Journal of Insect Science, (2019) 19(3): 22; 1–9
doi: 10.1093/jisesa/iez049
Molecular Entomological Genetics

A Comparative Whole Genome Sequence Analysis Leads to Identification of Repeat-Associated Evolutionarily Conserved miRNAs in *Bombyx mori* (Lepidoptera: Bombycidae)

Jyoti Singh[1,3,] and Uddhav B. Ambi[1,2]

[1]Lab-1 (New Building), National Centre for Cell Science, University of Pune Campus, Ganeshkhind, Pune, Maharashtra 411007, India, [2]Current address: Helmholtz Institute for RNA-Based Infection Research (H1R1), Genome Architecture and Evolution of RNA Viruses, Josef-Schneider-Str.2/D15, D-97080 Würzburg, Germany, and [3]Corresponding author, e-mail: jyoti.cdfd@gmail.com

Subject Editor: Margaret Allen

Received 27 February 2019; Editorial decision 30 April 2019

Abstract

MicroRNAs (miRNAs) are evolutionarily conserved small noncoding RNAs, which play important regulatory roles in various biological processes. In this study, we have developed a computational approach for detecting conserved miRNAs based on comparison of whole genome sequences of closely related species by considering various key features of experimentally validated miRNAs. By applying this approach, we have identified 34 new miRNAs from *Bombyx mori* (L.), which are also conserved in *Drosophila melanogaster* (Diptera: Drosophilidae) and *Anopheles gambiae* (Diptera: Culicidae). Most of these miRNAs were associated with repeat region of the genome. We did an expression analysis of the 34 newly predicted miRNAs and found that 30 of these miRNAs are expressing in different tissues of *B. mori*. Besides, we have also predicted the putative targets of these miRNAs in *B. mori* based on several known characteristic features of miRNA::mRNA duplexes and found that these targets include diverse range of functions, suggesting multiple layers of gene regulation of various important biological processes.

Key words: miRNA, repeat-associated miRNA, *Bombyx mori*, conserved miRNA

MicroRNAs (miRNAs) are endogenous small noncoding RNAs of ~22 nucleotides that have emerged as key regulators of posttranscriptional gene expression. miRNAs are derived from ~80 nucleotides long precursors (pre-miRNAs), which can fold back into typical stem-loop structure. The initial processing of pre-miRNAs takes place in the nucleus from primary transcripts called pri-miRNAs with the help of an RNase III enzyme Drosha (Lee et al. 2002, 2003). The pre-miRNAs are transported to the cytoplasm (Yi et al. 2003, Kim et al. 2009, Okada et al. 2009) and the subsequent processing of these pre-miRNAs into ~22 nucleotide miRNA::miRNA* duplexes takes place in the cytoplasm is mediated by another RNase III enzyme Dicer (Grishok et al. 2001, Hutvagner et al. 2001, Ketting et al. 2001). Based on the thermodynamic stability of each end of these duplexes, one of the strands is preferentially incorporated into the RNA-induced silencing complex (RISC), and this miRNA is considered to be the biologically active form (Schwarz et al. 2004, O'Toole et al. 2006). Subsequently, such RISC complexes with distinct miRNAs bind to the complementary site(s) on the target mRNAs. Depending on the extent of complementarity between the miRNA and its target mRNA, such binding regulates protein expression either by degrading the target mRNA or by blocking the translation machinery (Lau et al. 2001, Bartel 2004).

Almost a decade ago, the discovery of miRNAs (Lee et al. 1993, Wightman et al. 1993, Reinhart et al. 2000) has revolutionized our understanding of gene expression regulation. Regulatory miRNAs have now been identified in various species including viruses (Pfeffer et al. 2004).

Diverse experimental and computational approaches, with their inherent strengths and weaknesses, have been employed to predict miRNAs in different species. Cloning-based experimental identification methods suffer from the fact that they require a considerable amount of small RNA as starting material. This generally results in under-representation of temporally restricted and sparsely expressed miRNAs (Lai et al. 2003). In contrast, computational methods rapidly identify such miRNAs that are even expressed at low levels (Mendes et al. 2009). But many of these computation driven methods are often based on the known data sets (Mishra and Lobiyal 2010).

A widely used computational approach for the prediction of conserved miRNAs is based on homology search of known miRNAs reported in miRBase (the primary repository for published miRNA sequences) on to the complete genome of the query species

(Griffiths-Jones 2004), followed by scanning the secondary structures of their precursor sequences for characteristic stem-loop structure (Zhang et al. 2006, 2007, 2008; Singh and Nagaraju 2008; Han et al. 2010; Yin and Shen 2010). However, such predictions need to be constantly updated due to the dynamic and evolving nature of miRBase, where newly predicted miRNAs are added with each release. These approaches generally miss conserved miRNAs not reported in the known dataset.

To address these shortcomings, we have developed an approach to predict all the possible conserved miRNAs among closely related species by analyzing whole genome sequences for telltale signs of experimentally validated miRNAs like minimum free energy change of the stem-loop structure, asymmetric bulge size restriction and phylogenetic conservation of mature miRNA sequence, etc. Employing this approach, we predicted 76 conserved miRNAs in *Bombyx mori* (L.), by comparing whole genome sequence of *B. mori* with those of its two closely related Dipteran species, *Drosophila melanogaster* (D.) and *Anopheles gambiae* (D.). Out of these 76 miRNAs, 34 proved to be novel ones that had not been reported from any other species till date. Further, we have experimentally validated expression of all the newly predicted miRNAs in different tissues of *B. mori*. As a result, we found 30 out of 34 miRNAs expressions in various tissues of *B. mori*. Additionally, we also predicted putative targets of these novel miRNAs by making use of the 3′UTR of mRNA sequences from *B. mori*.

Materials and Methods

Genome Sequence Data

The genome sequence of *B. mori* was downloaded from the ftp site of Silkbase (http://silkbase.ab.a.u-tokyo.ac.jp/). Genome sequences of all other insects were downloaded from the genome database of NCBI (http://www.ncbi.nlm.nih.gov/).

Known miRNA Data

Known conserved miRNA sequences of *B. mori* were downloaded from Release 21.0 of miRBase database (http://microrna.sanger.ac.uk/sequences/).

Prediction of miRNAs

Fragmentation

The whole genome sequence of *B. mori* was first fragmented into 80 nucleotides (approximate size of pre-miRNA) long overlapping sequences with the difference of four nucleotides, i.e., 1 to 80, 5 to 85, 10 to 90 and so on, using the in house program.

Characteristics of Pre-miRNAs

Vienna RNA package (http://www.tbi.univie.ac.at/RNA/) was employed for checking the auto-folding of each of the 80 nucleotide long fragments along with the following criteria based on previous reports (Ambros et al. 2003, Lai et al. 2003, Lim et al. 2003, Joshi et al. 2010) for characterizing a typical stem-loop structure of pre-miRNAs:

a) Minimum free energy change (ΔG) should be less than and equal to –25 kcal/mol.
b) Number of base pairs in the miRNA::miRNA* duplex region should be more than or equal to 16.
c) The asymmetric bulge size in the stem-loop structure should be restricted to five bases.

Phylogenetic Conservation

To confirm the phylogenetic conservation of the predicted miRNAs of *B. mori*, BLASTn search of all the predicted pre-miRNAs was done against the genomes of *D. melanogaster* and *A. gambiae* with *e-value* less than 0.1. Generally, mature miRNA sequences are more conserved as compared to the rest of the fragment of their pre-miRNAs (Lau et al. 2001). Hence, instead of searching for the conservation of complete pre-miRNA sequence we considered only the matured miRNA sequences. Therefore, the *identity* parameter was taken as more than or equal to 18 nucleotides. The pre-miRNA sequences of the conserved hits in *D. melanogaster* and *A. gambiae* were extracted using upstream and downstream flanking regions of the hit position into consideration. Each of the extracted pre-miRNA sequences was also subsequently scanned through the first filter based on sequence and structural characteristics of a typical pre-miRNA as described earlier.

Randomization Test

Robustness of each of the 682 conserved pre-miRNA structures was assessed by randomization test using Randfold program downloaded from the following ftp site: http://bioinformatics.psb.ugent.be/supplementary_data/erbon/nov2003/, and the parameters assigned for calculating *P*-value for the pre-miRNA secondary structures were:

a) Number of randomization equal to 1,000 and
b) Type of shuffling was dinucleotide.

Nomenclature

The mature miRNAs and their precursor sequences were designated as 'miR', and 'mir' respectively, with the prefix 'bmo-new' for *B. mori* newly predicted miRNAs. Identical miRNAs originated from distinct precursors located at different places in the genome were suffixed with 'dash numbers', e.g., as *bmo-new-miR-10–1* and *bmo-new-miR-10–2*. Whereas, miRNA sequences with one or two mismatches were denoted by suffixing with an additional lower case alphabet, e.g., *bmo-new-miR-17a* and *bmo-new-miR-17b*.

RNA Isolation From Tissues and BmN Cells

RNA was isolated from midgut and fat body tissues of 5th instar second day *B. mori* larvae as well as the ovary derived BmN cells (maintained in TC-100 insect medium (Sigma) supplemented with 10% FBS at 25°C) using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Subsequently, DNA contamination was removed by treating with DNase I (Invitrogen).

Stem-Loop Reverse Transcription Polymerase Chain Reaction

All the stem-loop reverse transcription polymerase chain reaction (RT-PCR) primers are listed in Supp Table S2 (online only).

RT Reaction

cDNA was synthesized from the small RNA by using eight nucleotides miRNA specific stem-loop primers as previously described (Yu et al. 2008), with some modifications. The 20 µl reverse transcriptase reaction included 3 µg RNA, 1 µl 10 mM dNTPs (Fermentas) and 1 µl 50 nM stem-loop primer. The reaction mixture was heated at 65°C for 5 min and then immediately placed on ice. The contents were collected by brief centrifugation and then 4 µl 5× first strand buffer (Invitrogen), 1 µl 0.1 M DTT and 1 µl 200 U/µl Superscript III reverse transcriptase (Invitrogen) were added and this mixture was incubated at 55°C for 60 min. The reverse transcriptase enzyme was finally inactivated at 70°C for 15 min.

PCR Reaction

Twenty-microliter PCR mixture contained 1.5 µl cDNA, 2 µl 10× PCR buffer, 0.2 µl 10 mM dNTPs (Fermentas), 1.2 µl 25 mM $MgCl_2$, 1 µl each of 5 µM forward and reverse primers, 0.2 µl 5 U/µl Taq polymerase (Fermentas) and nuclease-free water to make up the volume. PCR reaction was performed using Applied Biosystems Gene Amp 9700 Thermal Cycler in 200 µl micro-tubes for 3 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 60°C, 1 min at 72°C and final extension at 72°C for 7 min. Ten microliters of PCR products were resolved on a 15% Native Polyacrylamide gel electrophoresis (PAGE), stained with ethidium bromide and photographed under UV light.

Sequencing and Cloning of PCR Product

Cloning of stem-loop RT–PCR products was done by using TOPO TA Cloning System (Invitrogen) based on manufacturer's protocol followed by sequencing of incorporated PCR products.

Target Prediction

We employed the previously reported target prediction algorithm (Singh and Nagaraju 2008) with a few minor improvisations as described below.

miRanda Program

The targets of 34 novel miRNAs were searched on the 3′UTR sequences of *B. mori* mRNAs by using miRanda program (Enright et al. 2003), and the parameters assigned were:

a) Smith-Waterman hybridization default alignment score (Smith and Waterman 1981) greater than or equal to 80.
b) Minimum free energy change (ΔG) between each miRNA::mRNA duplex less than or equal to −20 kcal/mol, and the rest of the parameters were kept as default.

Screening of miRNA-Target Alignments

Each of the selected miRNA-target alignments from miRanda was then scanned based on the following criteria to ensure less false positives:

a) No mismatches at the *seed region*, i.e., consecutive Watson–Crick matches on the position 2 to 8 at the 5′ end of a miRNA (Krek et al. 2005, Lewis et al. 2005).
b) Not more than one G:U pairing in the *seed region*.
c) Not more than three continuous gaps in the miRNA-target alignment.

Results and Discussion

Nearly 80 nucleotides long, characteristic stem-loop secondary structure of pre-miRNA is considered to be central feature for computational prediction of miRNAs. To predict all the possible stem-loop structures from *B. mori*, the whole genome was fragmented into several 80 nt long sequences. Each of these sequences was then scanned based on the different sequence and structural characteristics of pre-miRNAs, which showed more than 85% sensitivity when applied to the known miRNA data of miRBase. This criterion resulted in prediction of 253,322 hits from *B. mori* including Mirton precursors (Ruby et al. 2007). An overview of different steps involved in the miRNAs prediction is represented in Fig. 1.

The next important and most stringent filter was to check the phylogenetic conservation of these hits (Ambros et al. 2003). miRNAs are generally found to be evolutionarily conserved among closely related species and that is why phylogenetic conservation is considered to be one of the most reliable parameter to identify functional miRNAs (Stark et al. 2007a, b). When this study was initiated, *B. mori* was the only Lepidopteran species for which the complete genome sequence was available. Hence we searched for the phylogenetic conservation of selected pre-miRNAs from *B. mori* against

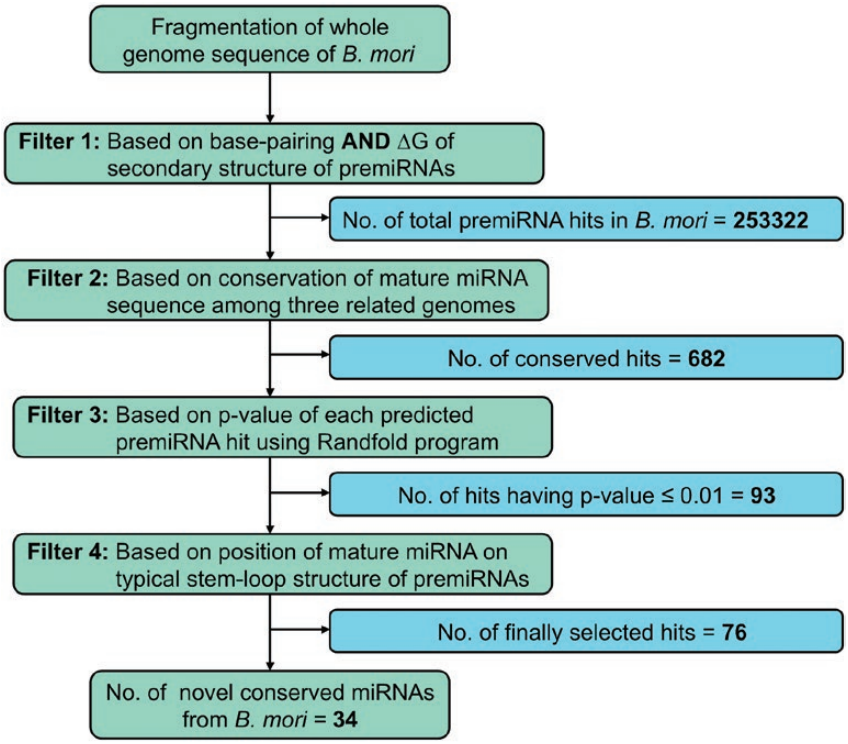

Fig. 1. Computational pipeline of different steps involved in genome-wide conserved miRNAs prediction.

Table 1. Details of 34 newly predicted miRNA sequences from *B. mori*

Newly predicted miRNAs	Chromosome number	Location	Size	Sequence
bmo-new-miR-1	Chr3	8819156–8819175	20	AAGGGGAAGGGGAAAGGGAA
bmo-new-miR-2	Chr16	3203390–3203371	20	ACCAACAGUGCAGCAGCAGC
bmo-new-miR-3	Chr10	11947380–11947398	19	ACGAGGACGAGGACGAGGA
bmo-new-miR-4	Chr18	636113–636094	20	UGUCGUCGUCGCUGUCGUCG
bmo-new-miR-4-1	Chr12	9955783–9955802	20	UGUCGUCGUCGCUGUCGUCG
bmo-new-miR-5	Chr7	6986974–6986956	19	UUUGUUUGUUUGUUUGUUU
bmo-new-miR-5a	Chr4	880559–880578	20	UUUGUUUGUUUGUUUGCUUG
bmo-new-miR-6	Chr7	12361884–12361866	19	UAUAUACAUAUGUAUAUAU
bmo-new-miR-7	Chr12	1968048–1968028	21	AGGAUAAUGAUGAUGAUGAUG
bmo-new-miR-8	Chr8	15887503–15887485	19	UCGUGUGUGUGUGGUGCCU
bmo-new-miR-9	Chr6	99564–99582	19	CACGCAUGCACACAAGCGC
bmo-new-miR-10	Chr22	16475902–16475884	19	UGUCUGUCUGUCUGUCUGU
bmo-new-miR-10–1	Chr9	8693199–8693217	19	UGUCUGUCUGUCUGUCUGU
bmo-new-miR-10–2	Chr14	4938190–4938172	19	UGUCUGUCUGUCUGUCUGU
bmo-new-miR-10–3	Chr21	1182118–1182100	19	UGUCUGUCUGUCUGUCUGU
bmo-new-miR-10–4	Chr21	13723615–13723597	19	UGUCUGUCUGUCUGUCUGU
bmo-new-miR-11	Chr23	16224343–16224363	21	UCUGUCUGUCUGUCUGUCUGU
bmo-new-miR-12	Chr22	8869827–8869808	20	UGUGUGCGAUCAGUGCGGCA
bmo- new-miR-13	Chr13	1003162–1003143	20	CGGUGGUGGUGGUGGUCAUC
bmo- new-miR-14	Chr1	2281662–2281641	22	UUAUGCAUGCGUGUUUGUGUGU
bmo- new-miR-15	Chr8	3622558–3622538	21	UAUAUAUACAUAUGUGUGUGU
bmo- new-miR-16	Chr5	11623004–11622980	25	UAUGUAGGUAUGUAUGUAUGUAUGU
bmo- new-miR-17	Chr22	10990852–10990831	22	UGGUGGUAGUAGUGGUGGUGGU
bmo- new-miR-17a	Chr18	2695654–2695675	22	UGGUGGUAGUGGUGGUGGUGGC
bmo- new-miR-17b	Chr11	8638959–8638980	22	UGGUGGUAGUGGUGGUGGUAGU
bmo- new-miR-18	Chr17	13957216–13957234	19	CUACAACAACAACAACAUC
bmo- new-miR-19	Chr18	12864315–12864297	19	ACCUCUGCGACUUGCCCCG
bmo- new-miR-20	Chr16	1545194–1545214	21	UCAUCAUCAUCAUGAUCAUCA
bmo- new-miR-21	Chr22	7371279–7371300	22	UCUAGCACCAUUCGAAUUCAGU
bmo- new-miR-22	Chr9	751375–751354	20	UCGCACAUCUGCAUGUCGUC
bmo- new-miR-23	Chr12	2976900–2976921	22	UUCCCAGCCCCCUCCCCCUUCC
bmo- new-miR-24	Chr14	10703188–10703167	22	UGGUCGUAACCAUCAGCACAUC
bmo- new-miR-25	Chr20	6846017–6846036	20	UGCUGCUGCUGUUGCAUCUG
bmo- new-miR-26	Chr23	649256–649237	20	AGGAUGGUGGUGGUGGUGGU

the most popular insect model organism, *D. melanogaster* for which the maximum number of miRNAs were reported in the miRBase database. The search returned 13,293 hits. To limit the false positives and increase the specificity of the program we further crosschecked the conserved hits of *B. mori* and *D. melanogaster* in another related Dipteran species, *A. gambiae*. This search resulted in 2,686 hits. All these 2,686 conserved hits in *B. mori*, *D. melanogaster*, and *A. gambiae* were then stringently screened for characteristic features of a typical pre-miRNA through the same filters as *B. mori*. This search reduced the hits to 628 potential pre-miRNAs, which were found to be conserved in both *D. melanogaster* and *A. gambiae*.

The statistical significance of the folding potential of each of these pre-miRNAs was then assessed by randomization test to rule out any possibility of getting a typical secondary structures of predicted pre-miRNA with a ΔG equal to or less than –25 kcal/mol merely by chance. The p-values were calculated using Randfold program and only those pre-miRNA hits, which had a *P*-value less than or equal to 0.01 were selected. This criterion has been found to trap more than 90% of the known pre-miRNAs (Bonnet et al. 2004).

Finally, 93 pre-miRNA sequences sorted after randomization test were analyzed by MFold program (Zuker and Stiegler 1981) for their secondary structures and matured miRNA location. The loop region of the pre-miRNA secondary structure is generally considered to be a variable region (Ambros et al. 2003), which is eliminated during the Dicer processing of pre-miRNA into imperfect miRNA:miRNA* duplex. Hence the conserved miRNA hits originating from the loop region of the pre-miRNAs were considered as false positives and excluded from the final list of 76 miRNAs. To check how many of these 76 newly predicted conserved miRNAs are already reported, the latest release of miRBase (release 15) was downloaded. Based on miRBase data, we found that there are 43 known miRNAs of *B. mori*, which are conserved in both *D. melanogaster* and *A. gambiae*. Out of these 43 miRNAs, 41 were included in the 76 predicted miRNAs along with an extra copy of *bmo-miR-2*. Details of all these 41 miRNAs are given in Supp Table S1 (online only). The two known miRNAs, which could not be predicted by this

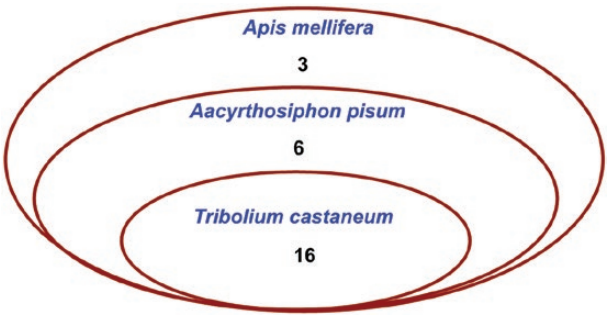

Fig. 2. Conservation of 34 newly predicted *B. mori* miRNAs in other insect species. *Apis mellifera* of order Hymenoptera, *Tribolium castaneum* of order Coleoptera and *Acyrthosiphon pisum* of order Homoptera.

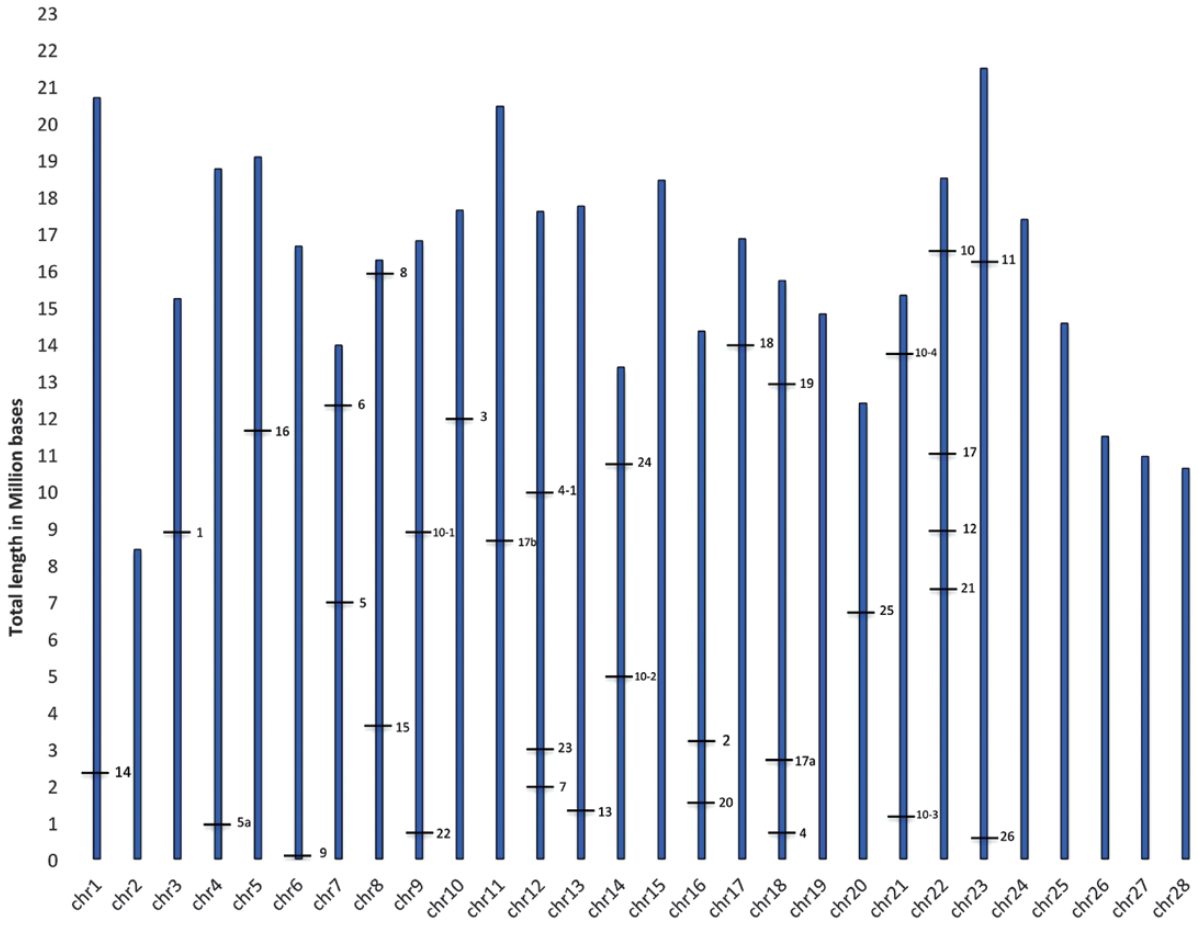

Fig. 3. Chromosomal distribution of 34 newly predicted miRNA genes in *B. mori*.

approach, were *bmo-miR-12* and *bmo-miR-276*. When the pre-miRNAs of these two miRNAs were analyzed, the pre-miRNA of *bmo-miR-12* was found to have ΔG much higher than –25kcal/mol whereas miRNA sequence of *bmo-miR-276* failed to map on to the genome of *B. mori*. The remaining 34 miRNAs of the total predicted 76 miRNAs have not been found to be reported in any other species.

Novel miRNAs

Details of the 34 newly predicted conserved miRNAs and their pre-miRNA sequences from *B. mori* are shown in Table 1 and Supp Table S3 (online only), respectively and secondary structures of their pre-miRNAs are presented in Supp File 2 (online only). These 34 pre-miRNAs accounts for 29 distinct miRNAs as the list includes two identical copies of *bmo-new-miR-4* and five copies of *bmo-new-miR-10*. Whereas *bmo-new-miR-5* and *bmo-new-miR-17* have two and three nearly identical copies with few mismatches respectively. An important characteristic feature of most of these miRNAs was its association with the repeat region of the genome. We also check the homologs of the newly predicted miRNAs in the available small RNA sequence data of *B. mori* but didn't found any hit. One of the plausible reasons could be masking of the repeat regions during small RNA sequence analysis.

Conservation of these 34 newly predicted miRNAs from *B. mori* were further analyzed in other insect species, i.e., *Apis mellifera* (Hymenoptera: Apidae), *Tribolium castaneum* (Coleoptera:

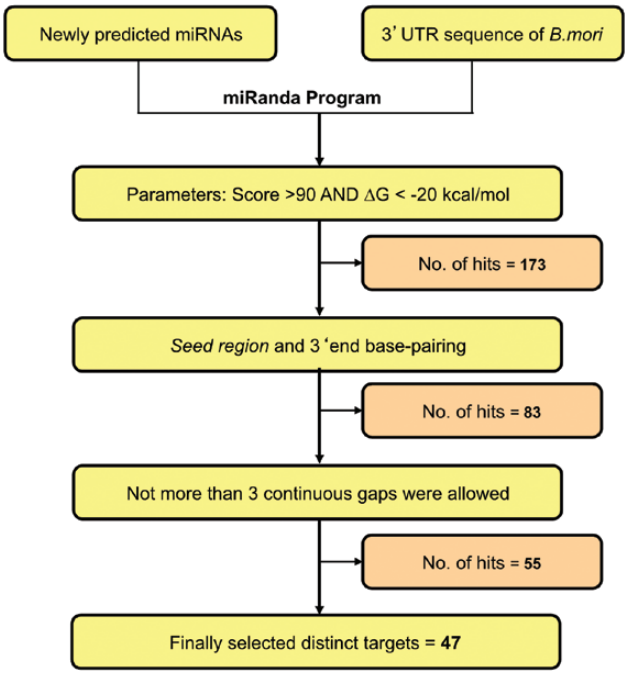

Fig. 4. An overview of different steps involved in target prediction of 34 newly predicted miRNAs in *B. mori*.

Table 2. List of putative mRNA targets of newly predicted miRNAs, and their known function in *B. mori*

Accession ID	miRNAs	3′UTR position	Known functions	ΔG (kcal/mol)	miRNA::mRNA alignments
DQ443281	new-miR-17	98 122	H+ transporting ATPase V0 subunit D mRNA.	−31.12	3′ TGGT--GGTGGTGA-TGATGGTGGT 5′ | || ||||||||	|:||||||| 5′ AGCATGCCACCACTCAGTGCCACCA 3′
	new-miR-17a	103 122		−33.15	3′ CGGTGGTGGTGGTGATGGTGGT 5′ |||||||||:|| |:||||||| 5′ GCCACCAC--TCAGTGCCACCA 3′
	new-miR-17b	100 122		−35.02	3′ tgATGGTGGTG-GTGATGGTGGT 5′ |:|||||||| || |:||||||| 5′ caTGCCACCACTCAGTGCCACCA 3′
DQ311148	new-miR-17	99 118	Chaperonin subunit 4 delta mRNA.	−27.55	3′ TGGTGGTGGTGATGATGGTGGT 5′ | || |||||	||:||||||| 5′ AACA-CACCAAT-CTGCCACCA 3′
	new-miR-17a	97 118		−27.55	3′ CGGTG-GTGGTGGTGATGGTGGT 5′ |: || ||||||: ||:||||||| 5′ GTAACACACCAAT-CTGCCACCA 3′
	new-miR-17b	96 118		−27.4	3′ TGATGGTGGTGGT--GATGGTGGT 5′ | || || ||||| ||:||||||| 5′ AGTAACA-CACCAATCTGCCACCA 3′
S77508	new-miR-4	242 261	Cysteine proteinase.	−22.54	3′ gcTGCTGTCGCTGCTGCTGt 5′ |	|| ||:||||||| 5′ gtAGGCCAACGGCGACGACc 3′
EF554692	new-miR-11	485 505	MSL3 protein mRNA.	−23.31	3′ TGT-CTGTCTGTCTGTCTGTCT 5′ ||: ||||:|| ||||||||| 5′ ACGCGACATGCA-ACAGACAGA 3′
DQ533987	new-miR-9	203 223	Transport protein Sec61 alpha subunit.	−24.17	3′ CGCGAAC--ACACGTACGCAC 5′ |:|| ||	||:||||||| 5′ GTGCGTGGCGGCGCGTGCGTG 3′
DQ443409	new-miR-26	75 96	Mariner Transposase.	−27.51	3′ TGGTGGTGGTG--GTGGTAGGA 5′ ||::| |||:| ||:||||||| 5′ ACTGCAACCGCCGCATCATCCT 3′
DQ443269	new-miR-26	10 29	Secreted protein acidic and rich in cysteine mRNA.	−27.08	3′ tGGTGGTGGTGGTGGTAGGA 5′ |:||	||||:||||||:| 5′ gCTACAAGCACCGCCATCTT 3′
DQ443199	new-miR-17a	14 36	Interleukin enhancer binding factor isoform 1.	−29.35	3′ CGGTGGTGGTGGTG-ATGGTGGT 5′ | ||:|:	|||| |||||:|| 5′ GACATCGGCTCCACTTACCATCA 3′
DQ443170	new-miR-26	1 17	Methylthioadenosine phosphorylase.	−25.18	3′ tggTGGTGGTGGTGGTAGGA 5′ |||:|| |||||||:| 5′ ---ACCGCCCGCACCATCTT 3′
DQ443140	new-miR-22	32 53	Septin mRNA.	−20.73	3′ ctGCTGTACG---TCTACACGCT 5′ ||:| |||	|||||||| 5′ ttCGGC-TGCCTGACATGTGCGA 3′
DQ311393	new-miR-1	585 607	p53-related protein kinase-binding protein.	−24.28	3′ aaGGG--AAAGG-GGAAGGGGAa 5′ ||| |||||::||||||| 5′ acCCCAATTTCCTTTTTCCCCTa 3′
DQ311391	new-miR-3	52 69	Proteasome 26S non-ATPase subunit 9.	−26.25	3′ aGGAGCAGGAGCAGGAGCa 5′ || ||| |||||||| 5′ aCC-GGTCACCGTCCTCGg 3′
DQ311384	new-miR-24	65 88	Nucleoplasmin isoform 2 mRNA.	−22.89	3′ ctACACGAC--TACCAATGCTGGT 5′ ||||| |: || |:||||||| 5′ atTGTGCAGGCGAGGGTGCGACCA 3′
DQ311315	new-miR-8	2 19	Stathmin mRNA.	−20.35	3′ tCCGTGGTGTGTGTGTGCt 5′ ||| ||	||||||| 5′ cGGC-CCTCCGACACACGc 3′
DQ311184	new-miR-21	30 55	Guanylate kinase mRNA.	−21.33	3′ TGAC--TTAAGCTT--ACCACGATCT 5′ ||||	|| ||| |||||:|||| 5′ ACTGCTATTTAGAAATTGGTGTTAGA 3′
AY970687	new-miR-13	16 35	Allototropin mRNA.	−24.91	3′ CTACTGGTGGTGGTGGTGGc 5′ |||	|||:||:|||||| 5′ GATAAGCACTACTACCACCc 3′

Table 2. *Continued*

Accession ID	miRNAs	3′UTR position	Known functions	ΔG (kcal/mol)	miRNA::mRNA alignments
AY753659	new-miR-4	435 453	Zinc finger protein.	–27.6	3′ gctGCTGTCGCTGCTGCTGt 5′ |||| ||| ||||||| 5′ ttgCGAC-GCGTCGACGACt 3′
AY297158	new-miR-9	37 54	Calreticulin mRNA.	–29.47	3′ cGCGAACACACGTACGCAC 5′ || | ||||||:|||||| 5′ aCG-TAGTGTGCGTGCGTG 3′
AF237663	new-miR-4	26 45	Nuclear orphan receptor (OR).	–26.43	3′ gCTGCTGTCGCTGCTGCTGT 5′ |||| | || |||||||: 5′ aGACGTCTGCATCGACGACG 3
AF013572	new-miR-9	13 31	Small GTP-binding protein (rabB).	–25.37	3′ cGCGAACACACGTACGCAC 5′ | | | ||||:|||||| 5′ tCACGAGCGTGCGTGCGTG 3′
AB302934	new-miR-4	452 473	BmJnk mRNA for c-Jun NH2-terminal kinase.	–32.26	3′ GCT---GCTGTCGCTGCTGCTGt 5′ ||| |||| |||:||||||| 5′ CGAACCCGAC-GCGGCGACGACt 3′
AB274989	new-miR-20	9 29	3-hydroxy-3-methylglutaryl-CoA synthase.	–22.1	3′ acTACTAGTACTACTACTACt 5′ | ||| ||||||||||| 5′ gcAGGATAATGATGATGATGt 3′
AB273625	new-miR-2	125 145	mRNA for histone H2A.Z.	–21.35	3′ CGACGACGA-CGTGACAACCA 5′ | ||: || | |||||||| 5′ GGTGTGACTGGAACTGTTGGT 3′
AB252485	new-miR-24	52 71	mRNA for ecdysteroid 22-phosphate.	–23.43	3′ CTACACGACTACCAATGCTGGT 5′ ||||| |||| | ||:||||| 5′ GATGT-CTGA-GCGTATGACCA 3′
AB206555	new-miR-11	1354 1377	Cyp307a1/spook mRNA for cytochrome P450.	–23.68	3′ TGTCTG---TCTGTCTGTCTGTCT 5′ ||:||| ||| | ||:|||||| 5′ ACGGACACAAGAAAAACGGACAGA 3′
AB188256	new-miR-11	1720 1739	Myosuppressin receptor.	–24.18	3′ tgtCTGTCTGTCTGTCTGTCT 5′ ||: ||| ||||||||| 5′ tatGATCGAC-GACAGACAGA 3′
AB186518	new-miR-22	1266 1284	Olfactory receptor-like receptor mRNA.	–22.74	3′ cTGCTGTACGTCTACACGCt 5′ |:| |||||:|||||| 5′ tATG-CATGCATGTGTGCGt 3′
AB183870	new-miR-8	150 168	MAP kinse-ERK kinase mRNA.	–28.26	3′ tCCGTGGTG-TGTGTGTGCT 5′ || |||:| ||||||:||| 5′ cGG-ACCGCAACACACGCGA 3′
AB072309	new-miR-4	152 173	mRNA for beta-tubulin (bmtub3)	–23.33	3′ gCTGCTGTCGCT--GCTGCTGt 5′ | | |:|||: ||||||| 5′ tGTCCCCGGCGGACCGACGACc 3′
AB030701	new-miR-16	25 48	mRNA for Promoting Protein.	–22.43	3′ TGTATGTATGTATGTATGGATGTAT 5′ |||:| |:::||::||||||||| 5′ ACACGC-TGTGTATGTACCTACATA 3′
AB013386	new-miR-19	205 224	mRNA for soluble alkaline phosphatase.	–23.31	3′ GCCCCGTTCAGCGTCTCCa 5′ || |||:: | |||||| 5′ CGCGGCGGTTGGCAGAGGg 3′

Tenebrionidae) and *Acyrthosiphon pisum* (Homoptera: Aphididae). Of the 34 miRNAs, 25 were found to be conserved in *A. mellifera*, out of which 22 and 16 miRNAs were also conserved in *A. pisum* and *T. castaneum*, respectively (Fig. 2). Details of these miRNAs are given in Supp Table S4 (online only), 16 miRNAs conserved in all the three species is shown in blue color whereas the six miRNAs that are conserved in *A. pisum and A. mellifera*, and three which are conserved only in *A. mellifera* are represented in green and red color, respectively.

Chromosomal Mapping

The newly predicted 34 miRNA genes were mapped on to *B. mori* chromosomes using the recent high-quality genome assembly of *B. mori* (Kawamoto et al. 2019) from Silkbase (http://silkbase.ab.a.u-tokyo.ac.jp). As shown in the Fig. 3, except for *bmo-new-8* and *bmo-new-9*, which were mapped towards ends of Chr8 and Chr6, respectively, rests all were showed distribution at random over the entire length of almost every chromosome (Fig. 3). Although, many of the miRNAs were found to be located on to a single chromosome like Chr12, Chr22, and Chr17, none of them were parts of any cluster. They all were found to be evenly spread along the chromosome that provides evidence for miRNAs mediated regulation coverage of diverse target genes.

Expression Validation of Newly Predicted miRNAs

To experimentally validate the expression of 34 newly predicted miRNAs, which actually account for 29 distinct miRNA sequences,

stem-loop RT–PCR was carried out using specific primers for each of the 29 miRNAs in BmN cell lines. Out of these 29 miRNAs, 23 were expressed in BmN cells (as shown in Supp Fig. S1 [online only]). Since miRNAs show differential spatial expression, we next checked the expression of rest six of the miRNAs, i.e., *bmo-New-miR-3*, *bmo-New-miR-4*, *bmo-New-miR-11*, *bmo-New-miR-21*, *bmo-New-miR-23* and *bmo-New-miR-24* in other tissues including fat body, silk gland, midgut, hemolymph, and the whole *B. mori* body. Only two of six miRNAs, *bmo-New-miR-3* and *bmo-New-miR-4* showed expression in fat body and midgut tissues. We did not observe expression of *bmo-New-miR-11*, *bmo-New-miR-21*, *bmo-New-miR-23,* and *bmo-New-miR-24* in any of the selected tissues and this could be because of differential temporal expression of these miRNAs. Altogether, we observe expression of 30 out of 34 newly predicted miRNAs in *B. mori,* indicating a very high sensitivity and specificity of our miRNA prediction algorithm.

Target Prediction

miRNAs primarily target 3′UTRs of the target mRNAs (Brennecke et al. 2003, Lin et al. 2003) and repress the protein production either by blocking the translation machinery or degrading the target mRNA. However, some of the reports by Vasudevan et al. 2007) suggest that miRNAs also up-regulate protein production. The basic hurdle in prediction of miRNA targets in animals is that unlike plants, they usually do not exhibit complete sequence complementarity. Hence, there is no perfect thumb rule for the prediction of miRNA targets in animals and it is still an unsolved problem. The most commonly used base-pairing rule, which is also well established through comparative genomic and experimental studies for the prediction of potential binding sites on target mRNA, is consecutive Watson–Crick matches on the positions 2 to 8 at the 5′ end of a miRNA with the 3′UTR of the mRNA sequence (Krek et al. 2005, Lewis et al. 2005), often called as *seed region*. The other empirical constraints include a positive balance of minimum free energy (ΔG) for local miRNA::mRNA interaction and limited gaps (Singh and Nagaraju 2008). Different target prediction methods use different sets of empirical rules and result in different rates of false positives.

When we used miRanda program for target prediction, we took all these constraints into consideration and used many stringent parameters to ensure minimum false positives. Besides, we have given a little flexibility in the *seed region* to increase the sensitivity of the program.

Mapping of 29 distinct miRNAs on to the 3′UTR sequences of *B. mori* using miRanda along with different parameters yielded 173 hits. These hits were then scanned for the *seed region* complementarity, which restricted the number of hits to 83. To give extra strength to the local miRNA::mRNA duplex, at least two base-pairing at 3′ end of the miRNA and less than three continuous gaps, were taken into consideration (Enright et al. 2003). Finally, we obtained 55 target hits of 23 miRNAs on 47 distinct mRNAs. An overview of complete computational pipeline of target prediction in *B. mori* is described in Fig. 4.

Many of these miRNAs were found to have multiple targets. A complete list of targets per miRNA is given in Supp Table S5 (online only). miRNA *bmo-new-miRNA-4* was found to have maximum target hits followed by *bmo-new-miRNA-8* and *bmo-new-miRNA-26*. Putative functions of all the predicted targets were assigned using BLASTn against the nonredundant nucleotides database of NCBI.

Details of the targets with known functions are tabulated in Table 2. These targets include diverse range of functions, suggesting multiple layers of gene regulation of various important biological processes.

To conclude, although we have predicted a considerable number of novel conserved miRNAs by applying approaches based on various characteristic features of miRNAs and their precursors, yet we still believe that there are many more conserved miRNAs to be uncovered, especially those that do not follow the canonical pathway of biogenesis. Besides, prediction of species-specific miRNAs is one of the biggest limitations of all the computational approaches. The next major step to do is a functional analysis of targets of these newly predicted repeat-associated miRNAs to understand their biological function.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

This work was supported by the Wellcome Trust/DBT India Alliance Fellowship (IA/E/16/1/503020) awarded to J.S. U.A. and J.S. also acknowledge financial support from the Department of Science and Technology, Ministry of Science and Technology, Government of India under INSPIRE-Faculty fellowship (IFA12-LSBM-25). We declare no conflict of interest.

References Cited

Ambros, V., B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen, G. Dreyfuss, S. R. Eddy, S. Griffiths-Jones, M. Marshall, et al. **2003**. A uniform system for microRNA annotation. RNA. 9: 277–279.

Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 116: 281–297.

Bonnet, E., J. Wuyts, P. Rouzé, and Y. Van de Peer. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics. 20: 2911–2917.

Brennecke, J., D. R. Hipfner, A. Stark, R. B. Russell, and S. M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell. 113: 25–36.

Enright, A. J., B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks. 2003. MicroRNA targets in *Drosophila*. Genome Biol. 5: R1.

Griffiths-Jones, S. 2004. The microRNA registry. Nucleic Acids Res. 32: D109–D111.

Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell. 106: 23–34.

Han, Y., B. Zhu, F. Luan, H. Zhu, Y. Shao, A. Chen, C. Lu, and Y. Luo. 2010. Conserved miRNAs and their targets identified in lettuce (Lactuca) by EST analysis. Gene. 463: 1–7.

Hutvagner, G., J. McLachlan, A. E. Pasquinelli, E. Bálint, T. Tuschl, and P. D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 293: 834–838.

Joshi, T., Z. Yan, M. Libault, D. H. Jeong, S. Park, P. J. Green, D. J. Sherrier, A. Farmer, G. May, B. C. Meyers, et al. **2010**. Prediction of novel miRNAs and associated target genes in Glycine max. BMC Bioinf. 11 (Suppl 1): S14.

Kawamoto, M., A. Jouraku, A. Toyoda, K. Yokoi, Y. Minakuchi, S. Katsuma, A. Fujiyama, T. Kiuchi, K. Yamamoto, and T. Shimada. 2019. High-quality genome assembly of the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 107: 53–62.

Ketting, R. F., S. E. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, and R. H. Plasterk. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes Dev 15: 2654–2659.

Kim, V. N., J. Han, and M. C. Siomi. 2009. Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. 10: 126–139.

Krek, A., D. Grün, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, et al. **2005**. Combinatorial microRNA target predictions. Nat. Genet. 37: 495–500.

Lai, E. C., P. Tomancak, R. W. Williams, and G. M. Rubin. 2003. Computational identification of *Drosophila* microRNA genes. Genome Biol. 4: R42.

Lau, N. C., L. P. Lim, E. G. Weinstein, and D. P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science. 294: 858–862.

Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 75: 843–854.

Lee, Y., K. Jeon, J. T. Lee, S. Kim, and V. N. Kim. 2002. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. 21: 4663–4670.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Rådmark, S. Kim, et al. **2003**. The nuclear RNase III Drosha initiates microRNA processing. Nature. 425: 415–419.

Lewis, B. P., C. B. Burge, and D. P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 120: 15–20.

Lim, L. P., M. E. Glasner, S. Yekta, C. B. Burge, and D. P. Bartel. 2003. Vertebrate microRNA genes. Science. 299: 1540.

Lin, S. Y., S. M. Johnson, M. Abraham, M. C. Vella, A. Pasquinelli, C. Gamberi, E. Gottlieb, and F. J. Slack. 2003. The *C elegans* hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev. Cell 4: 639–650.

Mendes, N. D., A. T. Freitas, and M. F. Sagot. 2009. Current tools for the identification of miRNA genes and their targets. Nucleic Acids Res. 37: 2419–2433.

Mishra, A. K., and D. K. Lobiyal. 2010. Prediction of miRNAs in *Bombyx mori* through computational approaches. Int. J. Comput Theory Eng. 2: 1793–8201.

O'Toole, A. S., S. Miller, N. Haines, M. C. Zink, and M. J. Serra. 2006. Comprehensive thermodynamic analysis of 3' double-nucleotide overhangs neighboring Watson-Crick terminal base pairs. Nucleic Acids Res. 34: 3338–3344.

Okada, C., E. Yamashita, S. J. Lee, S. Shibata, J. Katahira, A. Nakagawa, Y. Yoneda, and T. Tsukihara. 2009. A high-resolution structure of the pre-microRNA nuclear export machinery. Science. 326: 1275–1279.

Pfeffer, S., M. Zavolan, F. A. Grässer, M. Chien, J. J. Russo, J. Ju, B. John, A. J. Enright, D. Marks, C. Sander, et al. **2004**. Identification of virus-encoded microRNAs. Science. 304: 734–736.

Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature. 403: 901–906.

Ruby, J. G., C. H. Jan, and D. P. Bartel. 2007. Intronic microRNA precursors that bypass Drosha processing. Nature. 448: 83–86.

Schwarz, D. S., Y. Tomari, and P. D. Zamore. 2004. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr. Biol. 14: 787–791.

Singh, J., and J. Nagaraju. 2008. In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). Insect Mol. Biol. 17: 427–436.

Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. J. Mol. Biol. 147: 195–197.

Stark, A., P. Kheradpour, L. Parts, J. Brennecke, E. Hodges, G. J. Hannon, and M. Kellis. 2007a. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. Genome Res. 17: 1865–1879.

Stark, A., M. F. Lin, P. Kheradpour, J. S. Pedersen, L. Parts, J. W. Carlson, M. A. Crosby, M. D. Rasmussen, S. Roy, A. N. Deoras, et al.; Harvard FlyBase curators; Berkeley Drosophila Genome Project. **2007b**. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. Nature. 450: 219–232.

Vasudevan, S., Y. Tong, and J. A. Steitz. 2007. Switching from repression to activation: microRNAs can up-regulate translation. Science. 318: 1931–1934.

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

Yi, R., Y. Qin, I. G. Macara, and B. R. Cullen. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17: 3011–3016.

Yin, Z. J., and F. F. Shen. 2010. Identification and characterization of conserved microRNAs and their target genes in wheat (*Triticum aestivum*). Genet. Mol. Res. 9: 1186–1196.

Yu, X., Q. Zhou, S. C. Li, Q. Luo, Y. Cai, W. C. Lin, H. Chen, Y. Yang, S. Hu, and J. Yu. 2008. The silkworm (*Bombyx mori*) microRNAs and their expressions in multiple developmental stages. PLoS One 3: e2997.

Zhang, B. H., X. P. Pan, and T. A. Anderson. 2006. Identification of 188 conserved maize microRNAs and their targets. FEBS Lett. 580: 3753–3762.

Zhang, B. H., Q. L. Wang, K. B. Wang, X. P. Pan, F. Liu, T. L. Guo, G. P. Cobb, and T. A. Anderson. 2007. Identification of cotton microRNAs and their targets. Gene. 397: 26–37.

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

Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9: 133–148.