Avrupa Bilim ve Teknoloji Dergisi
Cilt. 5, No. 9, S. 1-6, Aralık 2016

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Vol. 5, No. 9, pp. 1-6, December 2016

Research Article

Identification of Conserved miRNAs and Their Target Genes in Faba Bean by EST Based Homology Analysis

Dilek KOPTEKİN[1], Lale YILDIZ AKTAS[2*]

[1] Ege University, Faculty of Science, Department of Biology, Botany Section, 35100 Bornova / Izmir Turkey
[2] Ege University, Faculty of Science, Department of Biology, Botany Section, 35100 Bornova / Izmir Turkey, lale.yildiz@ege.edu.tr

(First received 27 September 2016 and in final form 23 December 2016)

Abstract

MicroRNAs (miRNAs) are a class of endogenous, non-coding short RNAs, around 21 nucleotides (nt) in length found in eukaryotic cells and some viruses. To date, miRNAs are identified in several plant species through experimental and computational approaches where they play important roles in growth and development, metabolism, stress responses by guiding mRNA cleavage or repressing translation. Although the faba bean (Vicia faba) is an important source of protein widely used for human and animal nutrition, not a single miRNA has been identified in it up till now. Evolutionary conserved characteristics of plant miRNAs allow the identification of conserved plant miRNAs by homology analysis. The aim of this study is the identification and characterization of miRNAs in faba bean using the EST based homology analysis approach. For computational identification of novel miRNAs in faba bean, 8496 known and unique plant miRNAs from 73 plant species were searched for homology against 20697 expressed sequence tags (EST) and 577 genome survey sequences (GSS). Candidate miRNAs including protein coding sequences were recognized following the miRNA criteria of secondary structure and biogenesis. In this study, 262 candidate miRNAs belonging to 143 miRNA families have been identified for the first time in faba bean. Moreover, psRNATarget server predicted 712 potential target genes of these candidate miRNAs from faba bean. Predicted target genes seem to be involved in the regulation of several important biological processes. The results of this study will contribute to further research on miRNAs, leading to an improved understanding of the role miRNAs play in biological processes and the underlying environmental stress related molecular mechanisms of faba bean.

Keywords: Computational identification; Expressed sequence tag (EST); miRNA; Target genes; Faba bean; Vicia faba.

1. Introduction

MicroRNAs (miRNAs) are non-coding, short RNAs found in eukaryotic cells and some viruses. A mature plant miRNA is around 21 nucleotides (nt) in length which derive from longer, self-complementary transcript called primary miRNA (pri-miRNAs). Pri-miRNAs are mostly transcripted by RNA polymerase II from miRNA genes (MIR). Pri-miRNAs that fold into hairpin structures are cleaved by RNase III enzyme, Dicer-like 1 (DCL1), into a precursor miRNA (pre-miRNAs) whose length ranges from 43 to many hundreds of bases (miRBase, 2014- http://www.mirbase.org/). Processed pre-miRNAs are cleaved again by DCL1 into a miRNA / miRNA* duplexes that tend to have around 21 nt in length (Park, et al., 2002; Reinhart, et al., 2002) and then exported from the nucleus to the cytoplasm (Park, et al., 2005). In the cytoplasm, one strand of the duplex called mature miRNA is incorporated into an AGO protein, the component of RNA-induced silencing complex (RISC), while the other strand is degraded. Once a mature miRNA is associated to RISC, it interacts with its target mRNA. The base complementarities between miRNA and target mRNA determine either the mRNA cleavage or repression of translation. Plant miRNAs usually show perfect or near-perfect complementarities to their targets that generally cause the target mRNAs to break down (reviewed in Voinnet, 2009; Bologna, et al., 2013). Numerous studies have demonstrated that miRNAs play crucial roles in many biological and metabolic processes, including growth, development, metabolism, and transport as well as biotic and environmental stress responses.

The main characteristics of miRNAs like the high conservation of plant miRNAs among different species, the knowledge of plant miRNA biogenesis and the stem loop hairpin structure of pre-miRNAs with lower minimum folding energy provide foundation for the identification of miRNAs by computational approaches such as the EST based homology analysis. This approach allows for the prediction of miRNAs whose genomes have not been sequenced yet and it provides a better understanding miRNA evolution of plants.

While the first miRNA, lin-4, has been identified in Caenorhabditis elegans in 1993 (Lee, et al., 1993; Wightman, et al., 1993), the first plant miRNA was discovered in Arabidopsis in 2002 (Reinhart, et al., 2002). Since then, 28645 precursor miRNAs and 35828 mature miRNAs have been deposited in miRBase, an online database of published miRNA sequences. Some of those, 8496 mature and 6992 precursor miRNAs have

[2] **Corresponding author**: Ege University, Faculty of Science, Department of Biology, Botany Section, 35100 Bornova / Izmir Turkey, lale.yildiz@ege.edu.tr, +90 505 9352872

been identified in 73 different species in Viridiplantae (miRBase - Release 21, June 2014).

Faba bean (Vicia faba L.) is an important crop that has been cultivated from early Neolithic times (Duc, 1997). The crop is also an important source of protein not only for human but also for animal nutrition. Faba bean has high yield potential under optimum environmental conditions, but its high sensitivity to environmental stresses causes considerable reduction in the yield in comparison with other grain legumes (Hanafy, et al., 2013). According to FAO statistics, world annual production of faba bean is about 3.5 x 106 tons in 2013 and they are widely used especially in developing countries in Asia (~1.6 x 106 tons) and Africa (~0.7 x 106 tons) (FAOSTAT, 2015). Although faba bean is a good source of protein, its genome has not been sequenced yet because of its large genome size. Nevertheless, faba bean has some sequenced data i.e. expressed sequence tags (ESTs) and genome survey sequences (GSS) which are available in the NCBI-GenBank database (NCBI, 2014).

To date, numerous miRNAs have been identified by EST based homology analysis approach from both plant and animal species (Zhang, et al., 2005; Xie, et al., 2007; He, et al., 2008; Yang, et al., 2012; Panda, et al., 2014). Genome-wide comparative analysis was used to identify the conserved miRNA in various species of the Fabaceae family and determined a total of 1379 pre-miRNAs and 1545 functional miRNAs belonging to 9 species: Acacia auriculiformis (7 precursor, 7 mature), Arachis hypogaea (23 precursor, 32 mature), Acacia mangium (3 precursor, 3 mature), Glycine max (573 precursor, 639 mature), Glycine soja (13 precursor, 13 mature), Lotus japonicus (62 precursor, 67 mature), Medicago truncatula (672 precursor, 756 mature), Phaseolus vulgaris (8 precursor, 10 mature), Vigna unguiculata (18 precursor, 18 mature) deposited in miRBase (Release 21, June 2014). However, not a single miRNA has been identified in faba bean up till now. Considering the economic importance of faba bean we identified potentially conserved miRNAs and their putative target genes by using the EST based homology analysis approach to search against ESTs and GSS datasets with the previously known plant miRNA sequences as query. A total of 262 candidate miRNAs belonging to 143 miRNA families have been identified for the first time in faba bean. Moreover, 712 potential targets of these candidate miRNAs were also identified and their putative functions were analysed to improve the understanding of the role of miRNAs in metabolism, signal transduction and biotic - abiotic stress responses as well as the underlying molecular mechanisms of faba bean.

2. Materials and Experimental Procedure

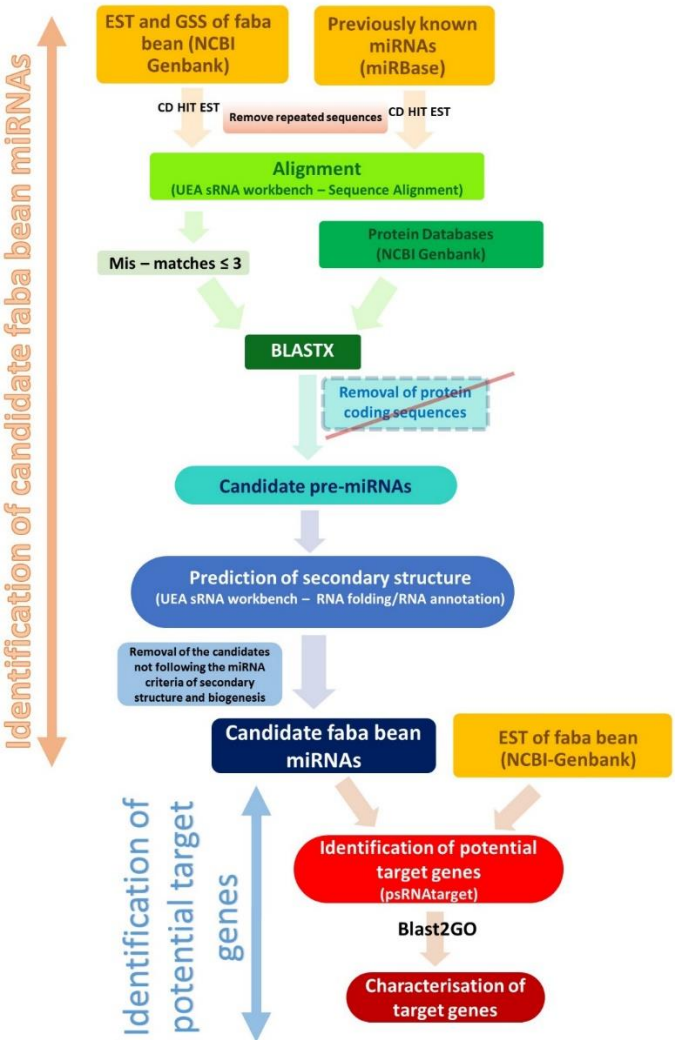
2.1. Sequence data base and reference miRNAs

A total of 8496 previously known mature miRNA sequences from 73 species in Viridiplantae were downloaded from miRBase database (Release 21, June 2014 -http://www.mirbase.org). Out of those known miRNAs, non-redundant miRNA sequences were selected as reference miRNAs using the CD-HIT-EST web server (http://weizhongli-lab.org/cd-hit/) (Huang, et al., 2010) with identity cut-off of 1.0 while other parameters were kept in the default state. Unique miRNA sequences were blasted against the 20697 EST and 577 GSS of Vicia faba which were downloaded from NCBI-Genbank nucleotide databases (http://www.ncbi.nlm.nih.gov). Reverse complement of EST and GSS sequences were produced using Biophyton (Cock, et al., 2009) script. The non-redundant EST and GSS (both positive and negative strand) were developed using CD-HIT-EST web server (Huang, et al., 2010) with default parameters and used for conserved miRNA search in faba bean.

2.2. Identification of potential (candidate) miRNAs

The major steps of the identification process were summarized in Figure 1.

Figure 1. *Workflow of the prediction of Vicia faba potential miRNAs by EST based homology analysis.*



The non-redundant mature miRNA sequences were aligned against the non-redundant EST and GSS (both positive and negative strand) of faba bean using the UEA sRNA workbench – Sequence Alignment tool (Stocks, et al., 2012) to identify novel faba bean miRNAs. UEA sRNA workbench – Sequence Alignment tool provides a graphical platform for the PatMaN (Prüfer, et al., 2008) sequence alignment tool. PatMaN makes it possible to perform a search on short patterns in large DNA databases, allowing for approximate matches with the parameters set as follows: maximum mis-matches allowed for 3 nt and maximum gaps allowed for 0. Based on the alignment results, ESTs and GSS that closely matched the query miRNAs with less than 3 mismatches (n/n, n-1/n, n-2/n and n-3/n nucleotide matches, where n equals the previously known miRNA length) were selected and used for secondary structure prediction using the UEA sRNA workbench-RNA folding/RNA annotation tool.

This tool generates a secondary structure from an RNA sequence and highlights regions of interest using RNAplot and also reports the minimum free energy of the structure (Stocks, et al., 2012).

All candidate miRNAs were recognized following the miRNA criteria of secondary structure and biogenesis as described earlier (Ambros, et al., 2003; Zhang, et al., 2005; Zhang, et al., 2006; Meyers, et al., 2008):

1. The pre-miRNA should be folded into appropriate stem loop hairpin secondary structure.
2. The minimum length of the pre-miRNA is to be 60 nt.
3. The mature miRNA sequence should be located in one arm of the hairpin structure.
4. The mature miRNA sequence and its opposite miRNA strand (miRNA*) should not have more than 6 mismatches. Moreover, the maximal bulge of miRNA and miRNA* should not be more than 4 nt.
5. No loops or breaks should be allowed between the mature miRNA sequence and its opposite miRNA strand (miRNA*)
6. The predicted secondary structure should have higher minimal folding free energy index (MFEI) and negative minimal folding free energy (MFE) (below than -18).

Studies using BLASTX against the Uniprot database remove protein coding sequences, because mature miRNAs are mostly located in non-coding regions of genomes (Zhang, et al., 2005; Song, et al., 2009; Panda, et al., 2014). However, recent studies reported that some miRNAs can be located in exonic regions (Li, et al., 2011; Yang, et al., 2012; Li, et al., 2012). Thus, sequences matching the protein coding ones were also included in candidate miRNAs in this study.

2.3. Nomenclature of candidate miRNAs

Candidate miRNAs were named according to the method used by miRBase (Kozomara & Griffiths-Jones, 2014). These miRNAs were named with the first letters of Vicia faba (vfa) and with the gene families of homologous miRNAs that they show resemblance with (such as vfa-miR156). miRNAs which belong to the same gene family but whose secondary structure are different or which are different on mature sequence were diversified by small letters coming after gene family codes (such as vfa- miR156a, vfa-miR156b).

2.4. Prediction of potential target genes of faba bean miRNAs

After identifying the candidate faba bean miRNAs, the targets of these miRNAs were predicted using the plant small RNA target analysis server (psRNATarget: http://plantgrn.noble.org/psRNATarget/). psRNATarget server allows to reverse complementary matching between a miRNA and its target transcript. Determination of target site accessibility is calculated with unpaired energy (UPE) which is required for opening the secondary structure around the miRNAs target and the server distinguishes between translational and post-transcriptional inhibition (Dai and Zhao, 2011).

The non-redundant ESTs of faba bean were used for prediction of potential target of the candidate faba bean miRNAs because they provide a direct evidence for the transcript sequences. The “user-submitted small RNAs/user-submitted transcripts” option from psRNATarget was used with the parameters that follow:

1. Maximum expectation: 2.0
2. Length for complementarities scoring (hspsize): 20 bp
3. 25 of top target genes for each small RNA
4. Target accessibility – allowed maximum energy to unpair the target site (UPE): 25.0
5. Flanking length around target site for target accessibility analysis 17 bp in upstream / 13 bp in downstream
6. Range of central mismatch leading to translational inhibition: 9 – 11 nt

2.5. Functional annotation and pathway analysis of miRNA target genes

The putative functions and metabolic pathways of EST sequences of potential targets of faba bean miRNAs were determined using Blast2GO program (Release: 3.0, November 2014). Blast2GO program uses several traits to characterize functions of miRNA target sequences (Conesa and Götz, 2008). The program was developed to make Gene Ontology (GO) identification and at the same time it is a tool supported by enzyme codes (EC), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and InterPro motives.

First, the putative functions of the ESTs were used as query sequences for BLASTX searches against NCBI non-redundant NR protein databases (QBlast-NCBI) with the parameters E-value being 1e-10 and the number of blast hits being 20. Based on BLASTX results, homolog ESTs with the previously identified proteins in NCBI were accepted as target genes. To gain a better understanding of the functional roles of the target genes in faba bean all potential target genes were obtained through functional enrichment analysis against GO databases by using InterProScan5, AmiGO2 and PlantGOSlim.

Finally, to determine the metabolic pathways of target genes, EC and KEGG pathway were produced. Thus, the enzyme codes and the metabolic pathways of target transcripts were determined by using the GO outputs which were summarized and visualized by CateGOrizer (Zhi-Liang, et al., 2008) and REVIGO webtools (Supek, et al., 2011).

3. Results and Arguments

The high degree of conservation of plant miRNAs among plant species is the basis of computational identification of conserved miRNAs from various plants (Cuperus, et al., 2011). In this study, EST based homology analysis approach was used to search novel candidate faba bean miRNAs from EST and GSS of this important crop plant whose miRNAs have not been identified yet.

3.1. Identification of candidate faba bean miRNAs

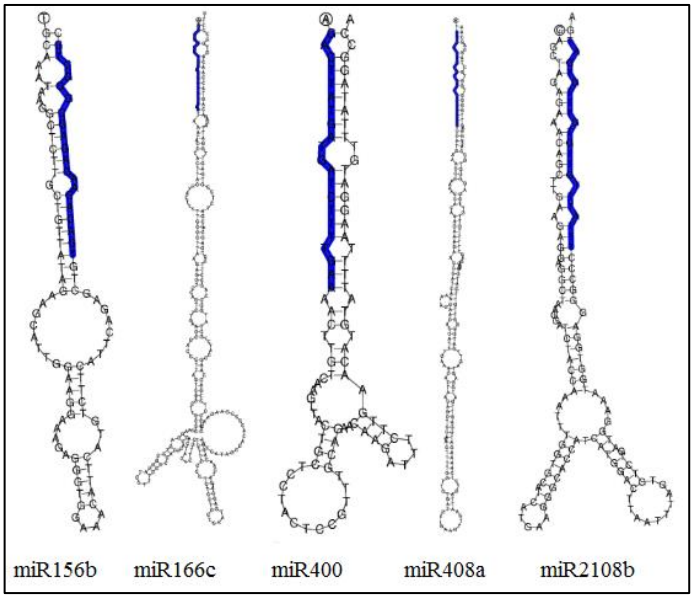
Most of the previous studies used BLAST to align previous known miRNAs against EST or GSS in their homology analysis (Zhang, et al., 2005; Panda, et al., 2014). In this study, PaTMaN was used instead of BLAST for alignment. It gives faster and better results for short sequences alignment than BLAST does (Prüfer, et al., 2008).

Previous studies used BLASTX against the Uniprot database to remove the protein coding sequences, because it was reported that mature miRNAs have been located mostly in non-coding regions of genomes (Zhang, et al., 2005; Song, et al., 2009; Panda, et al., 2014). However, recent studies demonstrated that some miRNAs were located in exonic regions of the genes because of alternative splicing (Li, et al., 2011; Yang, et al., 2012; Li, et al., 2012), which is shown for osa-miR6256 by Liu and Zhang (2012). Thus, in this study, sequences which matched those of protein coding were not removed and evaluated as candidate miRNAs as presented in Online Resource 3. (Online Resource 3- Table S2).

According to the alignment results, a total non-redundant 9208 EST and 447 GSS sequences of faba bean were selected by using of UEA sRNA workbench – Sequence Alignment tool (Stocks, et al., 2012). The non-redundant 4624 mature miRNAs from Viridiplantae were used as a reference for obtaining potential miRNA sequences with the following criteria; mis-matches ≤ 3 and gap = 0.

These potential miRNA sequences fold to the secondary structures (stem loop/hairpin) according to the validation parameters described in Materials and Methods section by using UEA sRNA workbench - RNA folding/RNA annotation tool (Stocks, et al., 2012). According to the folding results a total of 262 candidate miRNAs belonging to 143 miRNA families have been identified in faba bean (Fig. 2, Online Resource 1, Supp. Figure S1).

Figure 2. *Predicted secondary hairpin structures of some Vicia faba miRNAs precursors. Mature miRNA sequences are highlighted with blue colour.*



Within the candidate miRNAs, 253 miRNAs have been identified from the ESTs and remaining 9 miRNAs the GSS (miR845, miR1044c, miR1888, miR5140f, miR5506, miR6182b, miR6281d, miR6281e and miR8714) from the GSS. Among predicted miRNAs; 3 miRNAs have 0 (zero) mismatch, 33 miRNAs 1 (one) mismatch, 33 miRNAs 2 (two) mismatches and the 223 miRNAs 3 (three) mismatches (Supp. Table S1). All candidate miRNAs searched against non-reduntant protein database by BLASTX and 31 miRNA seems to be located on the protein-coding genes such as V. faba mRNA for phloem specific protein, V. faba cultivar Mammoth Sat5 gene (Online Resource 3, Supp. Table S2). The location of candidate faba bean miRNAs on protein coding genes were showed in Online Resource 5. (Online Resource 5- Table S1).

The distinguishing characteristics of the predicted miRNAs from other small RNAs such as miRNA mature sequences, length of mature miRNAs, miRNA location, length of precursors, MFEs, and MFEIs were analysed. miRBase data analysis shows that mature plant miRNAs' length is between 17-26 nt with a high frequency for 21 nt in length (62.68 %) based on the biogenesis criteria of miRNAs. Candidate faba bean miRNAs have been identified and were found to be 17-24 nt in length and 56.44 % of them were 21 nt in length, which is consistent with the size scale of previously known plant miRNAs of Arabidopsis thaliana (Zhang, et al., 2005), Glycine max (Zhang, et al., 2008), and Phaseolus vulgare (Han, et al., 2014) and so on.

The nucleotide distribution of mature faba bean miRNAs displayed a tendency to be first base towards a 5' uracil, of which 128 candidate mature miRNAs (48,48%) showed biases towards 5' uracil as characteristic features of miRNAs (Dhandapani, et al., 2011). miRBase data analysis with known plant miRNAs also indicated a tendency to be first base towards a 5' uracil and 55.36% of previously known plant mature miRNAs started with 5' U (Dhandapani, et al., 2011; miRBase, 2014).

To determine pre-miRNAs, the criteria of the minimum length of pre-miRNAs was stated as 60 nt (Ambros, et al., 2003; Zhang, et al., 2005; Zhang, et al., 2006; Meyers, et al., 2008). However, data about the limitation of the maximum length of pre-miRNAs was not indicated in the literature so far. The 6992 pre-miRNAs belonging to the Viridiplantae are stored in miRBase ranging from 60-350 nt in length. Only 4.32% of them were found outside of this range. Thus, wherever available the EST and GSS sequences of 700 nt (350 nt upstream and 350 nt downstream) were extracted and used for secondary structure prediction. The length of the candidate pre-miRNAs ranged from 60 nt to 346 nt (average 170 nt) were previously reported in which is in agreement with the previously reported results in Arabidopsis thaliana (Zhang, et al., 2005), Glycine max (Zhang, et al., 2008), Brassica rapa (Dhandapani, et al., 2011), Phaseolus vulgare (Han, et al., 2014) and Allium sativum (Panda, et al., 2014).

Mature miRNA sequence locations were found at either 5' end or the 3' end of the stem loop hairpin structure of potential miRNAs. Predicted121 sequences of faba bean miRNAs were located at the 3′ end, while the other 141 sequences were found at the 5′ end of the miRNA precursors. While 112 sequences of miRNAs were located only in positive strand and 128 sequences of miRNAs were located only in negative strand, 22 sequences of miRNAs seem to be located both positive and negative stand (Supp. Table S1; Supp. Fig. S1). The predicted faba bean pre-miRNA sequences were assessed for their G+C contents, and results showed that G+C contents ranged from 25 to 65% in the faba bean miRNA precursors with an average of ~ 31,92 (Online Resource 2, Supp. Table S1). These findings match the results of previously known plant miRNAs (Ambros, et al., 2003; Zhang, et al., 2005; Zhang, et al., 2006; Meyers, et al., 2008). The minimum free energy (MFE) value is a significant criterion for considering the stability of the secondary hairpin structures (Prabu and Mandal, 2010). The negative MFE value of the predicted secondary structures of faba bean pre-miRNAs ranged from -18.2 kcal/mol to -116.1 kcal/mol (average ~ 43.37) as being consistent with pre-determined criteria. The adjusted minimal folding free energy (AMFE) (average ~ -25.71), and minimal folding free energy index (MFEI) (average ~ -0.62) values of the candidate miRNA precursors are both calculated by using MFE values,

respectively. The results indicated that the predicted miRNAs in faba bean are probably true miRNAs in agreement with criteria of secondary structure and biogenesis of miRNAs (Ambros, et al., 2003; Zhang, et al., 2005; Zhang, et al., 2006; Meyers, et al., 2008).

3.2. Identification of potential target genes of miRNAs

The identification of the targets for the candidate miRNAs contribute to evaluation of the function and regulation of the newly predicted miRNAs in faba bean. Plant miRNAs show perfect or near-perfect complementarities with their target mRNAs as well as functional genes (Nodine and Bartel, 2010). Non-redundant 9208 ESTs of faba bean were used for identification of the potential targets of candidate faba bean miRNAs by using psRNATarget webserver. While 712 potential targets have been predicted from non-redundant EST sequences of faba bean for 99 miRNA families, however no potential target was predicted for 44 miRNA families (miR171, miR319, miR399, miR774, miR775, miR816, miR848, miR858, miR1028, miR1032, miR1061, miR1153, miR1168, miR1507, miR1513, miR1533, miR1888, miR2084, miR2095, miR2592, miR2662, miR3513, miR3520, miR4239, miR5077, miR5140, miR5246, miR5368, miR5384, miR5506, miR5565, miR5640, miR6182, miR6202, miR6300, miR6473, miR7534, miR7542, miR7817, miR8123, miR8610, miR8656, miR9489, miR9674). These 712 ESTs match with faba bean miRNA sequences from 99 miRNA gene families which were regulated by either cleavage or repression of translation. The predicted target genes for the faba bean miRNAs indicated that individual miRNA may regulate more than one gene in consistence with the reports for other studies about plant miRNAs (Zhang, et al., 2005; Zhang, et al., 2008; Song, et al., 2009; Dhandapani, et al., 2011; Han, et al., 2014; Panda, et al., 2014).

Blast2GO program was performed to gain a better understanding of the functions of these predicted target sequences in faba bean. The putative functions of the ESTs of potential target sequences of faba bean miRNAs were used as query sequences for BLASTX searches against NCBI non-redundant NR protein databases. The EST sequences regulated by 4 miRNA gene families, namely miR835, miR845, miR8005 and miR9559 showed no resemblance with any protein; however, the EST sequences regulated by the other 95 miRNA families resembled the genes stored in NCBI. The sequences identified with the best hits and suitable in terms of e-value by BLASTX were accepted as target genes and their GO, EC and KEGG pathway analysis was produced using default settings.

According to the results of the function identification and metabolic pathway analysis, it is estimated that miRNA sequences belonging to 95 miRNA gene families of faba bean played important roles in biological and molecular processes (Online Resource 4, Supp. Table S3). GO terms give the processes in which the genes were analysed functionally in three categories: i) cellular component, ii) molecular function and iii) biological process. GO analysis of target sequences of candidate faba bean miRNAs showed that 34 different terms were enriched terms for cellular component, 77 different terms were significantly enriched terms for metabolic function and also 94 different terms were significantly enriched terms for biological processes (Supp. Table S4). The putative target sequences of miRNAs in molecular function category were related to binding, catalytic activity, oxidoreductase activity, hydrolase activity and so on. GO analysis indicated that 94 different terms were related with development, metabolism, transport, signal transduction and predicted 20 of them under the function of biological process associated with biotic stress, particularly abiotic stress responses (Table 1).

Table 1. *GO analysis of stress related miRNA targets in faba bean.*

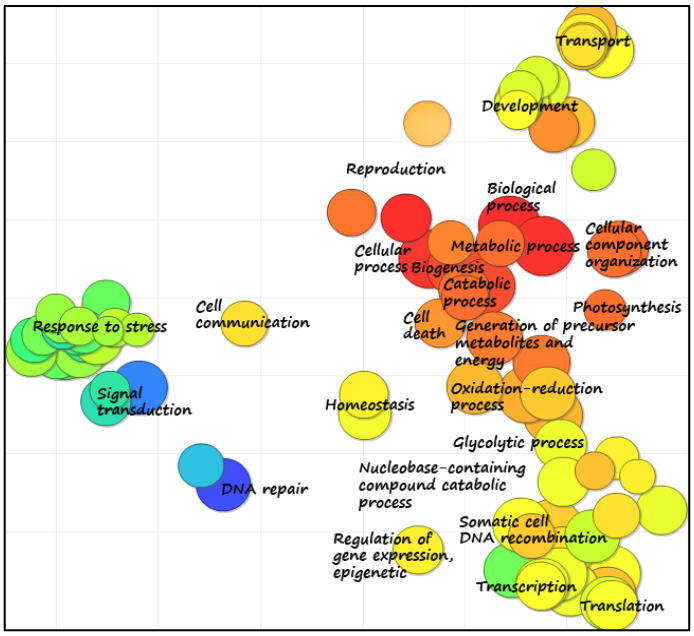
Faba bean miRNA families	GO terms related to stress
miR167, miR169, miR414, miR838, miR2611, miR2938, miR3631, miR5021, miR5198, miR5745, miR5998, miR6034, miR6281, miR6482, miR8709	GO:0006950 Response to stress
miR167, miR169, miR414, miR838, miR2611, miR2938, miR3631, miR5021, miR5198, miR5745, miR5998, miR6034, miR6281, miR6482, miR8709	GO:0006952 Defense response
miR400, miR5658, miR5658, miR7777, miR8709	GO:0006979 Response to oxidative stress
miR400, miR5658	GO:0009408 Response to heat
miR400	GO:0009409 Response to cold
miR400	GO:0009414 Response to water deprivation
miR400, miR5021, miR5998	GO:0009415 Response to water
miR167	GO:0009416 Response to light stimulus
miR2611, miR5745, miR6034	GO:0009605 Response to external stimulus
miR2611, miR2868, miR5745, miR6034	GO:0009607 Response to biotic stimulus

miR414, miR838, miR2938, miR3631, miR5021, miR5998, miR6281, miR6482	GO:0009628 Response to abiotic stimulus
miR400	GO:0009644 Response to high light intensity
miR400, miR5658	GO:0009651 Response to salt stress
miR400, miR5253	GO:0009737 Response to abscisic acid
miR6281	GO:0009753 Response to jasmonic acid
miR6281	GO:0009867 Jasmonic acid mediated signaling pathway
miR5021	GO:0009991 Response to extracellular stimulus
miR400	GO:0042542 Response to hydrogen peroxide
miR6281	GO:0042742 Defense response to bacterium
miR5021	GO:0043562 Cellular responses to nitrogen levels
miR5658	GO:0070370 Cellular heat acclimation

Moreover, based on KEGG biochemical pathway analysis 44 miRNAs' target transcripts involved in different cellular pathways including biosynthesis of secondary metabolites, purine metabolism, and thiamine metabolism were determined in faba bean (Supp. Table S4).

The GO outputs obtained by functional identification and metabolic pathway analyses were visualized by using CateGOrizer (Zhi-Liang, et al., 2008) and REVIGO web tools (Supek, et al., 2011). REVIGO distribution graph categorizes the GO outputs which show significant resemblance in two groups and show them in a two dimensional graph (Fig. 3).

Figure 3. *Characterization of the biological process category of the predicted miRNA target. The coordinates in the graph do not express quantified values. The principle of the graph is the distance between GO outputs showed resemblances.*



4. Conclusion

According to miRBase (Release 21, June 2014) 1379 precursor and 1545 mature miRNA sequences belonging to 9 species have been identified and deposited in Fabaceae. However, not a single miRNA has been identified in faba bean in literature up to now, which is an important source of protein widely used for human and animal nutrition. In this study, 262 candidate faba bean miRNAs belonging to 143 miRNA families have been identified from 20697 ESTs and 577 GSS using computational approach. This study revealed that homology based bioinformatics approaches like EST based homology analysis are effective to identify and make functional analysis of miRNAs of important agricultural plants especially those genome sequences are not known. The next aim of this study is to verify experimentally the predicted candidate faba bean miRNAs. The results of this study will contribute to further research on miRNAs, leading to an improved understanding of the role miRNAs play in biological processes and the underlying molecular mechanisms of faba bean. Thus, it will be possible to increase the quality and efficiency of important agricultural products by miRNA based approaches. As a matter of fact, transgenic approaches may provide higher stress tolerance to faba bean by either decreasing or increasing the expression of some miRNAs which are specific to some known environmental stimuli.

References

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., Tuschl, T. 2003. A uniform system for microRNA annotation. RNA 9(3), 277-279.

Bologna, N.G., Schapire, A.L., Palatnik, J.F. 2013. Processing of plant microRNA precursors. Briefings in Functional Genomics 12(1), 37-45.

Cock, P.J.A., Tiago. A., Jeffrey, T.C., Brad, A.C., Cymon, J.C., Andrew, D., Iddo, F., Thomas, H., Frank, K., Bartek, W., Michiel, J.L. de H. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11), 1422-1423.

Conesa, A., Götz, S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics, 2008, 1-13.

Cuperus, J.T., Fahlgren, N., Carrington, J.C. 2011. Evolution and functional diversification of MIRNA genes. The Plant Cell 23(2), 431-442.

Dai, X., Zhao, P. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Research 39, 155-159.

Dhandapani, V., Nirala, R., Parameswari, P., Joonki, K., Sun, H.C., Jeongyeo, L., Yoonkang, H., Yong, P.L. 2011. Identification of potential microRNAs and their targets in *Brassica rapa* L. Molecules and Cells 32, 21-37.

Duc, G. 1997. Faba bean (*Vicia faba* L.). Field Crops Research 53, 99-109.

FAOSTAT, 2015. Food and Agriculture Organizatıon of The United Nations. http://faostat3.fao.org/browse/Q/QC/E.

Han, J., Xie, H., Kong, M.L., Sun, Q.P., Li, R.Z., Pan, J.B. 2014. Computational identification of miRNAs and their targets in *Phaseolus vulgaris*. Genetics and Molecular Research 13(1), 310-322.

Hanafy, M.S., El-Banna, A., Schumacher, H.M., Jacobsen, H.S., Hassan, F.S. 2013. Enhanced tolerance to drought and salt stresses in transgenic faba bean (*Vicia faba* L.) plants by heterologous expression of the PR10a gene from potato. Plant Cell Reports 32(5), 663-674.

Hao, Y., Haiyang, Z., Lin, Z., Chenyu, Z., Donghai, L. 2012. Identification and characterization of microRNAs in *Macaca fascicularis* by EST analysis. Comparative and Functional Genomics ID: 957607. http://dx.doi.org/10.1155/2012/957607.

He, P.A., Nie, Z., Chen, J., Chen, J. L.Z., Sheng, Q., Zhou, S., Gao, X., Kong, L., Wu, X., Jin, Y., Zhang, Y. 2008. Identification and characteristics of microRNAs from *Bombyx mori*. BMC Genomics 9(248), 1-17.

Huang, Y., Beifang, N., Ying, G., Limin, F., Weizhong, L. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26(5), 680-682.

Kozomara, A., Griffiths-Jones, S. 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research 42, 68-73.

Lee, R.C., Feinbaum, 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, T., Li, H., Zhang, Y.X., Liu, J.Y. 2011. Identification and analysis of seven H_2O_2-responsive miRNAs and 32 new miRNAs in the seedlings of rice (*Oryza sativa* L. ssp. indica). Nucleic Acids Research 39, 2821-2833.

Li, W., Xiao, C., Zhaolu, M., Xiahe, H., Qi, X., Heng, W., Hailing, J., Dabing, Z., Wanqi, L., 2012. Transcriptional regulation of *Arabidopsis* MIR168a and argonaute1 homeostasis in abscisic acid and abiotic stress responses. Plant Physiology 158(3), 1279-1292.

Liu, Q., Zhang, H., 2012. Molecular identification and analysis of arsenite stress-responsive miRNAs in rice. Journal of Agricultural and Food Chemistry 60, 6524-6536.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qio, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., Zhu, J.K. 2008. Criteria for annotation of plant MicroRNAs. The Plant Cell 20(12), 3186-3190.

miRBase. 2014 *miRBase.* http://www.mirbase.org.

NCBI. 2014. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov.

Nodine, M.D., Bartel, D.P. 2010. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes and Development, 24(23), 2678-2692.

Panda. D., Dehury, B., Sahu, J., Barooah, M., Sen, P., Modi, M.K. 2014. Computational identification and characterization of conserved miRNAs and their target genes in garlic (*Allium sativum* L.) expressed sequence tags. Gene 537(2), 333-342.

Park, M.Y., Gang, W., Alfredo, G.S., Hervé, V.R., Scott, P. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. PNAS 102, 3691-3696.

Park, W., Junjie, L., Rentao, S., Joachim, M., Xuemei, C. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Current Biology 12(17), 1484-1495.

Prabu, G.R., Mandal, A.K. 2010. Computational identification of miRNAs and their target genes from expressed sequence tags of tea *(Camellia sinensis*). Genomics Proteomics and Bioinformatics 8(2), 113-121.

Prüfer, K., Stenzel, U., Dannemann, M., Green, R.E., Lachmann, M., Kelso, J. 2008. PatMaN: rapid alignment of short sequences to large databases. Bioinformatics, 24(13), 1530-1531.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P. 2002. MicroRNAs in plants. Genes Development 16: 1616-1626.

Song. C., Fang, J., Li. X., Liu, H., Chao, C.T. 2009. Identification and characterization of 27 conserved microRNAs in citrus. Planta 230(4), 671-685.

Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T., Moulton, V. 2012. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics 28(15), 2059-2061.

Supek, F., Bosnjak, M., Skunca, N., Smuc, T. 2011. REVIGO Summarizes and visualizes long lists of gene ontology terms. PLoS One 6(7), e21800. http://dx.doi.org/10.1371/journal.pone.0021800

Voinnet, O. 2009. Origin, biogenesis, and activity of plant microRNAs. Cell 136(4), 669-687.

Wightman, B., Ha, I., Ruykun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75: 855-862.

Xie, F.L., Huang, S.Q., Guo, K., Xiang, A.L., Zhu, Y.Y, Nie, L., Yang, Z.M. 2007. Computational identification of novel microRNAs and targets in *Brassica napus*. FEBS Letters 581(7), 1464-1474.

Yang, G.D., Yan, K., Wu, B.J., Wang, Y.H., Gao, Y.X., Zheng, C.C. 2012. Genomewide analysis of intronic microRNAs in rice and *Arabidopsis*. Journal of Genetics 91(3), 313-324.

Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P., Anderson, T.A. 2005. Identification and characterization of new plant microRNAs using EST analysis. Cell Research 15(5), 336-360.

Zhang, B., Pan, X., Cox, S. 2006. Evidence that miRNAs are different from other RNAs. Cellular and Molecular Life Sciences 63, 246-254.

Zhang, B.H., Pan, X.P., Stellwag, E.J. 2008. Identification of soybean microRNAs and their targets. Planta 229, 161-182.

Zhi-Liang, H,. Bao, J., Reecy, J.M. 2008. CateGOrizer: A web-based program to batch analyze gene ontology classification categories. Online Journal of Bioinformatics 9(2), 108-112.