International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 7, No. 4 (2018): 1002-1017
Research Article
Open Access

ISSN: 2320-0634

A computational approach to study the ESTs, miRNA and SNPs in *Antheraea mylitta*

Anjali Singh Deo[1], Sneha Prasad[1], Mousam Kumar Ram[1], Ajit Kumar Sinha[2], Jay Prakash Pandey[2] and Dev Mani Pandey[1,*]

[1] *Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India.*
[2] *Central Tasar Research & Training Institute, Piska Nagari Ranchi-835303, Jharkhand, India.*

*Corresponding author: Dev Mani Pandey; email: dmpandey@bitmesra.ac.in

Received: 05 July 2018 Accepted: 10 August 2018 Online: 25 August 2018

ABSTRACT

Antheraea mylitta is a most economical wild type non-mulberry silkworm in India that produces the exotic variety of tasar silk. The whole genome sequence of *A. mylitta* is not yet available and also the numbers of ESTs are varying with time. Study on ESTs and SNPs in *A. mylitta* are still very less as compared to other species. miRNAs are about 18-22 nucleotides long noncoding small endogenous RNAs that have important role in various biological and metabolic processes. Presently not a single miRNA has been reported in *A. mylitta* that is most economical species for sericulture in India. In this study to predict the potential miRNA of *A. mylitta,* EST based homology search was applied. Here EST sequences retrieved from WildSilkBase were subjected to BLAST analysis against previously known *Bombyx mori* miRNAs. Six potential miRNAs with suitable fold back structures were identified adopting a series of filtration criteria *i.e* using BLAST and RNA structure prediction tools. The target genes of *A. mylitta* miRNAs were predicted depending on their sequence complementarities. The identified target genes mainly encode transport inhibitor like protein, transcription factor, immune response proteins. Identified target genes also exhibited important role in cellular, molecular as well as various biological processes. The prediction of SNPs in the identified miRNA was also performed to analyze the variations in the sequences. RNAsnp Web Server was used to identify the SNPs resulting the prediction of sixteen SNPs in five miRNA sequences. Gene ontology of miRNA target genes was also performed indicating their involvement in cellular components, biological processes, molecular functions, species distribution as well as percentage gene distribution.

Keywords: *Antheraea mylitta,* Expressed Sequence Tag, miRNA, Single Nucleotide Polymorphism, Gene Ontology, miRNA targets.

1. INTRODUCTION

The natural silks have been broadly classified as mulberry and non-mulberry. Non-mulberry sericulture is universally known as forest or wild sericulture. *A. mylitta* is one of the most economically important species for sericulture in India that produces the unique variety of tasar silk [1]. *A. mylitta* is an widespread insect species that produces the world famous tasar silk. Non-mulberry sericigenous creatures belongs to the family *Saturniidae* (superfamily *Bombycidea*) are mostly wild silkmoths. Thangavelu and Sinha [2] reported that *A. mylitta* species is wild in nature and spread in different geographical regions of the Indian sub-continent ranging between 60–88° longitude and 16–24° latitudes. These are in dense, humid tropical forest of central and southern plateau of India. Nearly 95% of the global production of non-mulberry silks is tasar.

Tasar silk fiber has its own distinct color and abrasive to feel, but higher tensile strength, elongation capaicty, and stress relaxation values as compared with

mulberry silk fiber secreted by *Bombyx mori*. These characteristics have made tasar silk as competent and desirable silk similar to mulberry silk. Total 34 eco-races of *A. mylitta* has been reported so far [3]. Now detailed generic analysis of Indian tasar silk moth (*A. mylitta*) populations and ecoraces has also been described elsewhere [4]. Four major stages in the life cycle of *Saturniidae* silkworm namely egg, larva, pupa and adult (moths) has been described. It has been described that Tasar silkworm cocoons have the highest capacity of silk production among all other known non- mulberry silk producing insects [5]. It *h*as been reported that *Terminalia tomentosa* and *Terminalia arjuna* are primary food plants for the *A. mylitta* larvae. And now jamun *(Syzygium cumini* L.) is being considered as another potential host of tropical tasar silkworm i.e. *A. mylitta* drury.

Now the importance of the silkworm as a source for genetic and bio-material has been increased. And *in silico* study on sequence similarities, conserve domain and 3D structure of cocoonase has also been studied previously [6]. A report on current and future perspective of cocoonase enzyme indicating its all possible role in tasar industry has also been published [7]. The whole genome sequence of the tasar silkworm *A. mylitta* is not yet available. Despite the economical and ecological importance of *A. mylitta*, only 1412 EST sequences are available from a cDNA library created from the fat bodies of bacteria infected *A*. *mylitta* larvae [8].

Expressed Sequence Tags (ESTs) are short DNA sequences (with 300 – 500 bp) that correspond to a fragment of a complimentary DNA (cDNA) molecule that may be expressed in a cell at a particular given time. ESTs are short DNA molecules that reverse-transcribed from a cellular mRNA population [9]. ESTs are being used as a fast and efficient method for profiling of genes expressed either in various tissues, cell types or developmental stages [10]. These ESTs are partial cDNA sequences of expressed genes that have been cloned into a plasmid [10, 11]. The knowledge of Lepidopteran biology and genetics is most advanced in the mulberry silkworm *B. mori* due to the availability of extensive genome and EST sequence data [12, 13].

An EST database (WildSilkBase) for wild silkmoths has also been constructed [14]. ESTs may be utilized for gene discovery in species that lack the draft of whole genome sequence [11], and many important genes have already been predicted by EST analysis [15-17]. ESTs have applications in the discovery of new genes, mapping of the genome and identification of coding regions in genome sequences.

MicroRNAs (miRNAs) are newly recognized and widespread class of endogenous non-coding small RNAs that are involved in the negative regulation of nearly all biological and metabolic processes [18]. miRNAs play very important roles in post-transcriptional gene regulation by degradation of target mRNAs or by repression of targeted gene translation in animals, plants, and fungi [19-23]. miRNAs are a class of endogenous small regulatory RNAs of about 18–22 nucleotides derived from their precursor sequences that negatively regulate gene expression at the post-transcriptional levels either by binding to target mRNAs for initiating mRNA cleavage or inhibition of mRNA translation [24]. miRNA expression is negatively correlated with their mRNA targets. Along with other non-coding RNA families, miRNA are of great importance in studying the detailed gene expression [25].

Computational approaches have been successfully applied to predict miRNA and their target genes in vertebrates [26, 27], insects [28], *Arabidopsis* and rice [29-33]. Earlier directional cloning and sequencing were adopted for miRNA identification, however, recently high-throughput sequencing becomes a very much effective stratetgy for miRNA prediction. Presently, prediction of novel miRNAs from the larger pool of small RNA sequences is still a great of challenge. Expressed Sequence Tag (EST) and a genome survey sequence (GSS) approach has been developed to predict miRNAs [34] either from plants or various animals.

Silkworm genome has been reported [12] that provides opportunities for detailed survey of miRNAs and their functionality. Till now only a few silkworm miRNAs have been deposited in miRBase (version 13.0; http://microrna.sanger.ac.uk) that belongs of *B. mori,* and the collection is very far to complete as compared with other model insects such as *Drosophila*. There are experimental evidences that miRNAs has important role in different biological and metabolic processes in plants as well as animals [35, 36].

Single Nucleotide Polymorphism (SNP) indicates a DNA sequence variation of a single nucleotide- A-T-C or G- in the genome of different members of a species. SNPs may be found in the coding, non-coding and intergenic regions of the genome. SNP analysis has varying applications in genetic linkage mapping and fine-mapping of candidate regions to determine haplotypes associated with traits of interest. Variations in the DNA sequences of any organisms can cause diseases in their body. SNPs study is also significant for crop and livestock breeding programs and SNP can cause the Mendelian disease. Cells sometimes make mistakes during the copying process that lead to variations in the DNA sequence at particular locations (SNPs). Earlier, genomic DNA based molecular markers were mainly used [37]. However, SNPs has been the excellent markers for association mapping of genes that are related in the controlling of the complex traits and provide the highest map resolution [38, 39]. The SNPs might occur in the coding and the non-coding sequences including the miRNAs. Polymorphisms in miRNA genes could potentially alter various biological processes by influencing the processing and/or target selection of miRNAs. The severity of illness and the

way body responds to treatments are also manifestations of genetic variations.

Functional genomics, advancements of computational tools and uses of bioinformatics approaches helps in silkworm biology for identifying genes involved in synthesis and other processes. Currently biologists are exploiting the uses of ESTs for prediction of miRNAs and SNPs in their research. Advancement of sequencing techniques and molecular biology research has generated huge amount of data. However, uses of these data for the meaningful biological information are scanty. Therefore, in current study we have used *A. mylitta* ESTs for finding the miRNAs and SNPs by applying various computational approaches.

2. MATERIALS AND METHODS

2.1 Retrieval of Sequences

Antheraea mylitta EST sequences were obtained from the WildSilkBase (http://www.cdfd.org.in/wildsilkbase/ESTorg=mylitta) database. A total of 1412 EST sequences of *A. mylitta* were retrieved. To find the probable miRNAs in *A. mylitta*, previously known 159 mature miRNA sequences of *Bombyx mori* were retrieved from miRNA Registry Database (miRbase-Release 18.0, May, 2012) [40], as miRNA sequences available in the database for silk worm was only for *B. mori*. Above retrieved miRNAs were named as reference set of miRNA sequences.

2.2 Identification of potential miRNAs

The reference miRNA sequences from *B. mori* were used as a query sequence for homology search against retrieved *A. mylitta* EST sequences. The mature sequences of all miRNAs from *B. mori* were used for BLASTn analysis [41] against the *A. mylitta* EST sequences. After BLASTn analysis total 3895 sequences were obtained. From these obtained sequences finally 27 EST sequences were found having minimum E-value (i.e. 8e-55). It is well described that lower the E – value, higher will be the matching score. Obtained EST sequences were considered as probable miRNA candidates only if they fulfill the following criteria with slight modification from a previous study [36]. These EST sequences were subjected to BLASTx and tBLASTn analysis using NCBI and WildSilkbase database against non-redundant Protein database for removing the protein-coding sequences and retaining only the non-protein coding sequences [41]. Total 21 EST sequences were predicted that belongs to non-protein-coding sequences.

2.3 Prediction of secondary structure

The secondary structure of candidate pre-miRNA sequences of potential miRNA homologs (having 100-140 bases) was further predicted using Zuker folding algorithm, MFOLD -3.2 [42] with the default settings and outputs of MFOLD-3.2 were recorded. Subsequently, adjusted minimal folding free energy (AMFE) and the minimal folding free energy index (MFEI) were calculated according to the following equation:

AMFE = (MFE/length of a potential pre-miRNA)*100

MFEI = ((100 * MFE)/Length of RNA/ (G + C))%

ESTs fulfilling the criteria described by Wang et al. [43] were considered as a potential miRNA.

2.4 Identification of SNPs in potential miRNAs

The selected potential miRNAs were analyzed for the prediction of SNPs using the RNAsnp Web Server. The parameters were set, mode to screen putative structure disruptive SNP. To filter SNPs threshold P-value was kept at default setting according to the mode1, mode2 and mode3. The result was obtained including the dot matrix graphs of wild type and mutant type, query sequence with the SNP alleles and the position on the sequence. Structural changes with the variation in the folding energies were also predicted.

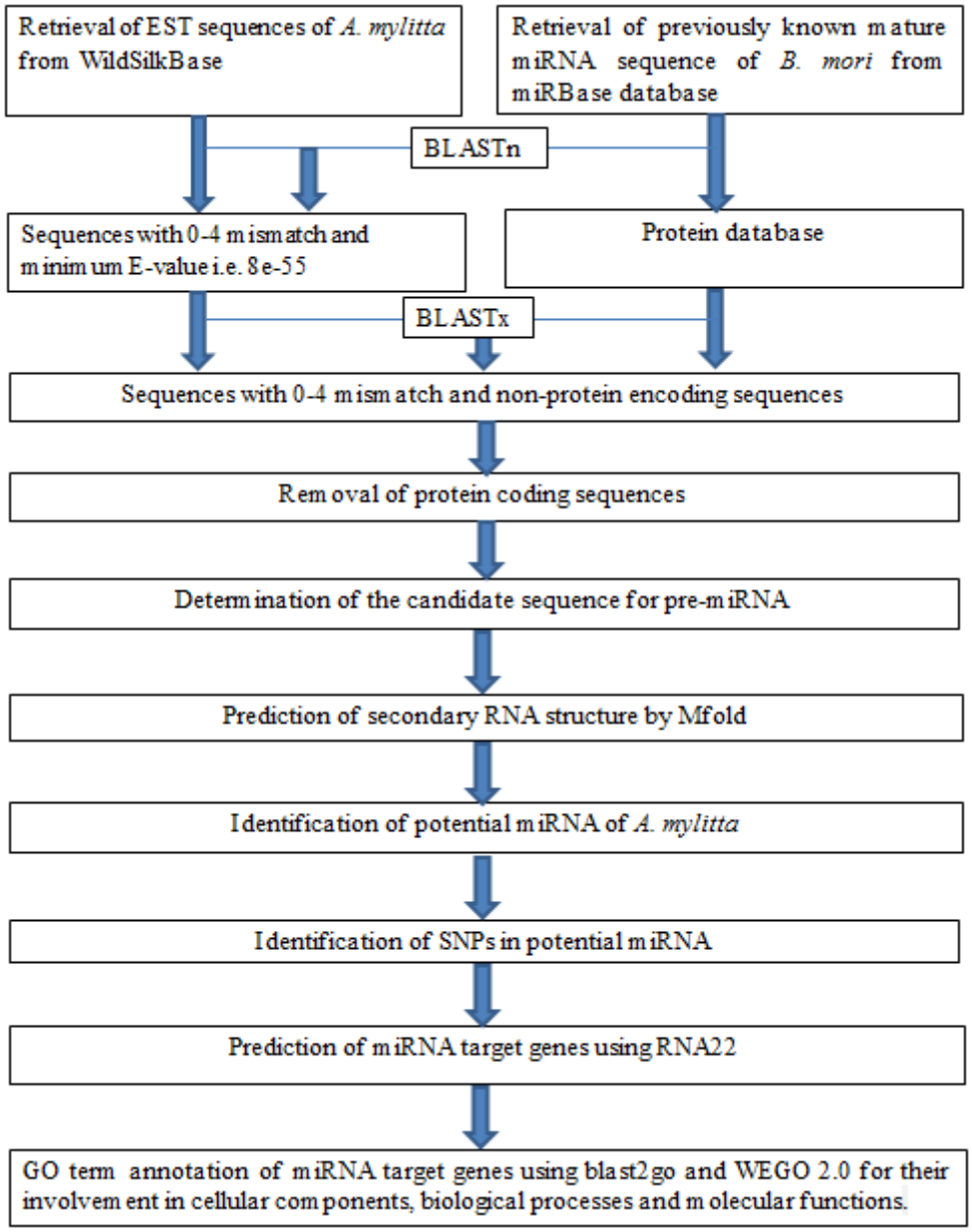
2.5 Potential miRNA targets

The whole genome sequence of *A. mylitta* is not available. Therefore, in this study *B. mori* mRNA and coding sequences were used as the reference sequences for finding the targets of the newly predicted miRNAs. The predicted *A. mylitta* miRNAs were used as query against the mRNA sequences of *A. mylitta* using the RNA22 online tool with a mismatch of 4 base pairs and the following parameters: (1) Sensitivity vs. Specificity setting, (2) Seed/nucleus region, (3) Minimum number of paired-up bases in heteroduplex, (4) Maximum folding energy for heteroduplex (Kcal/mol), (5) Maximum number of G:U wobbles allowed in seed region. The result was obtained in the text format with the target sequence, its target site and the MFE value.

2.6 Gene ontology term annotation

To understand the function of predicted miRNA, Gene ontology (GO) terms were annotated for the targets using blast2go. WEGO (Web Gene Ontology Annotation Plot) was also made for visualizing, comparing and plotting GO annotation results of miRNA targets using online WEGO 2.0 [44] for their involvement in cellular components, biological processes and molecular functions.

A flow chart indicating all the steps used in the present study in concise form have been described.



3. RESULTS AND DISCUSSION

3.1 Identification of miRNAs

To perform *in silico* study on the ESTs in *A. mylitta*, EST sequences of *A. mylitta* were retrieved from WildSilkBase, while mature miRNA sequences of *B. mori* were retrieved from the miRBase database. Total of 1412 EST sequences and 159 miRNA sequences were obtained from the databases. These EST sequences were mapped against mature miRNA sequences for the homology search using the NCBI BLASTn program. Total 3895 aligned sequences were obtained, out of these twenty-seven sequences were shortlisted on the basis of the miRNA criteria and were used for further analysis. The shortlisted EST sequences were screened against the protein database of NCBI using the BLASTx program. This analysis gives the non-redundant protein sequences that were aligned with the EST sequences. Obtained protein sequences were used for further comparing with the EST database of *A. mylitta* to remove the aligned sequences. The protein sequences obtained from the BLASTx program were saved and were used for the tBLASTn program of WildSilkBase. The EST database of *A. mylitta* was aligned. The selected EST sequences were compared with the BLAST result and the protein encoding ESTs were removed.

Out of the 27 sequences, 21 EST sequences were selected as non-protein coding sequences. Further, the pre-miRNA sequences having the length of 140 nucleotides were extracted and are listed in Table 1. Similar approaches have been used to predict the miRNAs in animals [25-27] and plants [24, 29-33, 36,

37]. Efforts have also been made towards the prediction of other plants like sorghum [45] and wheat [46] using this approach. However, to get more detailed information about novel miRNA and trans-acting siRNA in *Sorghum bicolor* using high-throughput sequencing approaches has also been applied [47].

Table 1: Extracted pre-miRNA sequences.

ID	SEQUENCE
>Amfb0022	CATGTCA**GATCGTCAAAATGGCTAA**GAAAAAAGGCGATGAAAAAGTGAATCAAAA CCCTCAAAGTGATAATGAGGAGCAACATTTTGACGAGGAGCCAGATTTTGATGATCC CGAGGGATTCGTCGATGATATCCCCGAC
>Amfb0039	GCTGAGAATTCGGACGAGGGCAACTATTCATATATGTATAAGTGT**TGAAACGTATAT GTGTAC**CAAGCATACACGCCTCAACAACTCCTCCATATATTTATTATCGGAAAAATG AGAGTTTTCAGTTCGATAAGGGAAAA
>Amfb0111	GTGCCCACATTCCGCACGA**GGAGAACACGAAGTAATA**TATCAAAATGTTCGCAAAG TTATTCCTCGTGTCCGTCCTTATGGTCGGAGTTAACAGCCGATACTTGCTCATAGAAG AACCTAATTACG
>Amfb0339	AATTAGTTTTTCTTCCTATTGCTGATACAGTCGAGAATATCAAGGGTGATTTATTCCCT GCTTTCGT**CGAGCCTTTTTTGACGGA**CAAAACGATGCCTTTGAGTTTGGGAAACAGA TACAGAATTTCTTCAGGAATTGGG
>Amfb0374	GACTATTATTATAAAGAAAAATTAAAACGA**CTAATTGTTGATTTGCCTATA**ATTTCG AAAGATGACTGGCCAAAAAATTCCGACACTTATGCCGGAACATATGTCAAGTTACAA CAACAGTAAACTGAGAATATACGAGA
>Amfb0405	TGACGCCTCTGGTTGCTCCGATGCTAGGCTTCGG**TAGCGCCGGAATAGCGGC**CGGC AGCACAGCCGCAGCTGCACAGGCTTATTACGGCAACCTTGTGGCCGGCAGTATTGTG TCTCAGTTGACATCTGCGGCAATGTTA
>Amfb0406	TCTTACTGTTATTTTCGTTGGTTTACGGGCTATTTGAATACG**AAAGAGAATTTGTTAG GTTG**ACCAAATCTTATGTTACGGATGCTAAAATGTGTATTGAAGTATATGAAAGTAA AAATGATATCGCTTGCTCCTCG
>Amfb0406(b)	TTGTTAGGTTGACCAAATCTTATGTTACGGATGCTAAAATGTG**TATTGAAGTATATG AAAG**TAAAAATGATATCGCTTGCTCCTCTTTTATAAAATATTATGGTATTTTTAAGGG TGTAGTTGAACGACTT
>Amfb0508	AATGTTGCTAGTATTAAAAACAACAGTTTA**CAATCACCTTTTGTCATA**GATNAATAA AGTGTATGTATTATATATAAGAAATCTGTGTAATTTTGTATAAATTCTATTCAAAGGA TCCCTAAATTTGATCTTTTCAATAC
>Amfb0556	AAGTACTTAAACTTTTTCAAGTAAAATTATCATTTGCAAACCCTTGTCTTTGCATTAAA AAAACAG**ATTGTTTTTCTGCGGTGG**CAAATTTCAAATACTATAATTTAAATGAAAT AGATTTGACAAAAACAATTATGGGT
>Amfb0591	CGTGACAACCATGGTAGTCGCAGAAACTACCATCGAAAGTTGATAAGGCAGACATTT GAAAGATGCGTCGCCGGTACTGGAC**CGTGCGATCGGCAAAAGT**TATCCAGATTCAT CAAAATTAACGACATCGGACGCGAGGC
>Amfb0611	TGTGGAGGGTTGGATTCTGTTTGTAAGCAATGTGCATGAGGAAGCCCAAGA**AGAGGA TATTCAAAATCA**ATTCTCAGAGTTTGGGGAAATAAAAAAACATACTTTTAAATTTGG ATCGACGCACAGGATTCCTC
>Amfb0656	CTCAAACGGGGAGGACGTGGGATGAAACTGGGTCTTTAATGCTTTGTTAGGAACATT T**ATTTTTTTTATATATCAA**ATGTTTCTAATATCGGAATTATAGACTACAAAGTTTAGG ATAGGGTTCCCACAGTCGCCGCCAT
>Amfb0785	TCTCATGAGCTCAGACTAAATATGAAGAATGTCGAACCAGGGAAGCTACATAAAGTG **GACGGATTTTATATTTTCA**AAAAGCCGCTTCTAGTANAAAATCTGGAGCCAACGTCC AGGTTGAAGGTCATGAATTTACTCAT
>Amfb0846	TACCATAATATTTTATAAAAGAGGAGCAAGCGATATCATTTTTACTTTCATATACTTC AATACACATTTTAGCATCCGTAACATAAGATTTGGT**CAACCTAACAAATTCTCTTT**C GTATTCAAA
>Amfb1092	ACTGAAGAAGAGTTATTACGATTCACTAATGAACGTTTATCCAAATCCTCACGACTTC ATGGTGGTATTATATTTTTGA**ACGATATACCAAAAAATC**CTTCGGGCAAAATATTAC GACGCTTGCTCAAGCAAAAACTTCA
>Amfb1159	GAAGATGGCTCAGGAAGTGTCTGTACACATGGACGAAACCAACAAGAAACTTGCACC CAAAGTG**AAGGAAGCTTATGACGAT**TTCGTGAAGAATGTACAGGAGCTGCAGAAGA AGTTGCACGAAGC
>Amfb1224	GACAGTATGTTGACGCAGTTTAGATATGCTTCTGAACGGTGAAAGCAAGGAGTTTAC ACAATC**AGTTGAAGATTATGAAGC**CGACAACCCTGTTTTAATCAGCGGAATTTGGA AGTGTGGAAACTGTACCGACCTCAGCG
>Amfb1332	CTAACCTAGATAACTATTCATATATGTATGAGTTGTT**GAAACGTATATGTGTACC**AA GCATACACGCCTCAACAACTCCTCCATATAATTATTATCGGAAAATTGAGAGTTTTTC ATTTCAATAAGAGAAACAAAGTAAA
>Amfb1340	AGAATCGGACAGGGCTAATAAGGAGTTAAAAATGC**ATCCATTAAGGTACGACG**CAC GTCCTTACGTGAACGGACAAGATCCATCGGTGATATCAAGACGCTCCGATCTCTTGCT GATTGAGGGTGAGGAGTT
>Amfb1340(b)	GACGCACGTCCTTACGTGAACGGACAAGATCCATCGGTGATATCAAGACGCTCCGAT CTCTTGCTGATTGAGGGTGAGGAGTTGACAC**TGTGTGTGTGTGGCGCA**CCTCTTGAG GATACATGCCTTGTGTCAGA

3.2 RNA secondary structure prediction

The 21 unique non-protein coding EST sequences that exhibited homology with the known mature miRNA of *B. mori* are subjected to secondary structure prediction using MFOLD. MFOLD result was critically examined to determine the sequence of a miRNA-precursor and appropriate stem-loop structure using the criteria described previously in Section 2.4. After critical analysis of all results, sequences with perfect hairpin loop structure and highly negative MFE values were termed as potential miRNA sequences. Total six potential miRNA sequences (Table-2; and Figure-1A to 1F) were predicted that followed all the criteria described previously [48].

Table 2: Properties of newly identified miRNAs in *A. mylitta.*

EST ID	MFE(kcal/mol)	AMFE	MFEI	Precursor length	Mature miRNA length
Amfb0022	-33.80	24.14	0.40	140	18
Amfb0039	-33.03	23.75	0.46	140	18
Amfb0405	-53.80	38.43	0.47	140	18
Amfb0556	-21.30	15.21	0.42	140	18
Amfb0591	-34.80	24.86	0.36	140	18
Amfb1332	-33.50	23.92	0.56	140	18

Note: MFE, minimal folding free energy; AMFE, adjusted minimal folding free energy; MFEI, minimal folding free energy index

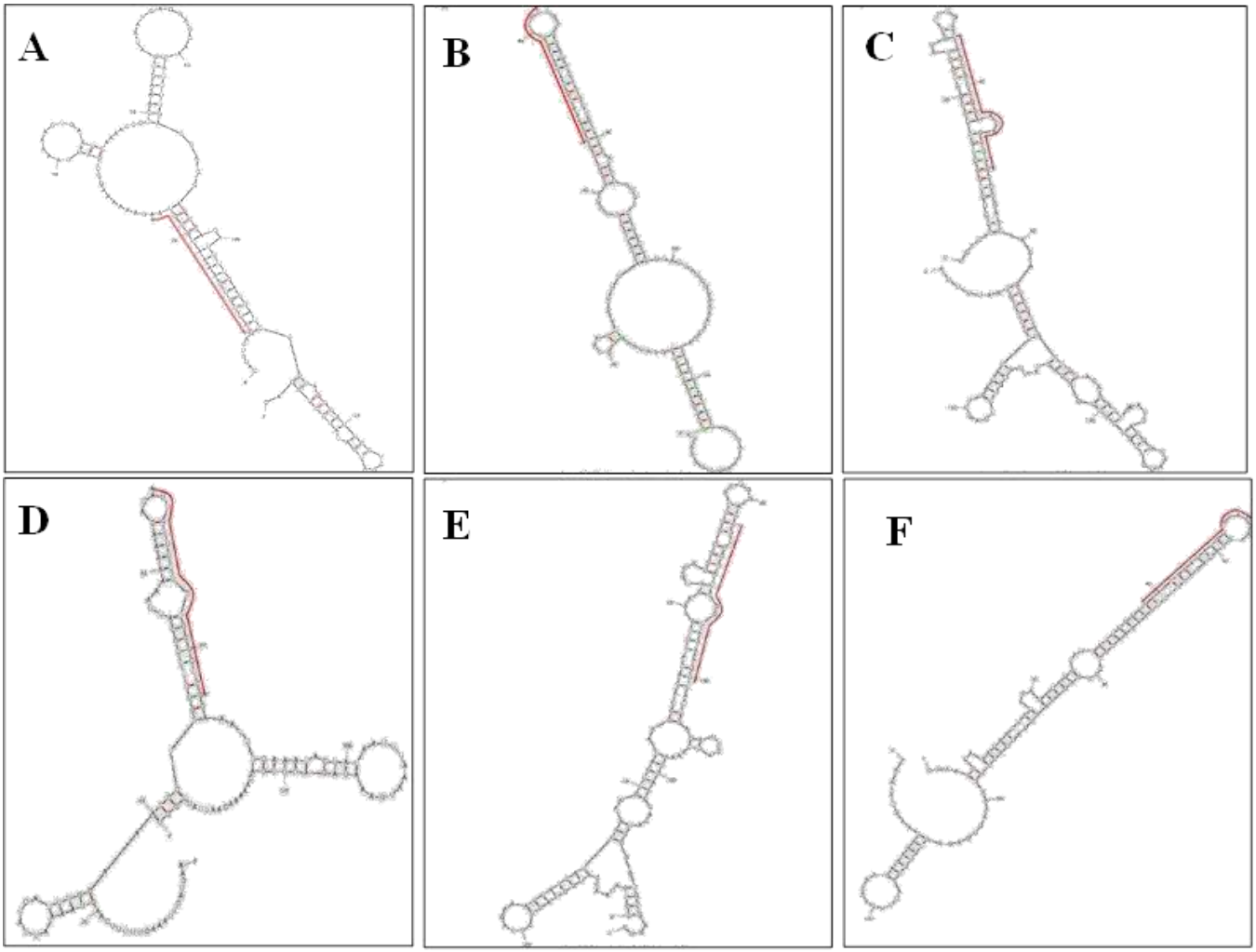


Figure 1. Predicted stem-loop hairpin secondary structure of the identified *A. mylitta* miRNAs: (A) Amfb0022, (B) Amfb0039, (C) Amfb0405, (D) Amfb0556, (E) Amfb0591, and (F) Amfb1332. Mature miRNA sequence is in red color.

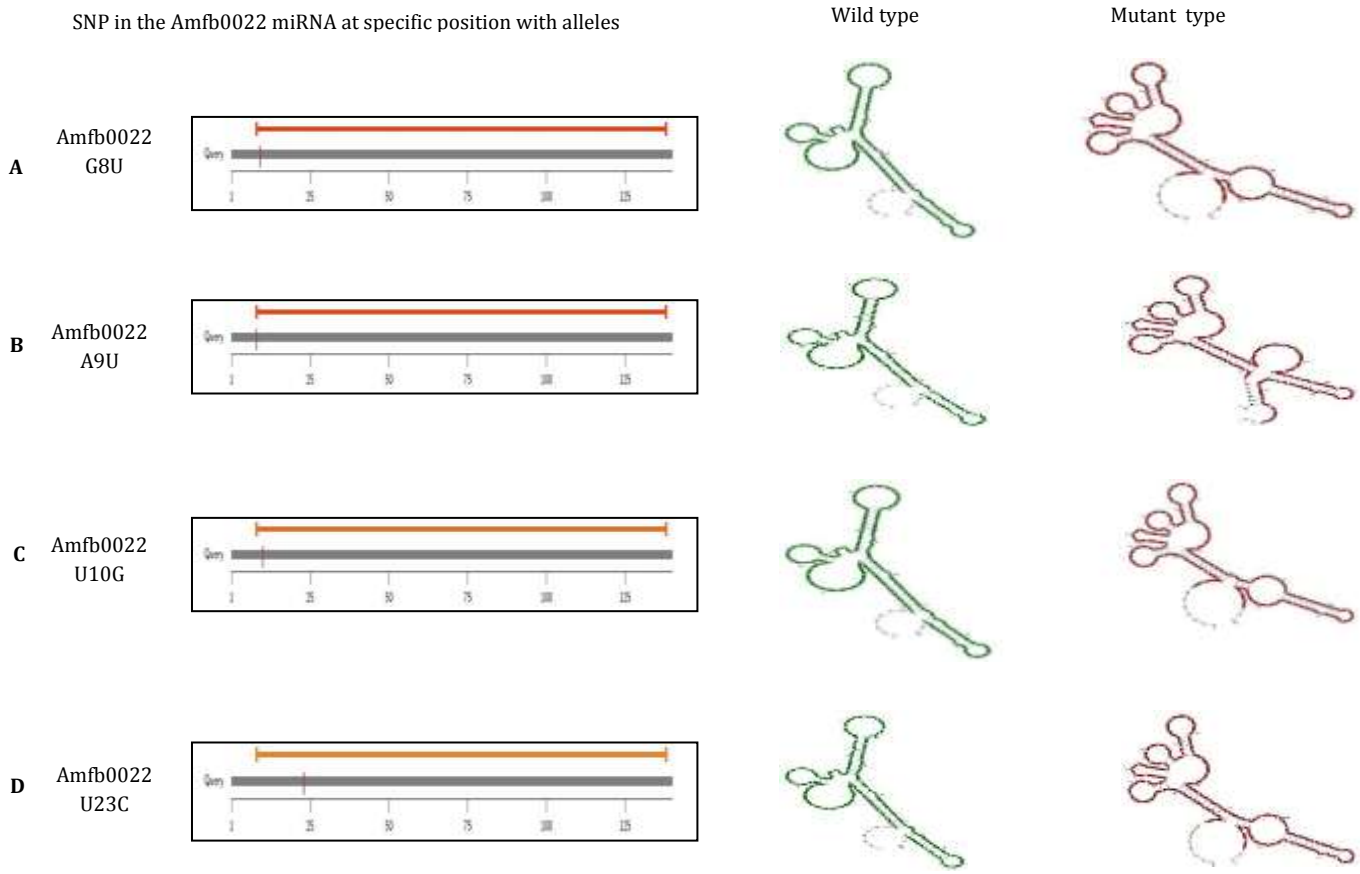


Figure 2.1: SNP in the Amfb0022 miRNA at: (A) position 8 with alleles G/U, (B) position 9 with alleles A/U, (C) position 10 with alleles U/G, (D) position 23 with alleles U/C, as well as structural variation of wild type and mutant type.

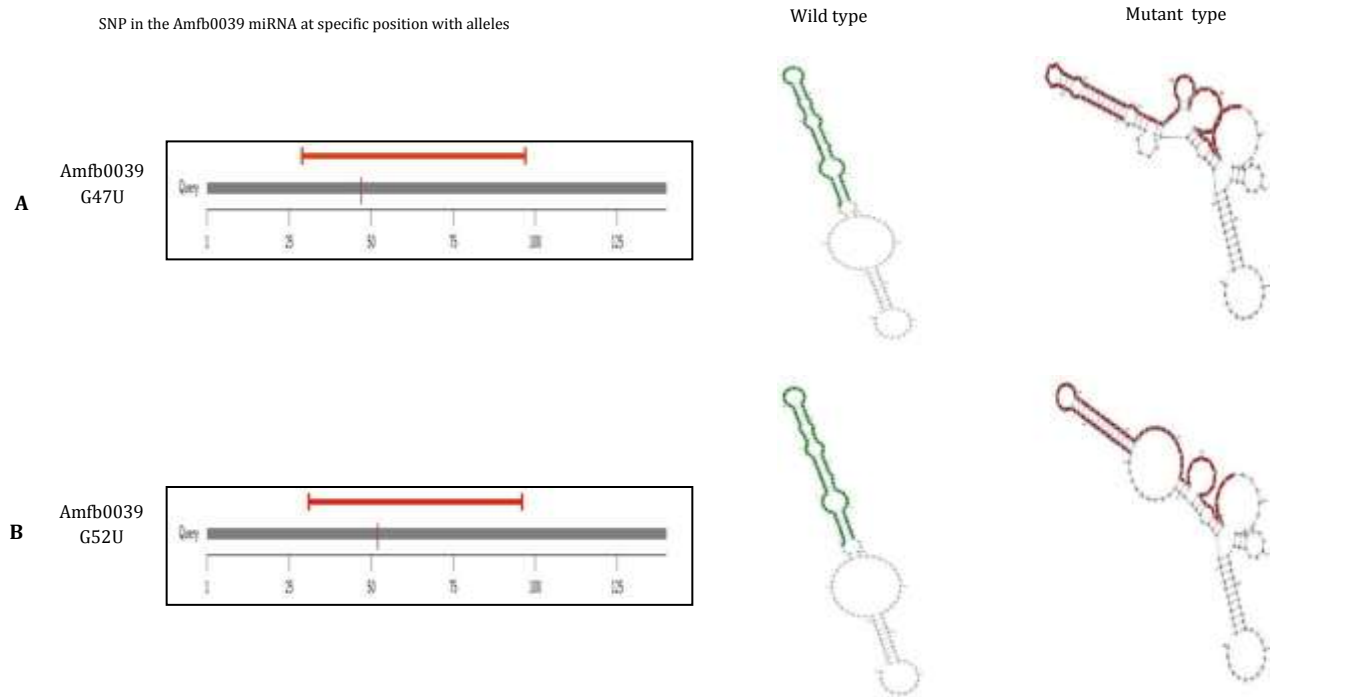


Figure 2.2: SNP in the Amfb0039 miRNA at: (A) position 47 with alleles G/U, and (B) position 52 with alleles G/U with alleles A/U, as well as structural variation of wild type and mutant type.

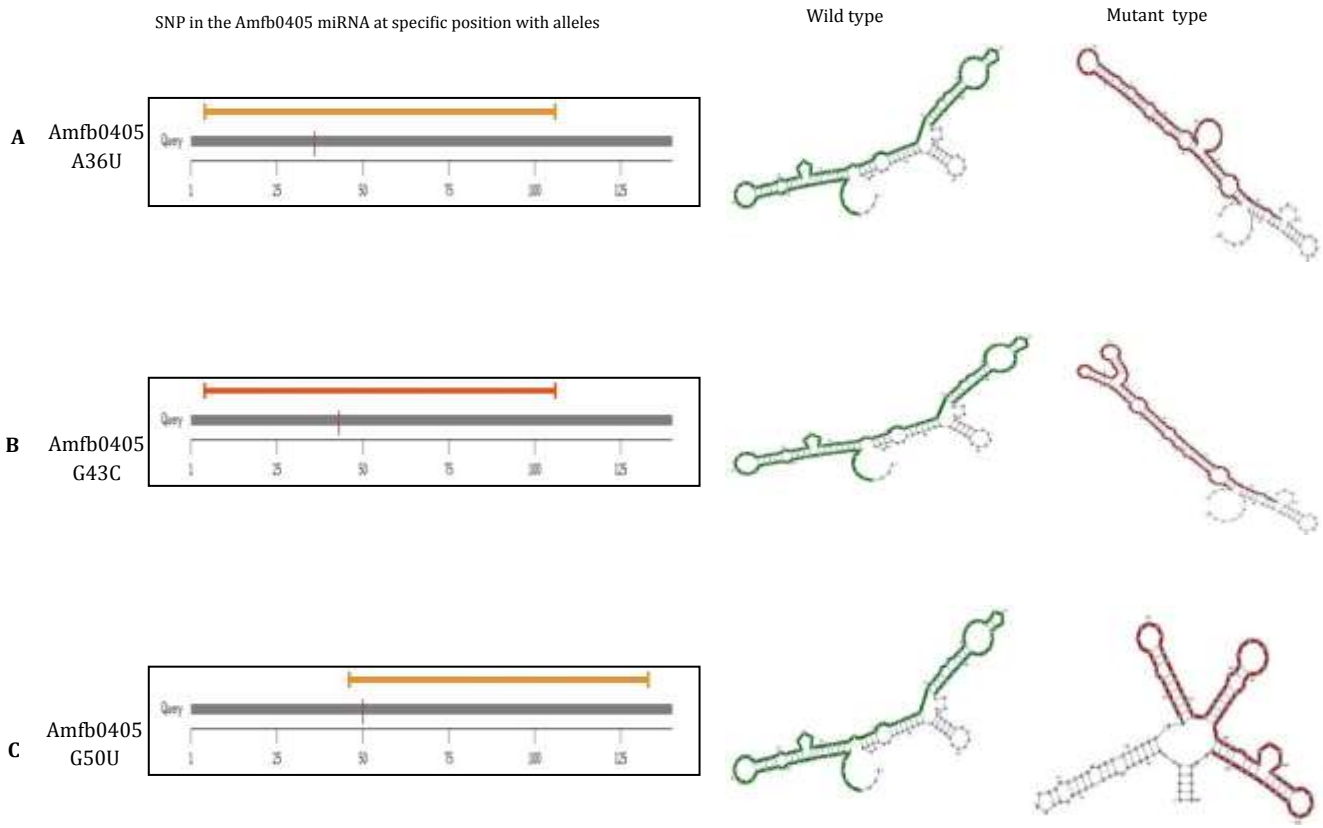


Figure 2.3: SNP in the Amfb0405 miRNA at: (A) position 36 with alleles A/U, (B) position 43 with alleles G/C, and (C) position 50 with alleles G/U, as well as structural variation of wild type and mutant type.

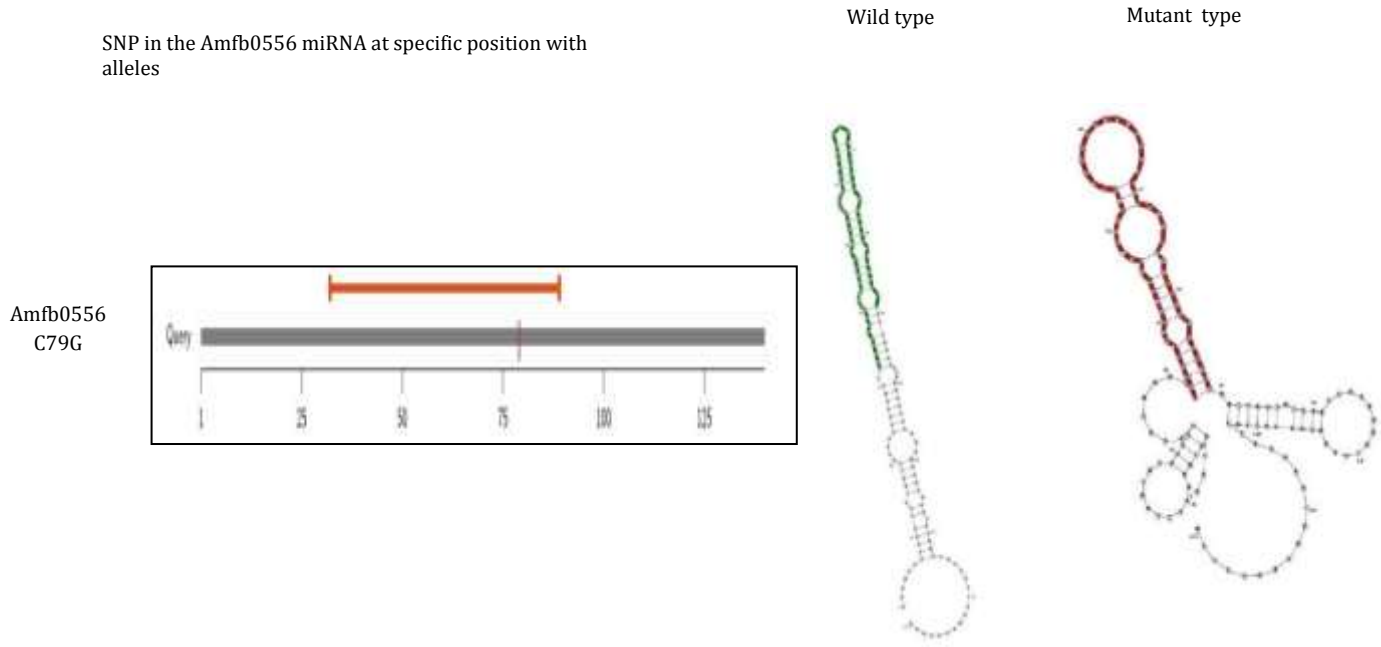


Figure 2.4: SNP in the Amfb0556 miRNA at position 79 with alleles C/G as well as structural variation of wild type and mutant type.

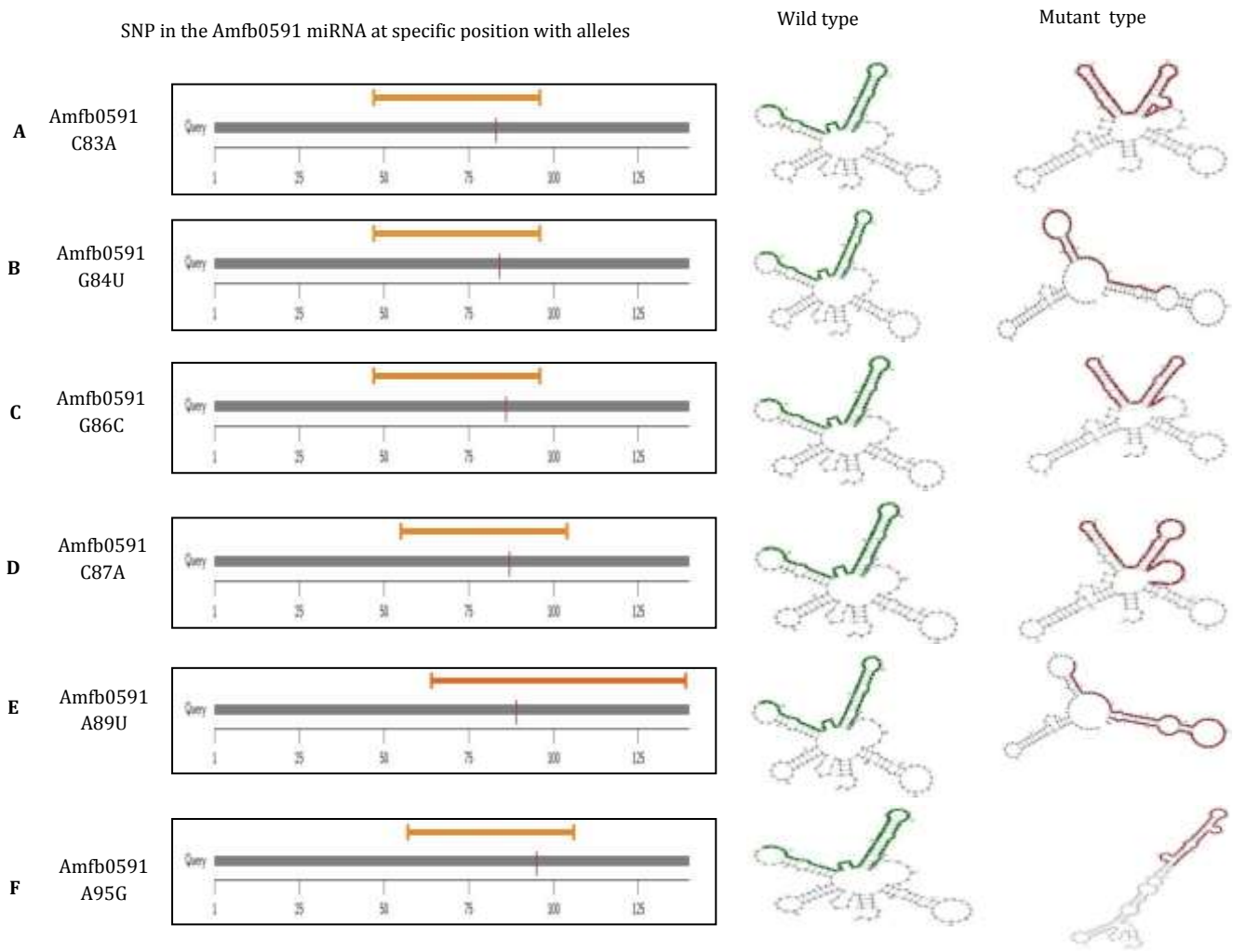


Figure 2.5: SNP in the Amfb0591 miRNA at: (A) position 83 with alleles C/A, (B) position 84 with alleles G/U , (C) position 86 with alleles G/C, (D) position 87 with alleles C/A , (E) position 89 with alleles A/U, and (F) position 95 with alleles, A/G as well as structural variation of wild type and mutant type.

3.3 Identification of SNPs associated with miRNA genes

Identified miRNA were further subjected to online tools for the prediction of polymorphism at different sites. To identify the SNPs associated with *A. mylitta* miRNAs, the identified miRNA sequences were uploaded in the RNAsnp Web Server. The SNPs with the position at which the polymorphism occurs and the structure of the wild type and mutant type variation were obtained. For the six potential miRNAs examined in this study, total 16 SNPs were predicted in only five miRNAs. Four SNPs that have been predicted in Amfb0022 miRNA at different positions of the sequences and their structural differences in wild type and mutant type are shown in Figure-2.1. Two SNPs that have been predicted in Amfb0039 miRNA at different positions of the sequences and their structural differences in wild type and mutant type are elustrated in Figure-2.2. Figure-2.3 illustrate about three predicted SNPs in Amfb0405 miRNA at different positions of the sequences and their structural differences. Only one SNPs was predicted in Amfb0556 miRNA at the position of 79 with allele C/G sequence and their structural differences in wild type and mutant type are shown in Figure-2.4. Maximum six SNPs have been predicted in Amfb0591 miRNA at different positions of the sequences and their structural differences in wild type and mutant type are illustrated in Figure-2.5. Obtained SNPs may control the expression and processing of miRNAs. Identification of SNPs in bread wheat using computational approaches has also been described [49].

Table 3: List of the potential targets of newly identified miRNAs in *A. mylitta*

miRNA	miRNA sequence	Target sequence ID	Protein name	Position	MFE
>Amfb0022	GAUCGUCAAAAUGGCUAA	>gi|755223228|gb|KJ938018.1|	methyltrasferase (AJK26816.1)	1194-1210	-13.60
		>gi|18699032|gb|AF374298.1|	RNA binding protein NSP38 (AAL78381.1)	107-124	-14.60
>Amfb0039	UGAAACGUAUAUGUGUAC	>gi|110347787|gb|DQ666492.1|	cecropin-like protein (ABG72696.1)	779-796	-19.20
		>gi|119066318|gb|DQ975382.1|	p60 (ABL60869.1)	1565-1583	-12.00
>Amfb0405	UAGCGCCGGAAUAGCGGC	>gi|302513048|gb|GU930287.1|	male-specific doublesex isoform M (ADL40855.1)	307-325	-13.40
		>gi|302513046|gb|GU930286.1|	female-specific doublesex isoform F2 (ADL40854.1)	307-325	-13.40
		>gi|302513044|gb|GU930285.1|	female-specific doublesex isoform F1 (ADL40853.1)	307-325	-13.40
		>gi|110347846|gb|DQ666522.1|	protease inhibitor-like protein (ABG72725.1)	98-117	-15.00
				123-139	-14.20
		>gi|110347836|gb|DQ666517.1|	protease inhibitor-like protein (ABG72720.1)	116-133	-12.10
		>gi|110347832|gb|DQ666515.1|	protease inhibitor-like protein (ABG72718.1)	240-257	-12.10
		>gi|110347811|gb|DQ666504.1|	similar to lectin-like protein	263-278	-15.90
		>gi|296396421|gb|GQ351286.1|	RNA-dependent RNA polymerase (ADH10220.1)	891-906	-14.70
		>gi|37786953|gb|AY212274.1|	Polyhedrin (AAP44508.1)	508-525	-17.40
		>gi|37786907|gb|AY212273.1|	Polyhedrin (AAP44507.1)	517-534	-17.40
		>gi|37786770|gb|AY212272.1|	Polyhedrin (AAP44506.1)	516-533	-17.40
>Amfb0556	AUUGUUUUUCUGCGGUGG	>gi|110347850|gb|DQ666524.1|	protease inhibitor-like protein (ABG72727.1)	267-284	-12.10
		>gi|110347840|gb|DQ666519.1|	protease inhibitor-like protein (ABG72722.1)	113-132	-14.60
		>gi|110347836|gb|DQ666517.1|	protease inhibitor-like protein (ABG72720.1)	220-237	-13.10
		>gi|110347832|gb|DQ666515.1|	protease inhibitor-like protein (ABG72718.1)	344-361	-13.10
		>gi|110347830|gb|DQ666514.1|	serpin-like protein (ABG72717.1)	33-51	-16.40
		>gi|110347801|gb|DQ666499.1|	lebocin-like protein (ABG72703.1)	842-858	-15.60
		>gi|110347799|gb|DQ666498.1|	hemolin-like protein (ABG72702.1)	85-103	-12.30
		>gi|110347785|gb|DQ666491.1|	attacin-like protein (ABG72695.1)	224-243	-13.60
		>gi|110347783|gb|DQ666490.1|	attacin-like protein (ABG72694.1)	113-131	-13.10
				175-193	-15.50
				3833-3850	-15.60
		>gi|755223228|gb|KJ938018.1|	Methyltrasferase (AJK26816.1)	940-956	-12.70
				3077-3094	-14.00
		>gi|18699032|gb|AF374298.1|	RNA binding protein NSP38 (AAL78381.1)	485-504	-14.70
		>gi|107122649|gb|DQ512389.1|	P61 (ABF83587.1)	1138-1154	-13.20
>Amfb0591	CGUGCGAUCGGCAAAAGU	>gi|119066318|gb|DQ975382.1|	p60 (ABL60869.1)	400-419	-12.90
>Amfb1332	GAAAACGUAUAUGUGUAC	>gi|110347846|gb|DQ666522.1|	protease inhibitor-like protein (ABG72725.1)	33-48	-13.20

3.4 Prediction of potential targets of putative miRNA in *A. mylitta*

To predict miRNA targets for identified *A. mylitta* miRNAs, RNA22 analysis server and *A. mylitta* mRNA/cDNA sequences were used. Various potential targets of predicted miRNA has been identified and listed (Table 3) having their name and identification number. Prediction of target genes for the *A. mylitta* miRNAs revealed that individual miRNA was involved in the regulation of more than one target genes. The main role of miRNA sequences is in the regulation of post transcriptional functions, blocks the synthesis of protein. Targeted genes by our identified miRNA contain inhibitor responses like proteins such as protease inhibitor-like protein, serpin-like protein, lectin-like protein, hemolin-like protein and other proteins. Similar studies have also been reported previously [50]. Obtained result in *A. mylitta* will help to understand the role of identified miRNA in the regulation of different target genes. Experimental validation of identified miRNA with gain-of-function and loss-of-function will help in better understanding of detailed role of these miRNA in the regulation of gene expression and post transcriptional functions of predicted target genes in *A. mylitta.* Identified *A. mylitta* miRNA target genes and their functions are listed in Table 3. Also, identification of miRNA and their targets in *B. mori* [51], cotton [48], wheat [49], coffee [50], sorghum [47], and *Humulus lupulus* [52] have also been well elucidated.

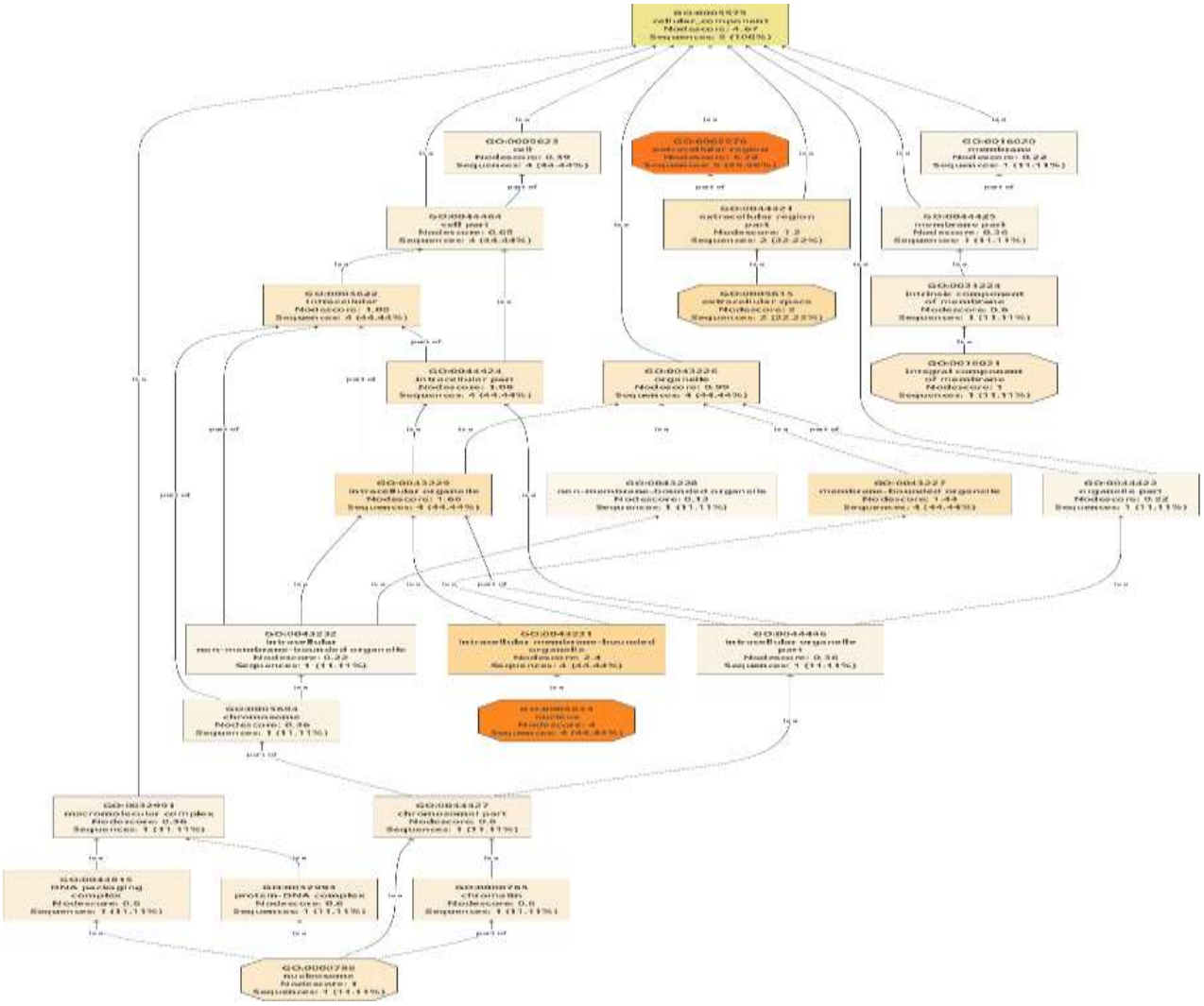


Figure 3: GO indicating the involvement of miRNA targets in cellular components.

3.5 Gene ontology analysis and functional characterization

To understand the function of predicted miRNA targets, GO terms were annotated for the targets using blast2go. WEGO plot indicated the detailed annotation results of miRNA targets for their involvement in cellular components (Figure-3), biological processes (Figure-4), molecular functions (Figure-5), species distribution (Figure-6) as well as gene distribution of cellular component, biological processes and molecular functions of miRNA targets (Figure-7).

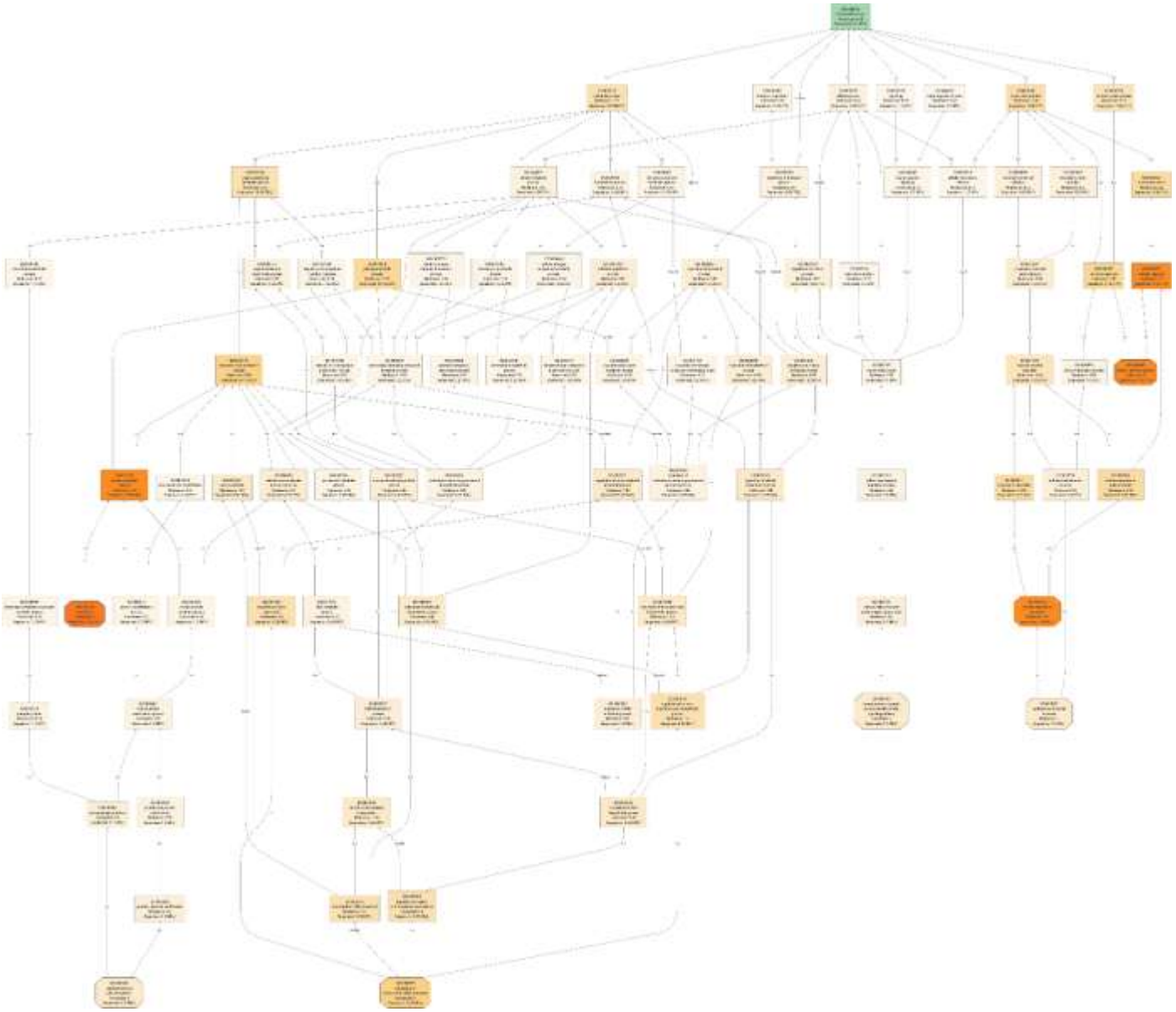


Figure 4: GO indicating the involvement of miRNA targets in biological processes.

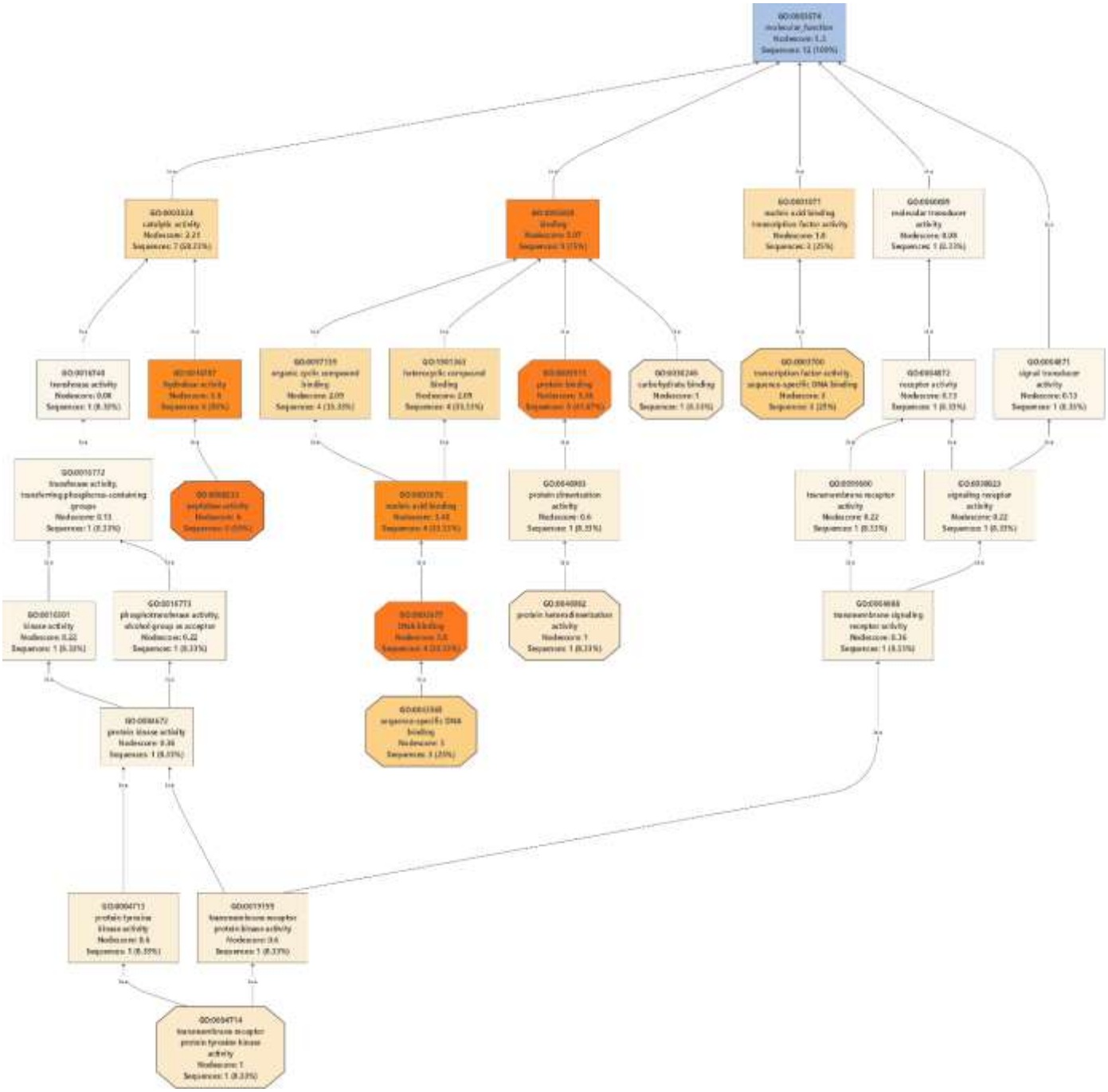


Figure 5: GO indicating the Molecular functions of miRNA targets.

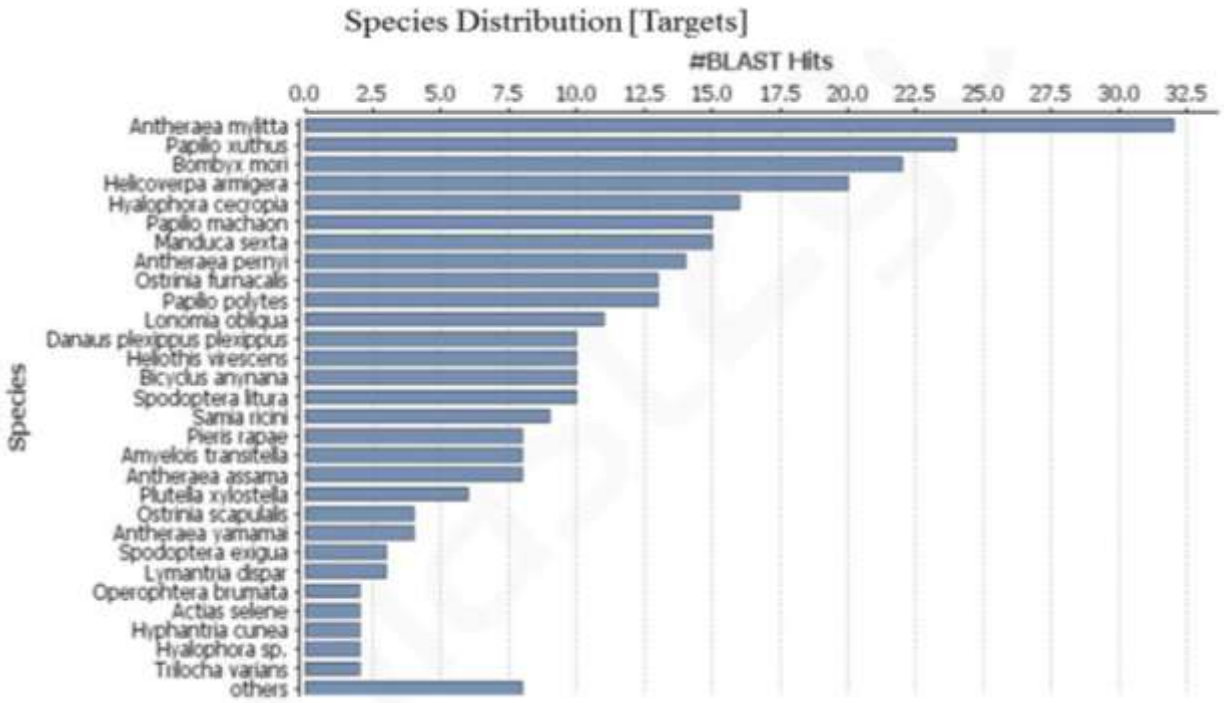


Figure 6: GO indicating the species distribution of miRNA targets.

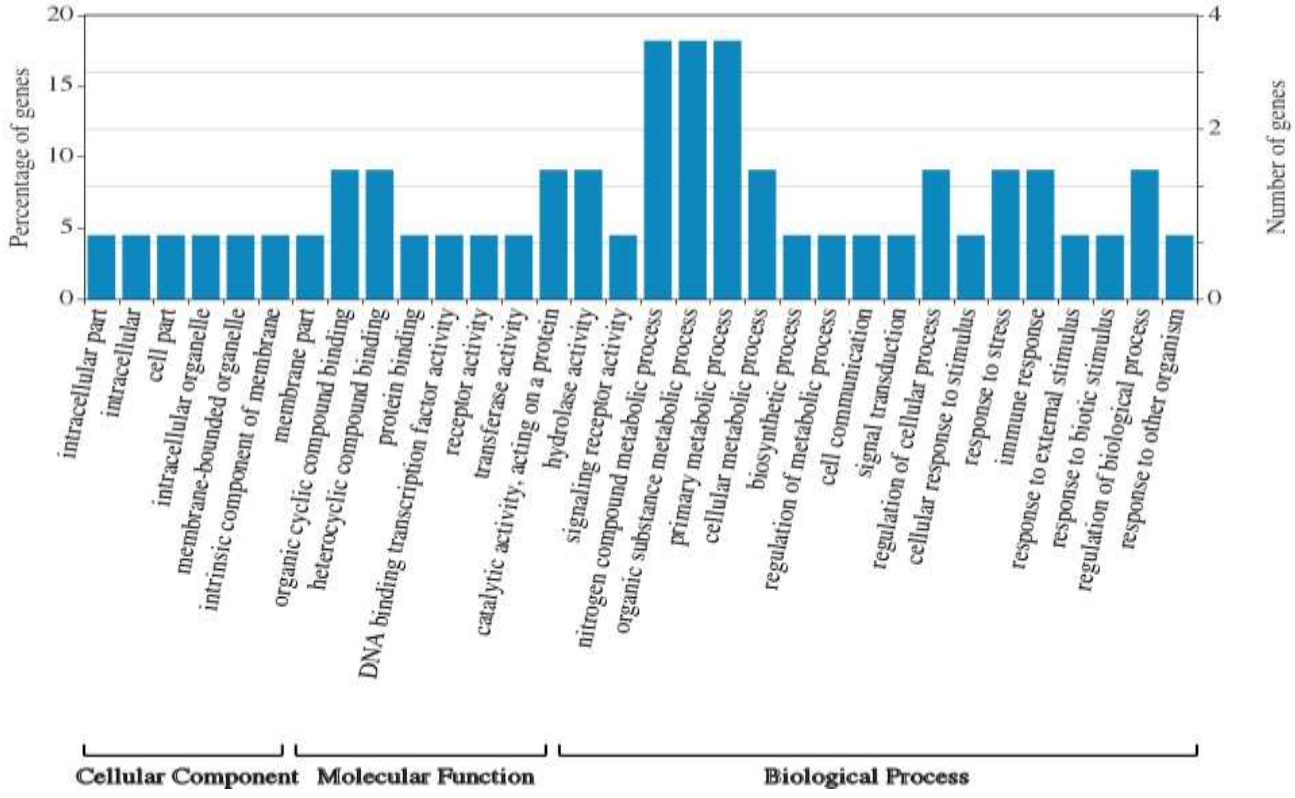


Figure 7: GO indicating the gene distribution in cellular component, Biological Processes and Molecular functions of miRNA targets.

4. CONCLUSION

In silico analysis of ESTs were carried out to identify potential miRNA and related SNPs in *A. mylitta*. ESTs based homology search was performed for finding the miRNAs from species whose genome is not yet fully known. In this study we identified the six potential miRNAs in highly economical silkworm species *A. mylitta* using the EST based approach and used to find the target genes. Identification of SNPs in the newly identified potential miRNAs were also carried out. Sixteen SNPs were predicted at different positions in five miRNA sequences. Potential miRNA targets and

their GO term analysis were also performed indicating their involvement in cellular component, biological processes and molecular functions. Overall, present study gives comprehensive information about use of ESTs for miRNA, SNPs, miRNA targets and their GO analysis particularly in *A. mylitta.*

Acknowledgements: Financial assistance provided by the Department of Biotechnology, Government of India, New Delhi through BITISNet SubDIC (BT/BI/0406504) for the Bioinformatics facility to carry out this project is gratefully acknowledged. Department of Biotechnology, New Delhi, GOI is thankfully acknowledged for supporting the JPP and DMP in the form Sponsored Research Project (Letter No. BT/PR5375/PBD/19/233/2012 Dated 10.06.2013 and Reference No. IFP, DBT, SAN no. 102/IFP/SAN/1162/2013-2014 dated 7/6/2013 serial no.1809).

Conflict of interest: We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. All the work was majorly performed by Anjali Singh Deo, Sneha Prasad and Mousam Kumar Ram under the supervision of Dr. D. M. Pandey, Associate Professor Department of Bio-Engineering, BIT Mesra, Jharkhand, India and with joint efforts of Dr. A. K. Sinha and Dr. J. P. Pandey from Central Tasar Research & Training Institute, Piska Nagari Ranchi, Jharkhand, India.

5. REFERENCES

1. Nagaraju, J, Goldsmith, M.R. (2002) Silkworm genomics-progress and prospects. Curr Sci, 83, 415–425
2. Thangavelu, K. and Sinha, A.K. (1993) Population ecology of Antheraea mylitta Drury (Lepidoptera : Saturniidae) in Wild silkmoths'92, (eds.) H. Akai, Y. Kato, M. Kiuchi, J. Inouchi. International Society for wild silkmoths. National Institute of Sericultural and Entomological Science, Tsukuba, Japan. 87– 92.
3. Jolly, M.S., Sen, S.K., Ahsan, M.M. (1974) Tasar culture. Ambika Publishers, Bombay
4. Chakraborty, S., Muthulakshmi, M., Vardhini, D., Jayaprakash, P., Nagaraju, J., Arunkumar, K.P. (2015) Genetic analysis of Indian tasar silkmoth (Antheraea mylitta) populations. Sci. Report, 5, 15728.
5. Akai, H. (2000) Cocoon filament characters and post cocoon technology. Int J Wild Silkmoth Silk, 5, 255-259.
6. Lata, S., Pandey, D.M., Pandey, J.P. (2013) Unraveling the sequence similarities, conserve domain and 3D structure of cocoonase to gain insights into their functional integrity. Int J Comput Bioinfo In Silico Model, 2, 141-146.
7. Pandey D.M., Pandey J.P. (2014) Cocoonase enzyme: Current and future perspectives. Austin J Biotechnol Bioeng, 1, 2.
8. Gandhe, A.S., Arunkumar, K.P., John, S.H., Nagaraju, J. (2006) Analysis of bacteria-challenged wild silkmoth, Antheraea mylitta (lepidoptera) transcriptome reveals potential immune genes. BMC Genomics, 7, 184.
9. MacIntosh, G.C., Wilkerson, C., Green, P.J. (2001) Identification and analysis of analysis of Arabidopsis expressed sequence tags characteristic of non-coding RNAs. Plant Physiol, 127, 765-776.
10. Adams, M.D., Kelly, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulus, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252, 1651-1656.
11. Matukumalli, L.K., Grefenstette, J.J., Sonstegard, T.S., Tassell, C.P.V. (2004) EST-PAGE - managing and analyzing EST data. Bioinformatics, 20, 286-288.
12. Xia, Q., Zhou, Z., and Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha X., Cheng, T., Chai, C., Pan, G., Xu, J. et al. (2004) A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science, 306, 1937–1940.
13. Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori H., Namiki N., Kitagawa, M., Yamashita, H., Yasukochi Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M. et al. (2004) The genome sequence of silkworm, Bombyx mori. DNA Res, 11, 27–35.
14. Arunkumar, K.P., Tomar, A., Daimon, T., Shimada, T., Nagaraju, J. (2008) WildSilkbase: An EST database of wild silkmoths. BMC Genomics, 9, 338.
15. Graham, M.A., Silverstein, K.A.T., Cannon, S.B., VandenBosch, K.A. (2004) Computational identification and characterization of novel genes from legumes. Plant Physiol, 135, 1179-1197.
16. Jung, J.D., Park, H.W., Hahn, Y., Hur, C.G., In, D.S., Chung, H.J., Liu, J.R., Choi, D.W. et al. (2003) Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. Plant Cell Rep, 22, 224-230.
17. Ohlrogge, J., Benning, C. (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol, 3, 224-228.
18. Sun, G. (2011) MicroRNAs and their diverse functions in plants. Plant Mol Biol 80, 17-36.
19. Bartel, B., Bartel, D.P. (2003) MicroRNAs: at the root of plant development? Plant Physiol., 132, 709-717.
20. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116, 281-297.
21. Mallory, A.C., Vaucheret, H. (2004) MicroRNAs: something important between the genes. Cur. Opin Plant Biol, 7, 120-125.
22. Carrington, J.C., Ambros, V. (2003) Role of microRNAs in plant and animal development. Science, 301, 336-338.
23. Hunter, C., Poethig, R.S. (2003) miSSING LINKS: miRNAs and plant development. Curr Opin Genet Developm, 13, 372-378.
24. Zhang, B., Pan, X., Cobb, G.P., Anderson, T.A. (2006) Plant microRNA: a small regulatory molecule with big impact. Dev Biol, 289, 3-16.
25. Oulas, A., Boutla, A., Gkirtzou, K., Reczko, M., Kalantidis, K., Poirazi, P. (2009) Prediction of novel microRNA genes in cancer-associated genomic regions-a combined computational and experimental approach. Nucleic Acids Res, 37, 3276-3287.
26. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., Bartel, D.P. (2003) Vertebrate microRNA genes. Science, 299, 1540.
27. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., Burge C.B. (2003) Prediction of mammalian microRNA targets. Cell, 115, 787-798.
28. Rajewsky, N., Socci, N.D. (2004) Computational identification of microRNA targets. Dev Biol 267, 529-535.
29. Bonnet, E., Wuyts, J., Rouze, P., Van de Peer, Y. (2004) Detection of 91 potential plant conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proc Natl Acad Sci USA, 101, 11511-11516.
30. Jones-Rhoades, M.W., Bartel, D.P. (2004) Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Mol Cell, 14, 787-799.
31. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., Bartel, D.P. (2002) Prediction of plant microRNA targets. Cell, 110, 513-520.
32. Wang, X.J., Reyes, J.L., Chua, N.H., Gaasterland, T. (2004a) Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. Genome Biol, 5, R65.
33. Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J., Qu, L.H. (2004b) Identification of 20 microRNAs from Oryza sativa. Nucleic Acids Res, 32, 1688-1695.
34. 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 Res, 15, 336-360.

35. Carrington, J.C., Ambros, V. (2003) Role of microRNAs in plant and animal development. Science, 301, 336-338.
36. Zhang, B., Wang, Q., Wang, K., Pan, X., Liu, F., Guo, T., Cobb, G.P., Anderson, T.A. (2007) Identification of cotton miRNA and their targets. Gene, 397, 26–37.
37. Gupta, P.K., Varshney, R.K., Prasad, M. (2002) Molecular markers: principles and methodology. In: Jain SM, Brar DS, Ahloowalia BS (eds). Molecular Techniques in Crop Improvement. Kluwer, Academic Publishers, Dordrecht, The Netherlands. pp 9-54.
38. Botstein, D., Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genet, 33, 228-237.
39. Brookes, A.J. (1999) The essence of SNPs. Gene, 234, 177-186.
40. Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J. (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res, 36, D154-158.
41. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25, 3389-3402.
42. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res, 31, 3406-3415.
43. Wang, L., Liu, H., Li, D., Chen, H. (2011). Identification and characterization of maize microRNAs involved in the very early stage of seed germination. BMC Genomics, 12, 154.
44. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L. and Wang, J. (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res, 34, W293–W297.
45. Katiyar, A., Smita, S., Chinnusamy, V., Pandey, D.M., Bansal, K.C. (2012) Identification of miRNAs in sorghum by using bioinformatics approach. Plant Signal Behav, 7, 246-259.
46. Pandey, B., Gupta, O.P., Pandey, D.M., Sharma, I., Sharma, P. (2013) Identification of new stress-induced microRNAs and their targets in wheat using computational approach. Plant Signal Behav, 8, e23932.
47. Katiyar, A., Smita, S., Muthusamy, S.K., Chinnusamy, V., Pandey, D.M., Bansal, K.C. (2015) Identification of novel drought-responsive microRNAs and trans-acting siRNAs from Sorghum bicolor (L.) Moench by high-throughput sequencing analysis. Front Plant Science, 6, 506.
48. Wang M, Wang Q, Wang B (2012) Identification and characterization of microRNAs in asiatic cotton (Gossypium arboreum L.). PLoS ONE, 7, e33696.
49. Pandey, B., Sharma, P., Pandey, D.M., Sharma, I., Chatrath, R. (2013) Identification of new aquaporin genes and single nucleotide polymorphism in bread wheat. Evol Bioinform, 9, 437-452.
50. Akter, A., Islam, M.M., Mondal, S.I., Mahmud, Z., Jewel, N.A., Ferdous, S., Amin, M.R., Rahman, M.M. (2014) Computational identification of miRNA and targets from expressed sequence tags of coffee (Coffea arabica). Saudi J Biol Sci, 21, 3–12.
51. Cai, Y., Yu, X., Zhou, Q., Yu, C., Hu, H., Liu, J., Lin, H., Yang, J., Zhang, B., Cui, P., Hu, S., Yu, J. (2010) Novel microRNAs in silkworm (Bombyx mori). Funct Integr Genomics, 10, 405–415.
52. Mishra, A.K., Duraisamy, G.S., Týcová, A., Matoušek, J. (2015) Computational exploration of microRNAs from expressed sequence tags of Humulus lupulus, target predictions and expression analysis. Comput Biol Chem, 59, 131-141.

*****