remaining pre-miRNA sequences were used as representative miRNAs in the hazelnut genome.

Expressed sequence tag (EST) analysis and target annotation 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 BLAST+ stand-alone toolkit version 2.2.25, pre-miRNA sequences were blasted to hazelnut

[2] http://www.cavellanagenomeportal.com/
[3] http://www.plantgdb.org/VvGDB/
[4] http://www.plantgdb.org/RcGDB/
[5] http://plants.ensembl.org/Arabidopsis_thaliana/Info/Index
[6] http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/

transcriptome sequences[7]. To filter out dubious miRNAs in this analysis, the strict criteria were used: only miRNA families who had hits above the threshold as 98% identity and 99% query coverage were kept.

Mature sequences were identified, and duplicates were removed. By using online web tool, psRNAtarget[8], mature miRNA sequences were blasted to hazelnut transcripts (Rowley et al. 2012). The results file was downloaded and then used as an input file for Blast2Go software[9] to analyze gene ontologies (Conesa and Götz, 2008). The predicted mature miRNA sequences were also searched in miRBase 21 database to confirm their experimentally validated targets.

Conclusion

With the advent of high-throughput sequencing technologies, the genome organization of different organisms is obtained efficiently and rapidly. By using genomics tools and computational methods, we have gained a better understanding of gene networks and their interactions. In this study, we identified the miRNA repertoire of *C. avellana* and some of the other plants chosen by the previous study (Rowley, 2016) including *R. communis*, *V. vinifera*, *A. thaliana*, and *P. trichocarpa* since their gene organization and proteins were found to be more similar to those of *C. avellana* than other plants. We then compared all of these putative miRNAs to gain insight into their evolutionary relationship and to enlighten whether these miRNA genes are conserved between the organisms. The miRNA families (miR156, miR159, miR160, miR164, miR166, miR172, miR319, miR398, and miR408) from the previous studies (Gupta et al. 2014; Zhao et al. 2012) that we also predicted in our study should be analyzed under EFB disease conditions to elucidate whether they have some roles on defense mechanisms or not, and how they are regulated in the plant cell primarily. The other predicted putative miRNAs of hazelnut should also be taken into account in further studies since they may play other important roles on different plant mechanisms and they may be transferred to other accessions through breeding approaches.

Acknowledgements

We acknowledge to Daniel Lee Calvey for his academic and professional English writing of the manuscript and providing valuable comments for its enrichment.

References

Akpinar BA, Avsar B, Lucas SJ, Budak H (2012) Plant abiotic stress signaling. Plant Signaling Behav. *7*:1450-1455.

Avsar B, and Aliabadi DE (2015) Putative microRNA analysis of the kiwifruit *Actinidia chinensis* through genomic Data. Int J Life Sci Biotechnol Pharma Res. 4:96-99.

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

Avsar B, Aliabadi DE (2017) In silico analysis of microRNAs in *Spinacia oleracea* genome and transcriptome. Int J Biosci Biochem Bioinforma. 7:84-92.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinf. 10:1-9.

Cao X, Wu Z, Jiang F, Zhou R, Yang Z (2014) Identification of chilling stress-responsive tomato microRNAs and their target genes by high-throughput sequencing and degradome analysis. BMC Genomics. 15:1-16.

Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008:1-12.

Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of miRNA genes. The Plant Cell. 23:431-442.

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of miRNA genes. PLoS One. 2:1-14.

Gokirmak T, Mehlenbacher SA, Bassil NV (2009) Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. Genet Resour Crop Evol. 56:147–172.

Gupta OP, Sharma P, Gupta RK, Sharma I (2014) Current status on role of miRNAs during plant–fungus interaction. Physiol Mol Plant Pathol. 85:1-7.

Kurtoglu KY, Kantar M, Lucas SJ, Budak H (2013) Unique and conserved microRNAs in wheat chromosome 5D revealed by next-generation sequencing. PLoS One. 8:1-14.

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

Li Y, Li C, Xia J, Jin Y (2011) Domestication of transposable elements into microRNA genes in plants. PLoS One. 6:1-13.

Lucas SJ, Budak H (2012) Sorting the wheat from the chaff: identifying miRNAs in genomic survey sequences of *Triticum aestivum* chromosome 1AL. PLoS One. 7:1-11.

Markham NR, Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. Bioinformatics: Structure, Function and Applications. 3-31.

Mehlenbacher SA, Smith DC, McCluskey RL (2011) "Jefferson" hazelnut. HortScience. 46:662–664.

Neale DB, Langley CH, Salzberg SL, Wegrzyn JL (2013) Open access to tree genomes: the path to a better forest. Genome Biol. 14:1-8.

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

Rowley ER, Fox SE, Bryant DW, Sullivan CM, Priest HD, Givan SA, Mehlenbacher SA, Mockler TC (2012) Assembly and characterization of the European hazelnut 'Jefferson' transcriptome. Crop Sci. 52:2679-2686.

Rowley ER (2016) Genomic resource development for European hazelnut (Corylus avellana L.). Dissertation, Oregon State University.

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D (2005) Specific effects of microRNAs on the plant transcriptome. Dev Cell. 8:517-527.

Sun J, Zhou M, Mao Z, Li C (2012) Characterization and evolution of microRNA genes derived from repetitive elements and duplication events in plants. PLoS One. 7:1-10.

Tang G (2010) Plant microRNAs: an insight into their gene structures and evolution. Semin Cell Dev Biol. 21:782-789.

Zhang B, Pan X, Wang Q, Cobb GP, Anderson TA (2006) Computational identification of microRNAs and their targets. Comput Biol Chem. 30:395-407.

Zhao JP, Jiang XL, Zhang BY, Su XH (2012) Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. PLoS One. *7*:1-13.

[7] http://hazelnut.data.mocklerlab.org/

[8] http://plantgrn.noble.org/psRNATarget/

[9] https://www.blast2go.com