

Prediction of miRNA targets, affected proteins and their homologs in *Glycine max*

Pramod Katara^{1*}, Budhayash Gautam², Himani Kuntal¹, Vinay Sharma¹

¹Department of Bioscience and Biotechnology, Banasthali University, P.O. Banasthali Vidyapith, India-304022; ²Department of Computational Biology & Bioinformatics, JSBB, SHIATS, Allahabad, India-211007. Pramod Katara - E-mail: pmkatara@gmail.com, Phone: +919413094705, *Corresponding author

Received August 9, 2010; accepted August 26, 2010; published September 20, 2010

Abstract:

microRNAs are small noncoding RNA gene products about 20-24nt long that are processed by Dicer from precursors with a characteristic hairpin secondary structure. As miRNAs affect the morphology of plants and animals by the posttranscriptional regulation of genes involved in critical developmental events, it has been proposed that precise regulation of miRNAs activity during various stages of growth and in specific cell types is of central importance for normal plant development. In our work we focus on the plant miRNAs and predict the miRNA targets, affected proteins by miRNA and miRNA homologs of *Glycine max*. Our analyses were based on sequence complementarities between miRNAs and mRNAs. As a result, we predicted 573 targets for 44 mature miRNAs sequences among 69 mature miRNAs sequences were published in database. Study of affected proteins revealed that for very less number of miRNAs, protein products are known and they mostly involved in diverse physiological process like as element of photosynthesis system. Homology analyses for miRNAs suggested that 22 miRNAs of *Glycine max* show 418 miRNA homologs for different plant species.

Keywords: EST; gene expression; miRNA targets; homologs and noncoding RNA.

Background:

MicroRNAs (miRNAs) are small ~20-24 nucleotide endogenous and non-coding regulatory RNAs that are processed from hairpin RNA precursors encoded within the genome of metazoans. In both animals and plants, the majority of the miRNA genes exist as independent transcriptional units, and they are transcribed by RNA polymerase II into long primary transcripts (termed pri-miRNAs) [1].

The first miRNA was discovered in *Caenorhabditis elegans* hundreds of miRNAs also have been cloned in various plant species [2, 3, 4], but the progress in prediction of the genes responsible for these miRNAs production, and cleavage target/sites of these miRNAs are still at very low rate which need to speed up to understand the fundamentals of gene regulation system at miRNA level, more accurately and completely [1].

In nature, miRNA acts as post-transcriptional regulators in animals and plants. miRNAs use two distinct post-transcriptional mechanisms to regulate gene expression. They act by binding to the complementary sites on the 3' untranslated region (UTR) of the target gene to induce cleavage with near perfect complementarity [2] or to repress productive translation; they also facilitate deadenylation, which leads to rapid mRNA decay.

miRNAs are very important regulators for development in plants, they code for proteins with a known or suspected role in developmental control [5, 3]. Abnormality or dysfunction of these miRNA has been shown to cause various defects in plants and its various development processes like floral and leaf-patterning defects, loss of organ polarity, altered vascular development and reduced fertility. Given these clear roles in plant development, it has been proposed that precise regulation of miRNA activity during various stages of growth and in specific cell types is of central importance for normal plant development [5, 1]. Thus to get insides about the post-translational regulation and importance of miRNA, it is very important and essential to understand about the interaction between miRNA and their complementary targets present on mRNA.

Several computational methods are developed and have been widely used by scientist for the prediction of miRNAs [6] and their target genes using

different approaches [5]. Several approaches have been used for miRNA target predictions in plants and animals. For plant, sequences similarity-based simple pattern matching approaches or BLAST similarity searches have shown high performance, because complementarity in miRNA and their targets is nearly perfect [5, 3], in case of animal they need different approaches.

Lot of work is already going on miRNA in plants but it mostly focus on world's most important crop Rice (*Oryza sativa*) [7] and model organism in plant biology, *Arabidopsis thaliana* [8] because of the availability of their complete genome sequence and other information that provide the basis for *in silico* analysis, but lot of work yet to have done to fully understand miRNA system in other plant species. This present study was undertaken to identify miRNAs targets, affected protein by miRNA and miRNA homologs of *Glycine max*, using simple sequence similarity search method.

Methodology:

Retrieval of miRNA sequences from miRBase database: For the current study, we used MiRBase database (<http://microrna.sanger.ac.uk/>) [9]. Its current release (12.0) contains 8619 miRNA loci from 87 species including *Homo sapiens*, flies, plants, animals and viruses. Among 87 species, 21 are plant species. Our work focus on *Glycine max*, for this, all available sequences of mature miRNAs of *Glycine max* (69) were downloaded from miRBase database.

Identification of miRNA targets: To find the targets for miRNAs, we compared them with their EST database dbEST (<http://blast.ncbi.nlm.nih.gov/dbEST/>) available at NCBI (National Center for Biotechnology Information) [10], with the help of BLAST tool [11]. Corresponding mRNA of EST, with high similarities (E-value less than 0.1) had been considered as potential targets for miRNA.

Identification of affected proteins: these targets encode specific protein whose translation is being suppressed due to binding of miRNA with corresponding mRNAs. In order to identify the miRNA affected proteins, with their functions, we had converted target sequence (nucleotide sequence) into protein query sequence. For this purpose BLASTX was

used, which uses nucleotide sequences as queries and translate them in all six reading frames to produce translated protein sequences. These translated protein sequences further used as protein query for protein sequence Database. Further, to get more specified results, target sequences with significant statistics values had been selected, and BLASTX was performed for these selected target sequences against SwissProt Database (highly annotated protein database). Hits with highest bit score and lowest E-value had been considered as affected proteins by miRNAs of *Glycine max*.

Identification of miRNA homologs: To identify miRNA homologs, sequences of mature miRNAs (which are conserved since their sequence are crucial for target-interaction) were taken as input for sequence similarity search against mature miRNA sequences of all other plants available in miRBase Database [12]. miRNAs of all other plant species with bit score more than 100 against query sequence had been selected as homolog miRNAs.

Result and Discussion:

Identification of miRNA targets: Previous studies suggest that miRNAs bind to their mRNA target with nearly perfect sequence complementary, and degrade the target mRNA [5]. On the basis of this concept we used sequence similarity search between our query sequence of miRNA and EST database and got 573 targets for 44 mature miRNA sequences among 69 mature miRNA sequences published in database. These targets were also showing significant sequence similarities with E-value ranging from 0.025 to 0.0004.

We observed that, among 573 miRNA targets of *Glycine max*, 61% of targets were showing E-value around 0.0004 and 17% were showing E-value around 0.001, thus most of the miRNA with their respective target sequence shared very high similarities which are significant and provide accuracy [5, 3]. High number of predicted targets (mRNAs) for the miRNA candidates of *Glycine max* (Fig. 1) were observed, as compared to other well studied plants i.e 49 for 13 miRNA in *Arabidopsis* [5], revealed that a broad range of genes were regulated by miRNA in *Glycine max*.

Identification of affected proteins: In *Glycine max* we got 458 targets sequences using suitable statistical parameter i.e. concerning bit score (40 and above) and E-value. Most of the targets encode for Maturase K (Intron maturase) and Photosystem I P700 chlorophyll a apoprotein A1 (PsaA/PSI-A). Probably, Maturase K assists in splicing of its own and other chloroplast group II introns [13]. Psa A proteins bind P700, the primary electron donor as well as the electron acceptor of photosystem I [14].

These proteins are among those protein whose formation is terminated or altered by the action of miRNA binding and interfere RNA splicing and photosynthesis respectively. Our prediction suggest that miRNA also control those mRNA whose proteins are involved in various process other than just as transcription factors, which was earlier suggested [5], and play broader regulatory role like in animal.

Identification of miRNA homologs: In *Glycine max* 22 miRNA targets were having total 418 homologs with-in species like *Physcomitrella patens*, *Triticum aestivum*, *Sorghum bicolor*, *Gossypium hirsutum*, *Vitis vinifera*, *Oryza sativa*, *Brassica napus* and many more.

Total number of miRNA homologs for *Glycine max* was 418 and 92 miRNA sequences were showing homology with miRNA of species *Populus trichocarpa*. Therefore miRNAs of *Glycine max* were more conserved with in *Populus trichocarpa*. miRNAs of *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*, *Zea mays* and *Brassica napus* were also showing homology with *Brassica napus* (Figure. 1). Analysis suggests that miRNA are evolutionary conserved and shared similarities with different plant species, over wide evolutionary distances [15, 16]. Here it is notable that *Glycine max* share maximum similarities (homology) with *Populus trichocarpa* and minimum similarities with species *Triticum aestivum* and *Saccharum officinarum*.

Conclusion:

Present study was focused on the prediction of the miRNA targets, affected proteins and miRNA homologs of *Glycine max*; as a result we found 573 mRNA targets for 44 mature miRNA sequences among 69 mature miRNA sequences published in database. Its most of the targets encode for Maturase K and PsaA/PSI-A that are very important elements of photosynthesis system, this suggest that miRNAs play major role in the regulation of photosynthesis process, thus their activity in *Glycine max* may cause major alteration in photosynthesis process. Homology analysis for miRNA suggest that 22 miRNA of *Glycine max* show 418 miRNA homologs from different plant species, these homologies indicates that miRNAs of most of the plants belongs to same family. Here we observed that *Glycine max* share maximum similarities with *Populus trichocarpa*. Mined potential miRNA targets and affected functions further can be validated by using available approach i.e. RT-PCR and 5'RACE which is based on PCR assay. Availability of microarray data in future can be used to asses the relative expression levels of potential mRNA targets in tissues in which their miRNAs were expressed, which will provide the means to understand and validate miRNA targets.

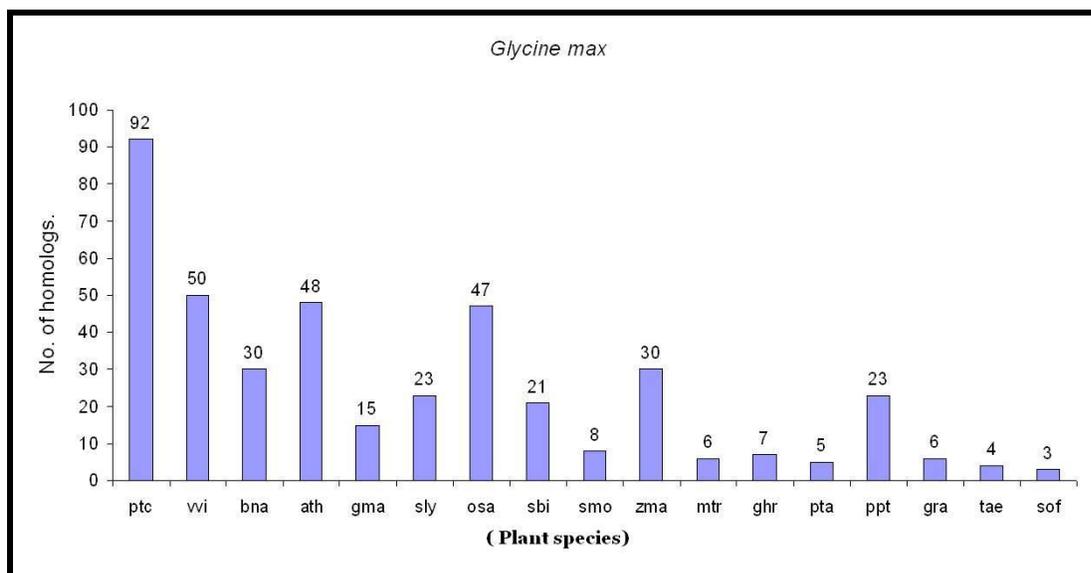


Figure 1: Plant share miRNA homologs with *Glycine max*

Acknowledgements

We acknowledge the bioinformatics facility at Department of Bioscience and Biotechnology, Banasthali University, Banasthali for providing essential facilities for completion of this research work.

References:

- [1] DP Bartel, *Cell* **116**: 281 (2004) [PMID: 14744438]
 [2] C Llave *C Science* **297**: 2053 (2002) [PMID: 12242443]
 [3] MW Jones-Rhoades & DP Bartel *Mol Cell*. **14**:787 (2004) [PMID: 15200956]
 [4] XJ Wang *et al. Genome Biol.* **5**(9):R65 (2004) [PMID: 15345049]
 [5] MW Rhoades *et al. Cell.* **110**:513 (2002) [PMID: 12202040]
 [6] JW Nam *et al. Nucleic Acids Res.* **33**:3570 (2005) [PMID: 15987789]
 [7] R Sunkar *et al. BMC Plant Biol.* **8**:25 (2008) [PMID: 18312648]
 [8] E Allen *et al. Cell.* **121**(2): 207 (2005) [PMID: 15851028]
 [9] S Griffiths-Jones *Methods Mol Biol.* **342**:129 (2006) [PMID: 16957372]
 [10] MS Boguski *et al. Nat Genet.* **4**:332 (1993) [PMID: 8401577]
 [11] SF Altschul *et al. J Mol Biol.* **215**(3):403 (1990) [PMID: 2231712]
 [12] T Dezulian *et al. Bioinformatics.* **22**(3):359 (2006) [PMID: 16317073]
 [13] J Vogel *et al. Nucleic Acids Res.* **27**: 3866 (1999) [PMID: 10481026]
 [14] L Krabben *et al. Biochemistry.* **39**(42):13012 (2000) [PMID: 11041867]
 [15] B Zhang *et al. Plant J.* **46**(2):243 (2006) [PMID: 16623887]
 [16] A Barakat *et al. BMC Genomics* **8**:481 (2007) [PMID: 18166134]

Edited by P. Kanguane

Citation: Katara *et al.* Bioinformatics 5(4): 162-165 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: List of plants species which was consider for comparative analysis in present study

<i>Abb.</i>	<i>ptc</i>	<i>vvi</i>	<i>bna</i>	<i>ath</i>	<i>gma</i>	<i>sly</i>	<i>osa</i>	<i>sbi</i>	<i>smo</i>	<i>zma</i>	<i>mtr</i>	<i>ghr</i>	<i>pta</i>	<i>ppt</i>	<i>gra</i>	<i>tae</i>	<i>sof</i>
<i>Plant name</i>	<i>Populus trichocarpa</i>	<i>Vitis vinifera</i>	<i>Brassica napus</i>	<i>Arabidopsis thaliana</i>	<i>Glycine max</i>	<i>Solanum lycopersicum</i>	<i>Oryza sativa</i>	<i>Sorghum bicolor</i>	<i>Selaginella moellendorffii</i>	<i>Zea mays</i>	<i>Medicago truncatula</i>	<i>Gossypium hirsutum</i>	<i>Pinus taeda</i>	<i>Physcomitrella patens</i>	<i>Gossypium raminidii</i>	<i>Triticum aestivum</i>	<i>Saccharum officinarum</i>