BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(2)

# Computational identification and characterization of putative miRNAs in *Heliothis virescens*

**Poonam Chilana, Anu Sharma*, Vasu Arora, Jyotika Bhati & Anil Rai**

Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Library Avenue, Pusa, New Delhi - 110012; Anu Sharma - Email: anu@iasri.res.in; *Corresponding author

Received November 16, 2012; Accepted November 24, 2012; Published January 18, 2013

**Abstract:**
*Heliothis virescens*, a polyphagous pest, is one of the most destructive pests of many crops and vegetables. Various insecticides and pesticides are used by agriculturalists to stop the growth and development of this pest. RNA interference is a new area for the management of pests / insects by inhibiting the growth related RNAs. This involves the miRNAs identification and its characterization. In the present study, computational approach is applied to predict putative miRNA candidates along with their possible target(s) in the *Heliothis virescens*. A total of 63,662 ESTs were downloaded from dbEST database and processed, trimmed and masked through EGassembler. The *H. virescens* contigs database obtained after assembly was now used to find the putative miRNA candidates by performing a local BLAST with the miRNAs of insects retrieved from miRBase. We have predicted putative miRNA candidates by homology search against all the reported insect miRNAs. These putative miRNAs candidates were further validated and filtered by different features. In addition, we have also attempted to predict the putative targets of these filtered miRNAs, by making use of 3' untranslated regions of mRNAs from *B. mori*. These miRNAs and their targets in *H. virescens* will help in improved understanding of molecular mechanisms of miRNA and development of novel and more precise techniques for better understanding some post transcriptional gene silencing.

**Background:**
The tobacco budworm (TBW), *Heliothis virescens* (F.), a pest which is responsible for substantial economic loss, environmental pollution therefore management of this is a great challenge to environment, researchers, cotton and tobacco producers etc. since decades **[1]**. Tobacco budworm is a polyphagous field crop pest, for crops such as alfalfa, clover, cotton, flax, soybean, and tobacco. However, it also attacks vegetables such as cabbage, lettuce, pea, pepper, pigeon pea, squash, and tomato. Recent attempts of development of insect resistance to transgenic crops provided new novel biotechnological solutions such as RNA interference (RNAi), gene silencing for pest management during 21st century. Recently, RNA interference (RNAi) efforts to identify endogenous small RNAs have led to the discovery of hundreds of miRNAs in nematodes, fruit flies and humans **[2-4]**. These small non-coding genes are typically transcribed by RNA polymerase II, processed into hairpins, and exported into the cytoplasm, where they are cleaved by the central enzyme of the RNAi pathway, Dicer, to form single-stranded mature microRNAs **[5, 6]**. MicroRNAs (miRNAs) are small endogenous RNA molecules (~21-25 nt) that regulate gene expression by targeting one or more mRNAs for translational repression or cleavage. The first two miRNAs (lin-4 and let-7) were identified from *Caenorhabditis elegans* and discovery of miRNA from various organisms has since accelerated with 21,264 miRNA known by Aug 2012 **[7, 8]**. Further, miRNAs are generally conserved in closely related species but also conserved in different taxonomic group. For example about 10% of miRNAs identified in invertebrates are also conserved in mammals and other higher animals, suggesting cross-species conservation **[9, 10].** In the recent years, with the availability of whole genome sequence data, linkage groups, expressed sequence tags (ESTs) and various genetic markers, research on insects miRNAs has extended gradually from *D. melanogaster* **[11, 12]** to other model insects, such as *Bombyx mori* **[13]**. The miRNA from *Apis mellifera* of order Hymenoptera and *Anopheles gambiae* of order Diptera have been predicted and submitted to the miRNA

 

registry miRBase **[14]**. The order Hymenoptera also includes natural enemies of a broad range of vector arthropods which are of medical, veterinary and agricultural significance. *Nasonia*, a parasitic wasp, is emerging as model for studies of complex genetic traits. It is well positioned phylogenetically to assist in identifying orthologs of important genes in insects and a genetically traceable system for functional analysis. Therefore, Sathyamurthy and Swamy (2010) identified putative miRNA gene sequences and predicted their possible targets in *N. vitripennis* species **[15]**. Singh and Nagaraju **[16]** attempted to predict miRNA from the important agricultural pest, *Tribolium castaneum* of order Coleoptera for which no data is available till date. Although, rapid progresses have been achieved in discovering new miRNAs and exploring their biological roles in model insects, studies on miRNAs in agricultural pest is still very slow. Keeping in view the importance of miRNAs in insects, we demonstrate a computational approach to predict putative miRNA candidates along with their possible target (s) in the polyphagous pest, *Heliothis virescens*.

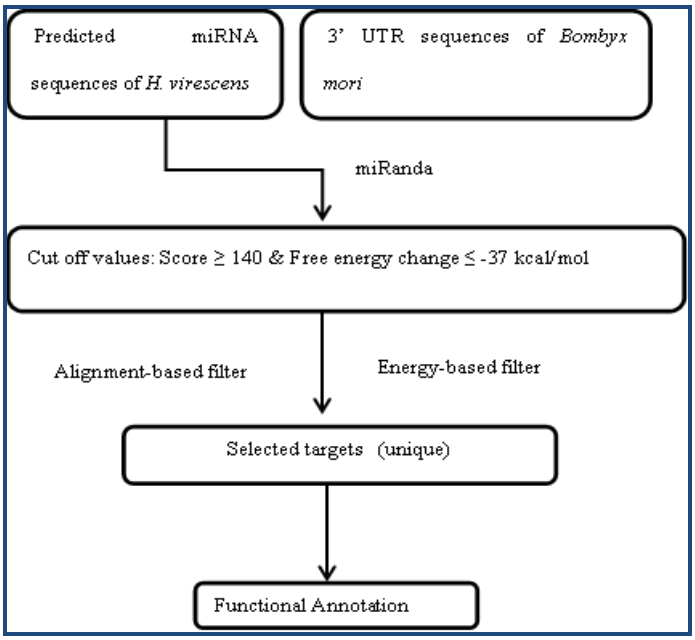


**Figure 1:** Steps involved in microRNA targets prediction in *Heliothis virescens*.

## Methodology:

### *EST mining and pre-processing*

A total of 63,662 ESTs of *Heliothis virescens* were downloaded from NCBI website (http://www.ncbi.nlm.nih.gov/est/). The sequence redundancy was removed using the sequence assembly program, EGassembler (http://egassembler.hgc.jp/). The program clustered the ESTs containing overlapping sequences as contigs and non-overlapping sequences as singletons. After removal of the repeated sequences, 63,314 were considered as reference set of *H. virescens* expressed sequence tag (EST).

### *Prediction of miRNAs by homology search*

In order to search potential miRNAs in *H. virescens*, previously known insect miRNAs including their precursor sequences were downloaded from the miRBase **[17]**. A BLASTn search of all the 3385 miRNA sequences of all insects with the EST sequences of *H. virescens* was first carried out with the e-value < 0.01 along with default parameters including low complexity filter. With the same parameters BLASTn search is carried out between pre miRNAs of insects and the match results of EST-miRNA blast. The two criteria used for screening the BLAST results were: (1) more than 90% identity between each potential *H. virescens* miRNA and the corresponding miRNA in the reference set (known miRNA homologue); (2) the length difference between each potential *H. virescens* miRNA and the corresponding miRNA in the reference set is not more than three bases.

### *Secondary structure validation*

Pre miRNA sequences were extracted using a sliding window of about 100nt in size (moving in increments of approximately 10nt) from the region ~80nt upstream of the beginning of the mature miRNA to ~80nt downstream of the miRNA. Extracted miRNA precursor sequences were then submitted to Mfold (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi) for checking of the fold-back secondary structure. The four criteria used for selecting pre miRNA structures were: (1) The RNA sequence folding into an appropriate stem-loop hairpin secondary structure that contains the ~22 nt mature miRNA sequence located in stem region of the hairpin structure; (2) maximum size of 7 nt for a bulge in the miRNA sequence was allowed; (3) miRNA precursors with secondary structures should have free energy change ($\Delta G$) less than or equal to –37kcal/mole; (5) no loop or break in miRNA sequences was allowed. These criteria significantly reduced false positives and ensure that the predicted miRNAs fit the criteria proposed by Ambros and coworkers **[18]**.

### *Identification of putative candidate miRNA sequences*

In order to distinguish the real pre-miRNAs from other hairpin sequences with similar stem-loops (pseudo pre-miRNAs), we explored with *MiPred* which decides whether it is a pre-miRNA-like hairpin sequence or not. If the sequence is a pre-miRNA-like hairpin, the RF classifier will also predict whether it is a real pre-miRNA or a pseudo one (http://www.bioinf.seu.edu.cn/miRNA/) **[19]**.

### *Target prediction using miRanda program*

In animals, employing computational approaches to identify miRNA are quite challenging because animal miRNAs are partially complementary to their target mRNAs, whereas, plants, miRNAs bind their targets by complete or nearly complete complementarity **[20-23]**. The primary target of miRNA is 3′UTRs **[24, 25].** We employed the miRanda program **[26]**, which utilizes thermodynamics and dynamic-programming alignments, along with statistical parameters, for target prediction in *H. virescens*. The parameters assigned for miRanda hybridization were default alignment score greater than or equal to 80, MFE of miRNA::mRNA duplex less than or equal to –37kcal/mol and the other parameters were kept as default **[27, 28]**. The different steps involved in this target prediction are shown in **(Figure 1)**. We also considered other stringent filters for screening targets to minimize the background matches, thus ensuring the least false positives.

## Results & Discussion:

### *Prediction of miRNAs*

The different steps involved in miRNA prediction are shown in **(Figure 2).** A BLASTn search of all the known mature miRNAs

from insect group **[7]** (miRbase Release: 9.2) against the EST sequences of *H. virescens* resulted in hits. These hits were subsequently scanned for their precursor sequences by taking a sliding window of about 100 nt (moving in increments of approximately 10 nt) from the region ~80 nt upstream of the beginning of the mature miRNA to ~80 nt downstream of the miRNA. The characteristic secondary structures of all of the 4 miRNA precursors were determined by the MFold program **[29]**, which computes the minimum free energy (MFE) contribution for various possible secondary structures. Further, miRNA precursor structures having an MFE less than –37 kcal/mol or a bulge size more than 7 bp or mature miRNA located on the loop region were excluded.

The four predicted miRNA in the present study are hvi-miR-750, hvi-miR-750-5p, hvi-miR-6497-5p and hvi-miR-6497-3p. The details on predicted *H. virescens* miRNAs including mature miRNA sequence, source contig and segment lengths, strand, % identity and A+U content are given in **Table 1 (see supplementary material)**.

***Target prediction***

Prediction of miRNA targets provides an alternative approach to assign biological functions. Since, high-throughput experimental methods for microRNA target identification have not been published yet, computational methods that try to identify target sites based on their partial complementarity with microRNAs have become increasingly important. For each of the validated miRNA–target pairs, functional target sites are located in the untranslated regions (UTR) of the mRNA and are conserved in the UTRs of the homologous genes from related species.

These UTRs have already been recognized as an important regulatory region even before the discovery of miRNAs, due to the presence of numerous regulatory signals involved in the control of nuclear export, subcellular localization, and transcript stability amongst other processes which contains multiple target sites more than one miRNA to interact **[9].** It is well known that animal miRNA targets are difficult to predict, unlike plant targets since miRNA: mRNA duplexes often contain several mis-matches, gaps and G+U base pairs in many positions **[16, 30, 31 & 32].** In the present study, pairwise comparison of the 1630 UTRs of *B. mori* the closest homologue of *H. virescens* against 4 mature miRNA of *Heliothis virescens* has been conducted. MiRanda algorithm, 32 which encompasses the thermodynamic stability of miRNA:mRNA duplex as one of the entity in detecting the potential binding site on the 3`UTRs has been used. We observed 6 potential targets from *hvi-miR-750* putative miRNA targeting different genes as shown in **Table 2 (see supplementary material)**. These potential targets are rich in genes that are expressed at specific developmental stages and that are involved in cell fate specification, morphogenesis and the coordination of developmental processes, as well as genes that are active in the mature nervous system. The predicted miRNAs revealed target multiplicity; hvi-miR-750 was found to have a maximum of 6 targets. In animals, cooperative binding of one or several distinct miRNAs on a single target gene is reported to be important for the functionality of miRNA-mediated gene regulation **[16, 26 & 33]**. As *Heliothis virescens* genome annotation is still under the way, so these predicted miRNA and their targets reported in the present study constitute an asset for further validation. Further, experimental evidences are required to validate these targets in *in vivo* conditions which are beyond the scope of our study.

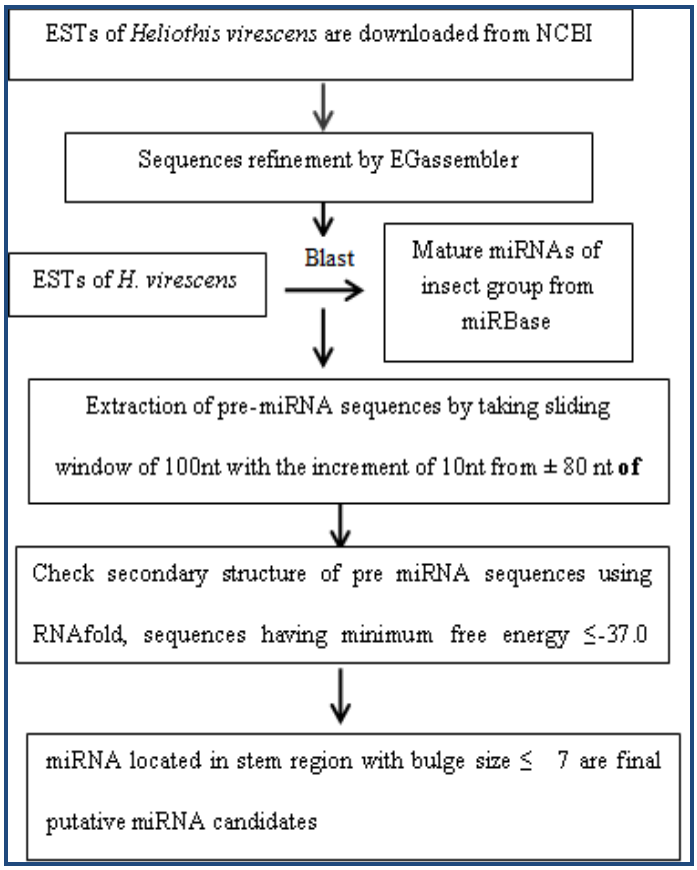


**Figure 2**: Steps involved in microRNA prediction in *Heliothis virescens.*

**Conclusion:**

Four novel putative miRNA are identified from *H. virescens* from ESTs sequences based on homology search. Their targeted proteins are also identified. These finding also strengthens the bioinformatics approach for new miRNAs identification from insect species whose genome is not yet sequenced. The ESTs based identification also confirmed the miRNAs expression. This approach holds great promise for the future as it allows a wide range of potential targets for suppression of gene expression in the insect. Additional genetic /molecular studies will be needed to understand whether miRNAs typically regulate only a handful of key targets or co-ordinately regulate multiple targets which are equally important. These miRNAs and their potential targets in *H. virescens* will help in improved understanding of molecular mechanisms of miRNA and development of novel and more precise techniques for better understanding of post-transcriptional gene silencing.

**Reference:**

**[1]** Blanco CA, *GM Crops Food.* 2012 **3**: 201 [PMID: 22892654]
**[2]** Moar WZ, *Nat Biotechnol.* 2003 **21**: 1152 [PMID: 14520396]
**[3]** Zeng Y & Cullen BR, *RNA.* 2003 **9:** 112[PMID: 12554881]
**[4]** Price DRG & Gatehouse JA, *Trends Biotechnol.* 2008 **26**: 393 [PMID: 18501983]
**[5]** Ambros V, *Nature.* 2004 **431**: 350 [PMID: 15372042]
**[6]** Bartel DP, *Cell.* 2004 **116**: 281 [PMID: 14744438]

**[7]** Kozomara A & Griffiths-Jones S, *Nucleic Acids Res.* 2011 **39**: D152 [PMID: 21037258]
**[8]** Griffiths-Jones S *et al. Nucleic Acids Res.* 2008 **36**: D154 [PMID: 17991681]
**[9]** Stark A *et al*. *PLoS Biol.* 2003 **1**: E60 [PMID: 14691535]
**[10]** Weber MJ, *FEBS J.* 2005 **272**: 59 [PMID: 15634332]
**[11]** Enright AJ *et al*. *Genome Biol*. 2003 **5**: R1 [PMID: 14709173]
**[12]** Lai EC *et al*. *Genome Biol*. 2003 **4**: R42 [PMID: 12844358]
**[13]** Zhang Y *et al*. *PLoS One.* 2009 **4**: e4677 [PMID: 19262741]
**[14]** Chen X *et al*. *Insect Mol Biol.* 2010 **19**: 799 [PMID: 20807255]
**[15]** Sathyamurthy G & Ramachandra SN, *Int J of Insect Sci.* 2010 **2**: 7
**[16]** Singh J & Nagaraju J, *Insect Mol Biol.* 2008 **17:** 427 [PMID: 18651924]
**[17]** Griffiths-Jones S *et al*. *Nucleic Acids Res.* 2006 **34**: D140 [PMID: 16381832]
**[18]** Ambros V *et al*. *RNA.* 2003 **9**: 277 [PMID: 12592000]
**[19]** Peng Jiang *et al*. *Nucleic Acids Res.* 2007 **35**: W339 [PMID: 17553836]
**[20]** Jones-Rhoades MW & Bartel DP, *Mol Cell.* 2004 **14**: 787 [PMID: 15200956]
**[21]** Bartel B & Bartel DP, *Plant Physiol*. 2003 **132**: 709 [PMID: 12805599]
**[22]** Vella MC *et al*. *Genes Dev.* 2004 **18**: 132 [PMID: 14729570]
**[23]** Rajewsky N, *Nat Genet.* 2006 **38**: S8 [PMID: 16736023]
**[24]** Brennecke J *et al*. *Cell.* 2003 **113**: 25 [PMID: 12679032]
**[25]** Lin SY *et al*. *Dev Cell.* 2003 **4**: 639 [PMID: 12737800]
**[26]** Enright AJ *et al*. *Genome Biol.* 2003 **5**: R1 [PMID: 14709173]
**[27]** Smith TF & Waterman MS, *J Mol Biol*. 1981 **147**: 195 [PMID: 7265238]
**[28]** Wuchty S *et al*. *Biopolymers.* 1999 **49**: 145 [PMID: 10070264]
**[29]** Zuker M & Stiegler P, *Nucleic Acids Res.* 1981 **9**: 133 [PMID: 6163133]
**[30]** Wightman B *et al*. *Cell*. 1993 **75**: 855 [PMID: 8252622]
**[31]** Moss EG *et al*. *Cell.* 1997 **88**: 637 [PMID: 9054503]
**[32]** Pasquinelli AE *et al*. *Nature.* 2000 **408**: 86 [PMID: 11081512]
**[33]** John B *et al*. *PLoS Biol.* 2004 **2**: e363 [PMID: 15502875]

**Edited by P Kangueane**
**Citation: Chilana *et al*.** Bioinformation 9(2): 079-083 (2013)

BIOINFORMATION

open access

## Supplementary material:

**Table 1:** Details of genomic locations and sequences of 4 predicted microRNAs (miRNAs) in *Heliothis virescens*

| S. No | New miRNAs | Source miRNA | Contig | Segment | Mature sequence | Strand | Identity (%) | ML | LP | A+U (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | hvi-miR-750 | bmo-miR-750 | Contig2791 | 698-719 | CCAGATCTATCTTTCCAGCTCA | +/+ | 100 | 22 | 100 | 45 |
| 2. | hvi-miR-750-5p | bmo-miR-750-5p | Contig2791 | 659-680 | AGTTGGACAGGGGATCTTGACA | +/+ | 100 | 22 | 100 | 44 |
| 3. | hvi-miR-6497-5p | bmo-miR-6497-5p | Contig3971 | 67-88 | GCTCTGAGGACCGGGGCGTGTC | +/+ | 100 | 22 | 100 | 29 |
| 4. | hvi-miR-6497-3p | bmo-miR-6497-3p | Contig3971 | 107-124 | GATGCGGCCGGTGCCGGG | +/+ | 100 | 18 | 124 | 26.83 |

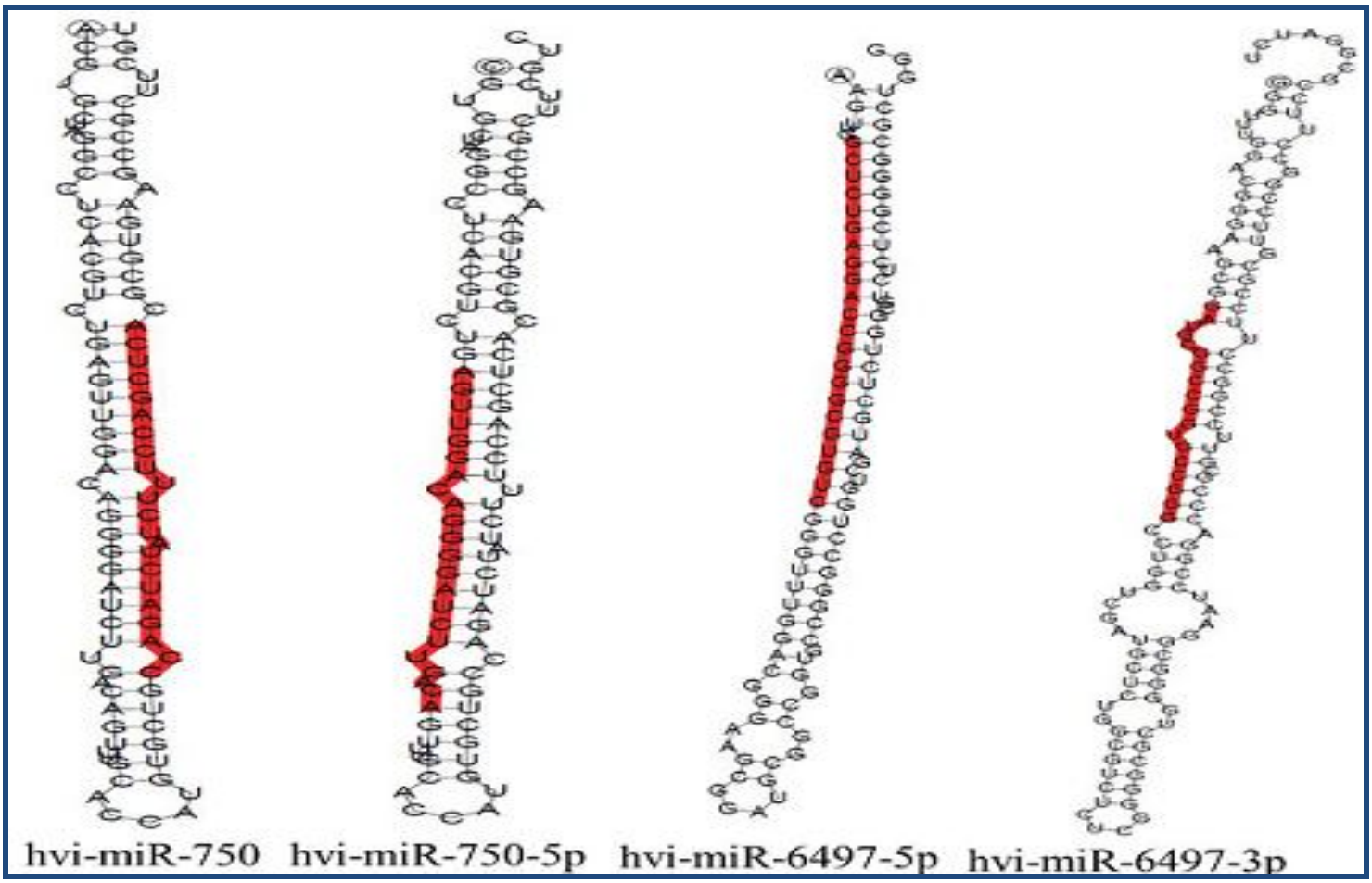


**Table 2:** Details of microRNA targets and their putative functions in *Heliothis virescens*

| S. No. | miRNA Name | Target Acc. | Target Gene | Target Description |
|---|---|---|---|---|
| 1. | hvi-miR-750 | NM_001046766.1 | anaphase promoting complex subunit 11 | anaphase promoting complex subunit 11 |
| 2. | hvi-miR-750 | NM_001044015.1 | chitinase | chitinase-like protein |
| 3. | hvi-miR-750 | NM_001046807.1 | Rak1 | receptor for activated protein kinase C RACK 1 |
| 4. | hvi-miR-750 | NM_001046837.1 | guanylate kinase GMPK | Guanosine monophosphate kinase (GMPK, EC 2.7.4.8),also known as guanylate kinase (GKase), catalyzes the reversible phosphoryl transfer from adenosine triphosphate (ATP) to guanosine monophosphate (GMP) to yield adenosine diphosphate (ADP) and guanosine |
| 5. | hvi-miR-750 | NM_001043866.2 | EcR / NR1H1 | 20-hydroxy-ecdysone receptor |
| 6. | hvi-miR-750 | NM_001046805.1 | ras small monomeric GTPase Rab6 | RAB6A, member RAS oncogene family |