In silico identification of miRNAs and their targets from the expressed sequence tags of *Raphanus sativus*

Charuvaka Muvva[1]*, Lata Tewari[2], Kasoju Aruna[3], Pabbati Ranjit[3], Zahoorullah S MD[4], K A Matheen MD[5], Hemanth Veeramachaneni[6]

[1]Muvva Biosolutions Pvt.Ltd, Bioinformatics Division, #301, No: 7-1-414/21, Kaushik Sai Rama Residency, Srinivasa colony (East) S.R.Nagar, Hyderabad-500038, India; [2]Department of Bioscience and Biotechnology, Banasthali University, P.O.Banasthali Vidyapith, Rajasthan- 304022, India; [3]Department of Biotechnology, JNTU College of Engineering, Pulivendula, Kadapa (Dist), Andhra Pradesh, India-516390; [4]Department of Biotechnology, Joginpally B R Engineering College, Yenkapally, Moinabad, Ranga Reddy (Dist), Andhra Pradesh, India-500075; [5]Department of Biotechnology, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad-500075, India; [6]Muvva Biosolutions Pvt.Ltd Bioinformatics Division, #301, No: 7-1-414/21, Kaushik Sai Rama Residency, Srinivasa colony (E), S.R.Nagar, Hyderabad-500038, India; Charuvaka Muvva – Email: charvakmuvva@gmail.com; Phone: 040-32584321; *Corresponding author

Abstract:
MicroRNAs (miRNAs) are a novel growing family of endogenous, small, non- coding, single-stranded RNA molecules directly involved in regulating gene expression at the posttranscriptional level. High conservation of miRNAs in plant provides the foundation for identification of new miRNAs in other plant species through homology alignment. Here, previous known plant miRNAs were BLASTed against the Expressed Sequence Tag (EST) database of *Raphanus sativus*, and according to a series of filtering criteria, a total of 48 miRNAs belonging to 9 miRNA families were identified, and 16 potential target genes of them were subsequently predicted, most of which seemed to encode transcription factors or enzymes participating in regulation of development, growth and other physiological processes. Overall, our findings lay the foundation for further researches of miRNAs function in *R.sativus*.

Keywords: MicroRNA, *in silico* Identification, Raphanus sativus, expressed sequence tag, minimal free energy index, Radish

Background:
MicroRNAs (miRNAs) are a novel growing family of endogenous, small, non- coding, single-stranded RNA molecules encoded in the genomes of plants and animals that repress mRNA translation or mediate mRNA degradation in a sequence-specific manner **[1].** The discovery of the first microRNA *lin-4* in *Caenorhabditis elegans* by Ambros laboratory emerged as biology's unusual or unique findings **[2].** These tiny bits of RNA play a major role in gene regulation, which involves in negative regulation of gene targets. In recent years, identification and functional studies of miRNA has made great

progress in research. The exposition of miRNA in plants is still a continuing process hence, till date a number of plant miRNAs have been discovered and functionally identified. Key roles of miRNA in biological processes are revealed by plant studies, which include regulation of leaf development **[3],** stem development **[4],** root development **[5],** signal transduction, developmental timing, floral differentiation and development, and defense response against every sort of stress. Plant miRNAs are usually evolutionary conserved and are observed in regions of the genome distinct from previously annotated genes.

BIOINFORMATION

Different approaches used for miRNA identification includes, gene cloning technology and Bioinformatics strategies. Gene cloning is a conventional method to identify the new miRNA accurately, even though it has disadvantages, such as difficulty in finding miRNAs which express at low levels, difficulty in cloning, degradation of RNA during sample separation etc **[6].** Rapid development in the field of bioinformatics has brought a number of computational programs and other tools to successfully predict the miRNA **[7, 8]**. This process is purely based on the genomic databases like expressed sequence tags EST and other. Since the miRNAs are more conserved in plant species, it is possible to identify novel miRNAs using computational techniques. Now a days miRNAs are identified using the computational or bioinformatics based approach, since it is very useful in predicting the novel miRNA, which cannot be done by cloning. Radish (*R.sativus*) is an edible root vegetable that has been found to possess a number of pharmacological and antioxidant properties. Radish is rich in folic acid, Vitamin C and anthocyanins. These nutrients make it a very effective cancer-fighting food. It is said that radish is effective in fighting oral cancer, colon cancer and intestinal cancer as well as kidney and stomach cancers.

In this study, all previously known plant miRNAs from *A. thaliana,* rice, and other plant species were used to search the *R.sativus* homologs of miRNAs in the publicly available expressed sequence tag (EST) (National Center for Biotechnology Information, NCBI, (http://www.ncbi.nlm.nih.gov/). A total of 48 potential miRNAs were detected. Using these potential miRNAs sequences, *R.sativus* mRNA database was further blasted to find 16 potential miRNA-targeted genes. Most of the target mRNAs were found to be coding transcription factors which are involved in regulating plant growth, development and metabolism.

Methodology:
Datasets of miRNAs, EST and mRNA sequences
To search potential miRNAs in *R.sativus*, previously known plant miRNAs including their precursor sequences from *Arabidopsis thaliana*, *Zea mays*, *Oryza sativa*, *Glycine max, Sorghum bicolor* and other plant species were downloaded from the miRBase (Release 17:April 2011) (http://www.mirbase.org/) **[7]**. After removal of the repeated sequences, 1,876 were considered as reference set *R.sativus* expressed sequence tag (EST) and mRNA database were downloaded from GenBank database (http://www.ncbi.nlm.nih.gov).

Availability of Computational Software
Comparative software BLAST-2.2.24 **[8]** was downloaded from NCBI. BLASTX (http://www.ncbi.nlm.nih.gov/BLAST/) was used for analysis of potential targets. RNA secondary structure and the free energy were calculated by web server Mfold 3.2 **[9].**

Prediction of R.sativus miRNAs
The Computational prediction of miRNAs in *R.sativus* is shown in **Figure: 1.** the plant miRNA sequences were initially scanned to remove repeats. The reference set was subjected to the BLAST **[10]** search for *R.sativus* homologs of miRNAs against EST database. The initial BLAST-2.2.24 search was carried out with default parameters. Mature miRNA sequence should be no less than 18 nt, and the mismatches should be less than 4. Later

these precursor sequences were BLASTXed online to remove the protein coding sequences. Pre-miRNAs secondary structure was run on MFOLD 3.2 (http://www.bioinfo.rpi.edu/). The following steps were considered for screening the candidate miRNA homologs: (1) The RNA sequence folding into an appropriate stem-loop hairpin secondary structure that contains the ~22 nt mature miRNA sequence located in one arm of the hairpin structure; (2) The predicted mature miRNAs with no more than 6 mismatches with the opposite miRNA* sequence in the other arm; (3) maximum size of 3 nt for a bulge in the miRNA sequence was allowed; (4) miRNA precursors with secondary structures had higher negative minimal free energies and minimal free energy index (MFEI) than other different type of RNAs; and MFEI of greater than 0.85 **[11]**; (5) The A+U content of pre-miRNA within 30-70% was considered; (6) no loop or break in miRNA sequences was allowed. These criteria significantly reduced false positives and required that the predicted miRNAs fit the criteria proposed by Ambros and co-workers **[12].**
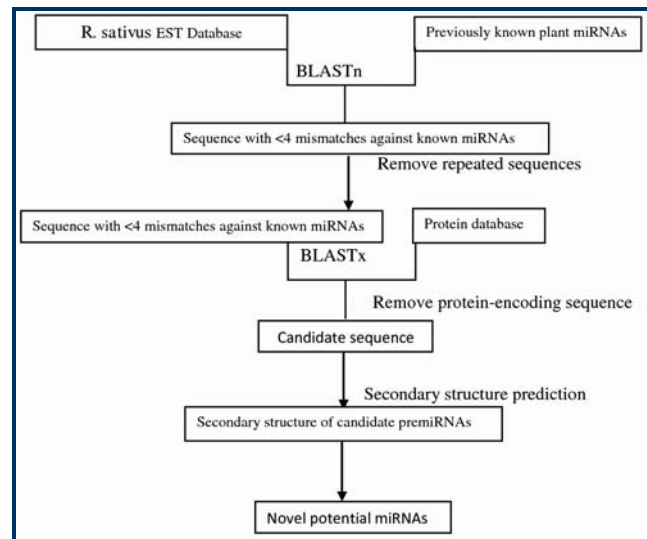

Figure 1: Flowchart of *R.sativus* miRNA prediction

Computational prediction of potential targets of miRNAs
miRNA targets prediction was performed by aligning the predicted miRNA sequences with mRNA sequences of *R.sativus* through the BLAST program. The targets were screened according to these criteria: the number of mismatches should be less than 4, and no gaps were allowed at the complementary sites. After removal of the repeated sequences, the potential target genes were BLASTed against protein databases to predict their function.

Results:
To signify new miRNAs in *R.sativus* by Bioinformatics strategy, the flowchart is shown in **Figure 1**. Setting the default e-value, we began our BLASTn search for homologous miRNA sequences against the EST databases of *R.sativus*. After the blast, all blasted hits with non-coding sequences were retained for analysis of secondary structure; those meeting the criteria, discussed in Methodology were termed as miRNA homologs. Finally, 48 potential *R.sativus* miRNAs belonging to 9 miRNA families were identified and they were named according to Ambros **[12].** The details on predicted *R.sativus* miRNAs

including family names, Sources, mismatches, mature miRNA lengths, precursors, A+U Content, and Minimal folding free energy index (MFEI) were listed in **Table 1 (see supplementary material**. During the screening of the potential miRNAs, candidate miRNAs were evaluated for A+U content. The sequences of the miRNA precursors had A+U content ranging from 31.57% to 61.90% **Table 1,** which is in agreement with the previous results **[13].** The length of the 48 predicted miRNAs ranged from 18nt to 22nt. All the MFEI of these hairpin structures were over 0.85, which was thought to be gold standard to differentiate miRNAs from other ones **[14].** The 48 miRNAs represent 9 miRNA families in *R.sativus* **(Figure. 2).** miRNA 156 has nine members; miRNA 160, 319 and 171 have 7 members; miRNA 396 and 162 have 6 members; miRNA 167, 170 and 399 have 4,1 and 1 members respectively. The current results confirm that the approach for EST analysis is a relatively efficient way to identify miRNAs.
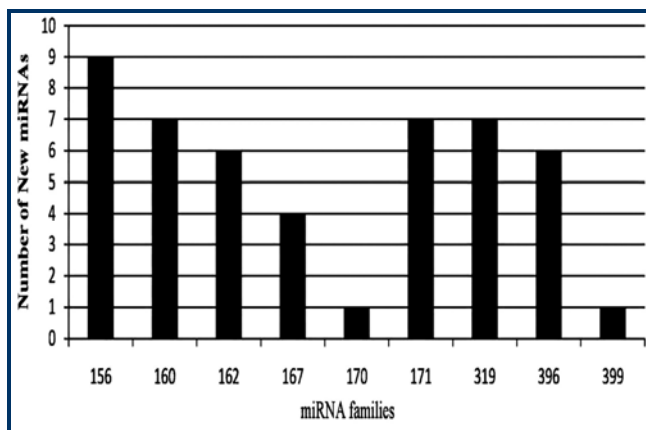

Figure 2: The distribution of new miRNAs in different miRNA families

According to Zhang **[13],** about 10000 ESTs contained one miRNA, so about 15 miRNAs should be predicted theoretically from the total of 150665 ESTs, but in this work we found total 48 miRNAs belonging to 9 families showing the higher value than the previous prediction results and different length of mature miRNAs from the same precursor were regarded as different ones, considering they corresponded to different target genes. Compared to the nucleotide number of animal miRNA precursors (typically with 70–80), the *R.sativus* miRNA precursors show more diversity in structure and size **Table 1.** The length of miRNA precursors in *R.sativus* varied from 77 to 181 nucleotides. The different size of the identified miRNAs within different families suggests that they may offer unique functions for regulation of miRNA biogenesis or gene expression **[13].** The diversity of the identified miRNAs can also be found in the location of mature miRNA sequences. The sequences of miR 156b/d/f/g/h/i, miR 160/b/c/d/e/f, miR 162/a, miR 167b/c/d, miR 170, miR 319c, miR396a/c/d and miR399a were located at the 5′ end of the miRNA precursors, while the miR 156a/c/e, miR160a, miR162b/c/d/e, miR167a, miR 171a/b/c/d/e/f/g, miR319a/b/d/e/f/g and miR 396b/e/f were found at the 3′ end. Based on the complementarity between miRNAs and their target genes in plants, the *R.sativus* EST database was searched for homology to the new miRNA sequences with a BLASTN and BLASTX algorithm for the discovery of miRNA targets. A total of 16

potential targets for *R.sativus* miRNAs were identified. These potential miRNA targets belonged to a number of gene families that had different biological functions **Table 2 (see supplementary material),** including the control of transcription factors, organizing and segregating chromosomes for partition. The miRNAs and their putative targets with known functions are listed in **Table 2.**

Discussion:
In plant kingdom most of the mature miRNAs are evolutionarily conserved from species to species. This information enables us to predict new miRNA homologs or orthologs by insilico method **[13].**Therefore, we used all previously known plant mature miRNAs from miRBase to search for homologs of miRNAs and their target genes in radish in the publicly available EST and GSS database of *R.sativus*. Finally, 48 potential *R.sativus* miRNAs belonging to 9 miRNA families were identified. In the present study, the length of predicted miRNA precursors varies from 77 to 181 NT. The different sizes of the identified miRNAs within the different families suggest that they may perform unique functions in the regulation of miRNA biogenesis or gene expression **[13].** MFEI is an important characteristic that distinguish miRNA from other non-coding and coding RNAs. The MFEI is a unique criterion to designate miRNAs. When the MFEI is more than 0.85, the sequence is most likely to be miRNA. We observed that the MFEIs of the hairpin structures ranged from 1.34 to 3.73 **Table 1**. All the mature sequences of Radish miRNAs are in the stem portion of the hairpin structures, as shown in Figure 3. According to the estimation approximately 10,000 ESTs in plants contain one miRNA **[13]** that means 150,665 ESTs in *R.sativus* should contain 15 miRNAs. But in this study, 48 miRNAs were detected, showing higher value than the previous prediction **(Figure. 3)**. The current results confirm that the approach of EST analysis is a relatively efficient way to identify miRNAs.

To understand the biological function of miRNAs in plant development, it is necessary to identify their targets. In miRNA target prediction, the screening criterion was set according to the description in Methodology. Finally, 16 potential targets for *R.sativus* miRNAs were identified **Table 2.** The miRNA 156 have been predicted to target Squamosa promoter Binding Proteins (SBP) or SBP-like (SPL) genes, **Table 2.** It is identified that, most of the predicted miRNA targets were coding genes for transcription factors mainly involved in the synthesis of enzymes participating in regulation of development, growth and other physiological processes.

The general characteristic of the miRNA sequence is, it is complementary to their target gene, and in some case single miRNA can be complementary to more than one target gene. For example, in this work we identified that the miR156 and miR396 have more regulatory targets, even though their structure and sequence is dissimilar. SMC proteins are recently identified important factors that influence chromosome structure during mitosis and development. SMC proteins were identified very first in the budding yeast through genetic analysis of chromosome segregation. These proteins are known to involve in both prokaryote and eukaryote chromosome structure maintainance **[15].** Thioredoxins (Trx) are proteins which play essential role in the cellular redox reactions. These

are small and molecular weight is approximately 120 KD. Trx are widely distributed in different forms in bacteria, fungi and plants etc. Especially plants have different Trx proteins found in chloroplast Trx f and Trx m, Trx h found in cytosol, ER and

mithochondria **[16].** The findings of this study considerably broaden the scope of understanding the function of miRNA in *R.sativus*.
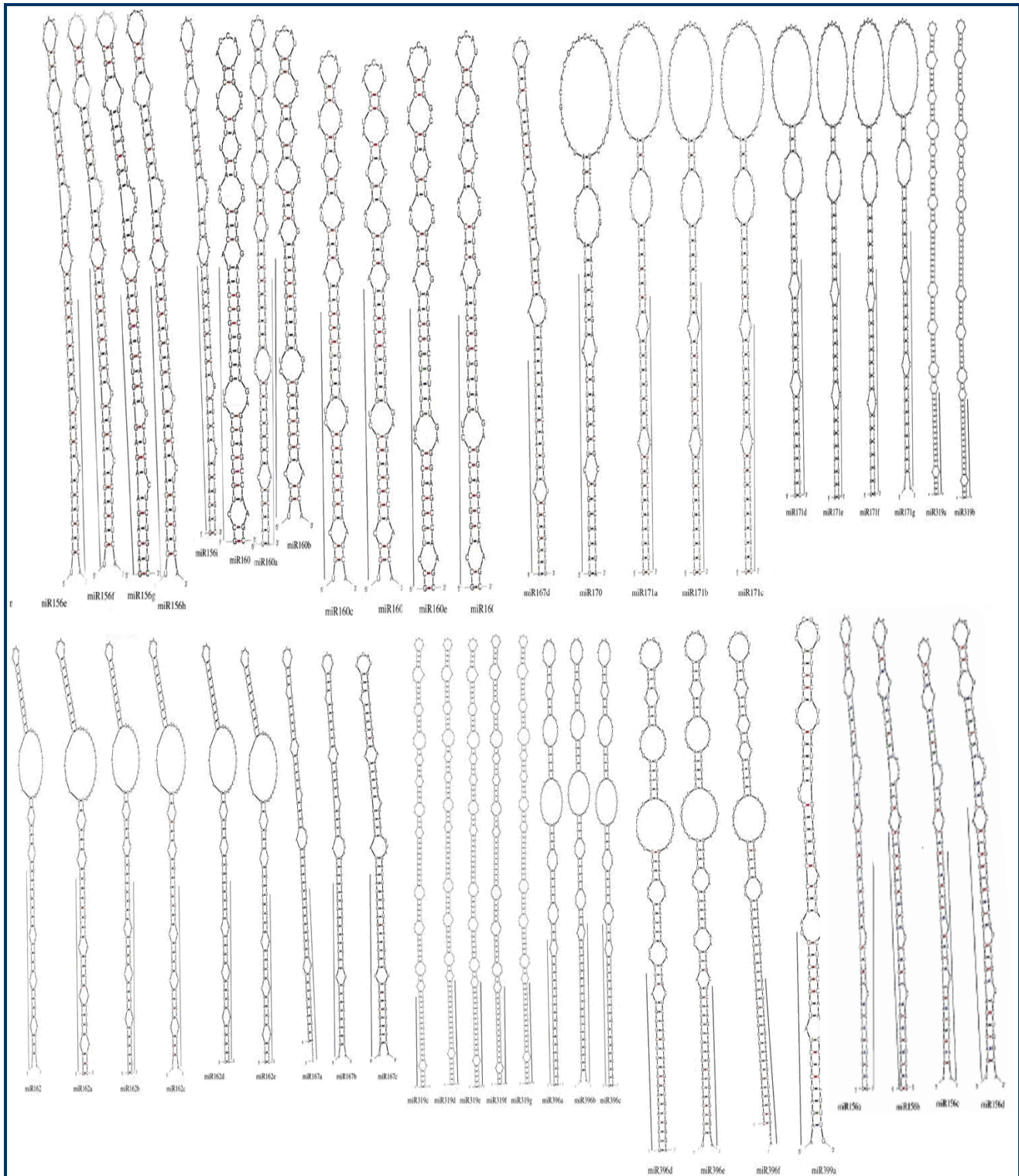

Figure 3: Mature and precursor sequences and the predicted stem-loop structures of newly identified miRNAs in *R.sativus*. The mature miRNAs are underlined.

BIOINFORMATION

Conclusion:

This paper, with a bioinformatics approach, 48 mature miRNAs along with 16 target genes were identified in *R.sativus*. *In silico* studies stand as initial point for understanding miRNAs role in gene regulation. Thus, identification of miRNAs and their target genes help in understanding function and regulatory mechanisms in *R. sativus*.

References:

[1] Unver T *et al. Int J Plant Genomics*. 2009 **2009**: 262463 [PMID:19834623]

[2] Lee RC *et al. Cell*. 1993 **75**: 843 [PMID:8252621]

[3] Palatnik JF *et al. Nature*. 2003 **425:** 257[PMID:12931144]

[4] Mallory AC *et al. Curr Biol*. 2004 **14**: 1035[PMID:15202996]

[5] Guo HS *et al. Plant Cell*. 2005 **17**: 1376 [PMID:15829603]

[6] Zhang BH *et al. Cell Mol Life Sci*. 2006 **63**: 246 [PMID:16395542]

[7] Griffiths-Jones S, *Methods Mol Biol*. 2006 **342**: 129[PMID:16957372]

[8] Altschul SF *et al. Nucleic Acids Res*. 1997 **25**: 3389[PMID: 9254694]

[9] Zuker M. *Nucleic Acids Res*. 2003 **31**: 3406 [PMID: 12824337]

[10] Mathews DH *et al. J Mol Biol*. 1999 **288**: 911 [PMID:10329189]

[11] Zhang BH *et al. Cell Mol Life Sci*. 2006 **63**: 246 [PMID:16395542]

[12] Ambros V *et al. RNA*. 2003 **9**: 277 [PMID:12592000]

[13] Zhang B *et al. Plant J*. 2006 **46**: 243 [PMID: 16623887]

[14] Zhang B *et al. Dev Biol*. 2006 **289**: 3[PMID: 16325172]

[15] Jessberger R *et al. Curr Opin Genet & Dev*. 1998 **8**: 254[PMID: 9610418]

[16] Marcus F *et al. Arch Biochem Biophys*. 1991 **287**: 195[PMID:1897989]

Edited by P Kangueane

Citation: Muvva *et al*. Bioinformation 8(2): 098-103 (2012)

BIOINFORMATION

Supplementary material:

Table 1: Predicted miRNAs of *R.sativus*

New miRNAs	EST ID	Mature sequence	NM	ML	LP	Location	A+U (%)	MFEI
rsa miR156a	332792699	GCUCACUGCUCUCUCUGUCAG	1	21	81	3	42.85	2.08
rsa-miR156b	332792699	CUGACAGAAGAGAGUGAGCAC	2	21	81	5	47	2.08
rsa-miR156c	332792699	GCUCACUGCUCUCUCUGUCAGA	1	22	83	3	45.45	2.16
rsa-miR156d	332792699	UGACAGAAGAGAGUGAGCACA	0	21	79	5	52	1.95
rsa-miR156e	332792699	GCUCACUGCUCUCUCUGUCA	0	20	79	3	45	1.95
rsa-miR156f	332792699	UGACAGAAGAGAGUGAGCAC	0	20	79	5	50	1.95
rsa-miR156g	332792699	GACAGAAGAGAGUGAGCA	0	18	77	5	50	1.87
rsa-miR156h	332792699	UGACAGAAGAGAGUGAGCA	1	19	79	5	52.63	1.95
rsa-miR156i	332792699	GACAGAAGAGAGUGAGCAC	0	19	77	5	47.36	1.87
rsa-miR160	158915937	GCCUGGCUCCCUGUAUGCCAU	0	21	77	5	38.09	2.05
rsa-miR160a	158915937	GCGUAUGAGGAGCCAAGCAUA	1	21	83	3	47	2.25
rsa-miR160b	158915937	UGCCUGGCUCCCUGUAUGCCA	0	21	79	5	38.09	2.14
rsa-miR160c	158915937	UGCCUGGCUCCCUGUAUGCC	0	20	79	5	35	2.14
rsa-miR160d	158915937	UGCCUGGCUCCCUGUAUGCCAU	1	22	79	5	40.90	2.14
rsa-miR160e	158915937	GCCUGGCUCCCUGUAUGCCA	0	20	77	5	35	2.05
rsa-miR160f	158915937	GCCUGGCUCCCUGUAUGCC	1	19	77	5	31.57	2.05
rsa-miR162	167433042	UGGAGGCAGCGGUUCAUCGAUC	0	22	101	5	40.90	1.73
rsa-miR162a	167433042	GGAGGCAGCGGUUCAUCGAUC	0	21	99	5	38.09	1.65
rsa-miR162b	167433042	UCGAUAAACCUCUGCAUCCAGC	0	22	105	3	50	2.03
rsa-miR162c	167433042	UCGAUAAACCUCUGCAUCCA	0	20	101	3	55	1.73
rsa-miR162d	167433042	UCGAUAAACCUCUGCAUCCAG	1	21	103	3	52.38	1.86
rsa-miR162e	167433042	UCGAUAAACCUCUGCAUCC	0	19	99	3	52.63	1.65
rsa-miR167a	158916067	GAUCAUGUUUGCAGUUUCACC	1	21	99	3	57.14	2.25
rsa-miR167b	158916067	UGAAGCUGCCAGCAUGAUCUA	0	21	97	5	52	2.23
rsa-miR167c	158916067	UGAAGCUGCCAGCAUGAUCUAA	0	22	97	5	54	2.23
rsa-miR167d	158916067	AAGCUGCCAGCAUGAUCUA	0	19	93	5	52.63	2.00
rsa-miR170	166132096	UAUUGGCCUGGUUCACUCAGA	0	21	83	5	52.38	1.34
rsa-miR171a	166132096	GAUUGAGCCGCGCCAAUAUC	0	20	87	3	45	1.54
rsa-miR171b	166132096	UGAUUGAGCCGCGCCAAUAUC	0	21	87	3	47.61	1.54
rsa-miR171c	166132096	UUGAGCCGCGCCAAUAUC	0	18	87	3	44.44	1.54
rsa-miR171d	166132096	GAUAUUGGCGCGGCUCAAUC	1	20	87	3	45	1.54
rsa-miR171e	166132096	UGAUUGAGCCGCGCCAAUAU	0	20	85	3	50	1.40
rsa-miR171f	166132096	GAUUGAGCCGCGCCAAUA	0	18	83	3	44.44	1.34
rsa-miR171g	166132096	UGAUUGAGCCGCGCCAAUAUCU	0	22	89	3	50	1.60
rsa-miR319a	332773864	UUGGACUGAAGGGAGCUCC	0	19	177	3	42	3.60
rsa-miR319b	332773864	UUGGACUGAAGGGAGCUC	1	18	175	3	44.44	3.44
rsa-miR319c	332773864	GGAGAUUCUUUCAGUCCA	0	18	177	5	55.55	3.60
rsa-miR319d	332773864	UUGGACUGAAGGGAGCUCCUU	0	21	181	3	47.61	3.73
rsa- miR319e	332773864	UUUGGACUGAAGGGAGCUCCU	0	21	179	3	47.61	3.68
rsa-miR319f	332773864	UUUGGACUGAAGGGAGCUC	0	19	175	3	47.36	3.44
rsa-miR319g	332773864	UUGGACUGAAGGGAGCUCCU	0	20	179	3	45	3.68
rsa-miR396a	166139020	UUCCACAGCUUUCUUGAAC	1	19	122	5	57	1.79
rsa-miR396b	166139020	GCUCAAGAAAGCUGUGGGAAA	0	21	126	3	52.38	1.88
rsa-miR396c	166139020	UUCCACAGCUUUCUUGAACUU	0	21	122	5	61.90	1.79
rsa-miR396d	166139020	UUCCACAGCUUUCUUGAACU	0	20	122	5	60	1.79
rsa-miR396e	166139020	UCAAGAAAGCUGUGGGAA	0	18	124	3	55.55	1.84
rsa-miR396f	166141534	GUUCAAGAAAGCUGUGGAA	0	19	120	3	57.89	2.53
rsa-miR399a	166146391	AAGAUCUCUAUUGGCAGG	0	18	89	5	55.55	2.08

NM=Number of Mismatch, ML=Length of Mature miRNAs, LP=Length of Precursor, MFEI=Minimal free energy index

Table 2: Potential targets of the identical miRNAs in *R.sativus*

miRNAs	Targeted gene	Targeted protein	Possible function
miR156	EY930450	squamosa promoter-binding protein	transcription factor
	EW715846	squamosa promoter-binding protein	transcription factor
	EX898839	squamosa promoter-binding protein	transcription factor
	FD950281	SMC (structural maintenance of chromosomes) protein	organizing and segregating chromosomes for partition
miR167	FY434046	DNA binding protein	Transcription Factor
	DN910345	DNA binding protein	Transcription Factor
	DY64102	DNA binding protein	Transcription Factor
miR396	EW713403	Thioredoxin super family	Transcription factor
	EW713404	Thioredoxin super family	Transcription factor
	EW718331	Thioredoxin super family	Transcription factor
	EW718591	Thioredoxin super family	Transcription factor
	EW733484	Thioredoxin super family	Transcription factor
	FD563793	Thioredoxin super family	Transcription factor
	FD941766	Thioredoxin super family	Transcription factor
	FD939455	Thioredoxin super family	Transcription factor
	FY442946	Thioredoxin super family	Transcription factor