Mol Biol Rep (2010) 37:3171–3176
DOI 10.1007/s11033-009-9897-4

Computational identification and characteristics of novel microRNAs from the silkworm (*Bombyx mori* L.)

Yong Huang · Quan Zou · Shun Ming Tang · Li Gang Wang · Xing Jia Shen

Received: 29 September 2009 / Accepted: 2 October 2009 / Published online: 13 October 2009

Abstract MicroRNAs (miRNAs) are a class of non-protein coding small RNAs that regulate expression of genes at post-transcriptional levels. Increasing evidence has shown that miRNAs play multiple roles in biological processes, including development, cell proliferation and apoptosis. Based on the conservation of miRNAs sequence, using a computational homology search based on genomic survey sequence analysis, a total of 16 novel miRNAs were identified and characteristics such as family and evolutionary conservation have been described. By using these newly identified miRNAs, the mRNA database of silkworm was blasted and 21 potential targets of miRNAs were detected. Most of these miRNA targeted genes were predicted to encode transcription factors. The semi-stem loop RT–PCR based assays were performed and found that silkworm miRNAs have diverse expression patterns during development.

Electronic supplementary material The online version of this article (doi:10.1007/s11033-009-9897-4) contains supplementary material, which is available to authorized users.

Y. Huang · S. M. Tang · L. G. Wang · X. J. Shen
Jiangsu University of Science and Technology, Zhenjiang City, Jiangsu Province 212018, People's Republic of China

Q. Zou
School of Information Science and Technology of Xiamen University, Xiamen City, Fujian Province 361005, People's Republic of China

Y. Huang · S. M. Tang · L. G. Wang · X. J. Shen (✉)
The Key Laboratory of Silkworm and Mulberry Genetic Improvement, Ministry of Agriculture, Sericultural Research Institute, Chinese Academy of Agricultural Sciences, Zhenjiang City, Jiangsu Province 212018, People's Republic of China
e-mail: shenxj63@yahoo.com.cn

Keywords MicroRNA · *Bombyx mori* · Computational prediction · Target gene

Introduction

MicroRNAs (miRNAs) are a large class of small non-coding RNAs approximately 22 nucleotides (nt) in length. They play important roles in post-transcriptional gene control in animals by binding on target mRNAs perfectly to induce mRNA cleavage or imperfectly within the 3′ untranslated region (3′ UTR) to repress protein synthesis [1]. Mature miRNAs are typically generated via a two-step processing pathway. Primary miRNAs (pri-miR) transcribed by polymerase II or III are processed by the nuclear enzyme Drosha to produce precursor-miRNAs (pre-miR), which are then exported into the cytoplasm and further processed into mature miRNAs by Dicer [2]. The accumulated evidence indicates that miRNAs play key roles in control a broad range of biological activities including embryonic development, cell proliferation, metabolic homeostasis and apoptosis [3].

There are several experimental approaches for identification of miRNAs. The genetic screening method is initially used in identification of lin-4 and let-7 [4, 5]. But it is expensive, time consuming and inefficient. The second one is direct cloning, by which the discovery efficiency of novel miRNAs was improved [6, 7]. However, this method also has several disadvantages, i.e., difficulty to find miRNAs expressed at a low level and degradation of RNAs during sample separation. Recently, computational approach has been developed and a large number of miRNAs from animals to plants and viruses were identified with this method [8–11]. The distinct advantage of the computational approach is that the miRNAs expressed in specific tissues at

certain stages of development or at low-copy number can be readily identified, whereas they are difficult to identify by cloning and sequencing methods [12].

The silkworm, as we all known, is one of the most importantly model insect for scientific discovery in the areas of microbiology, physiology and genetics. At present over 9,000 miRNAs have been identified in various species, including 152 ones from *Drosophila melanogaster*, 73 from *Drosophila pseudoobscura*, 62 from *Apis mellifera*, 66 from *Anopheles gambiae*, 155 from *Caenorhabditis elegans*, and 706 human miRNAs [miRBase release 13 March 2009]. Studies have been performed on identifying miRNAs in the silkworm by computational and experimental approaches [13–16] and totally 55 miRNAs have been identified on the miRBase database. But these miRNAs are less than those in other animal species. So much effort is still needed to find novel potential miRNA genes in the silkworm. In this study, by computational homology search approach, 16 potential miRNAs in silkworm were indentified.

Materials and methods

Referenced sequence data

The animal mature microRNAs, totally 8,478, used in the experiment were downloaded form the miRNA database (miRBase Sequence Database, http://microrna.sanger.ac.uk; released 13.0 March 2009). The genome assembly and some functional annotations were downloaded from the SilkDB (http://silkworm.genomics.org.cn/) and Silkworm Genome Research Program (http://sgp.dna.affrc.go.jp/index.html).

Computational prediction of microRNAs

Comparative software BLAST-2.2.20 was downloaded from NCBI GenBank. Procedure of search for potential miRNAs in the silkworm was showed in Fig. 1. Four criteria were used to predict the silkworm miRNAs and pre-miRNAs: (1) predicted mature miRNAs were allowed to have only 0–4 nucleotide mismatches in sequence with all previously known animal mature miRNAs; (2) pre-miRNAs sequence can fold into an appropriate hairpin secondary structure that contains the ~22 nt mature miRNA sequence within one arm of the hairpin; (3) miRNA precursors with secondary structures had higher negative minimal free energies (MFEs) and minimal free energy index (MFEIs) than other different types of RNAs by RNA-fold prediction; and (4) miRNA had 15–70% contents of A + U by SMV (support vector machine).

Prediction of targets of microRNAs

The mRNA database of the silkworm downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) were extracted and potential target sites for the silkworm miRNAs sequences were detected with the RNAhybrid program [17, 18]. The parameters employed were described as follows: *P* value cutoff of 0.05, target duplex free energy $\Delta G \leq -24$ kcal/mol. The miRNA sequences and potential mRNAs targets were no more than four gaps.

RT–PCR assay

The silkworm small RNA samples from four different development stages (embryo, 5th instar larva, pupa and moth) were isolated using mirVana^{TM} miRNA isolation kit (Ambion) according to the manufacturer's instruction. The cDNAs were synthesized from small RNAs by using miRNA specific stem–loop primers according to criteria mentioned as described previously [16, 19]. The stem–loop RT primers and gene specific primers were listed in Table S1. The DNA fragments were directly subcloned into pMD18-T vector (Takara) and sequenced.

Analysis of mature miRNA expression by RT–PCR was carried out with the method as described previously [16, 20]. The cDNAs were diluted 10 times to perform PCR for expression confirmation and expression pattern analysis. PCRs were performed, respectively, in 20 μl mixture containing 1 μl cDNA, 0.5 μM forward and reverse primers, 10× PCR buffer, 0.25 μM each of dNTPs (Takara) and 2 U *Taq* polymerase (Takara), under the following parameters: 94°C for 30 s, 55°C for 30 s and 72°C for 30 s. The PCR products were detected by electrophoresis with 3% agarose gel containing ethidium and photographed under UV light. Expression of U6 snRNA gene was performed for cDNA normalization.

Results and discussion

Identification and characteristics of potential miRNAs in the silkworm

The homology search procedure was adopted to identify candidate miRNA genes based on a query animal mature miRNAs. After removing the redundant sequences and the sequences derived form other non-coding RNAs and protein-coding sequences, 14 miRNA candidates were identified in 3′ or 5′ arm of the stem–loop hairpin structures (Fig. S1). Compared with those predicted by Zhang and Cao et al. in the silkworm [21, 22], all these 16 miRNAs are novel.

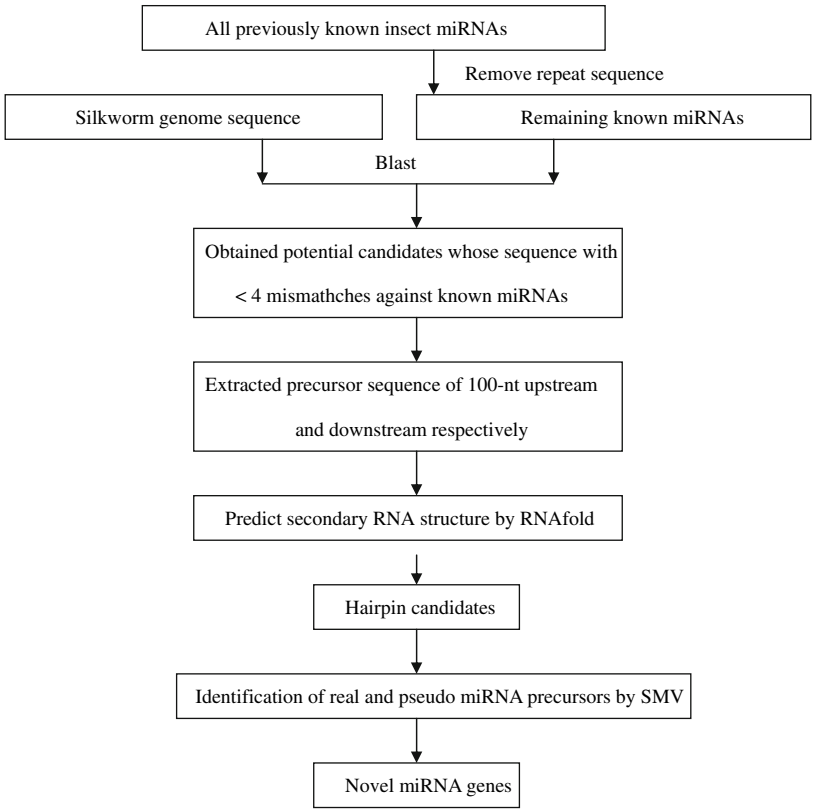

Fig. 1 Procedure of potential silkworm miRNA gene search by identifying homologs of previously known miRNA gene

The mature silkworm miRNA sequences predicted were 18–23 nt in length centering on 20–22 nt. This is coincident with the specificity of Dicer processing [23]. The length of these newly identified silkworm miRNA precursors varied from 64–123 nt with an average of 98 nt (Table 1). The A + U content of these miRNAs were evaluated which range from 17.0 to 68% in agreement with the point of view that miRNA precursors and mature miRNAs contain more A + U than G + C [24]. The miRNA precursors, unlike other non-coding RNAs, have lower folding free energy than random sequence. Minimal folding free energy is one of the important features to identify new miRNA genes [25–27]. Analysis by the SMV software revealed these newly identified silkworm miRNA precursors have negative minimal folding free energies ranging from −23.00 to −59.40 kcal mol^{-1} with an average of 45.93 kcal mol^{-1}, implying that these are silkworm miRNA precursors.

The 16 indentified novel miRNAs represent 12 families in the silkworm (Fig. 2). MiR-1268, miR-1638, miR-2285 and miR-1804 have only one member; miR-989 and miR-561 have two members; miR-iab-4as belongs to the miR-iab-4 family which contains 17 members; other miRNAs have more than two members. Noticeably, no homologous of miR-1516, miR-870 and miR-2285 was detected in other species by homology search, implying that these four miRNAs are unique in silkworm. Five of the predicted miRNAs, miR-989, miR-1000, miR-iab-4as, miR-993 and miR-316 are conservative in insects. These five miRNAs do not have orthologous in mammals and nematode, but are conserved in the insect species. Sequence alignment of pre-miR-iab-4 family strongly suggests that miRNAs of this family not only hold high similarity in sequences, but also present conserved organization pattern among insects in evolution (Fig. S2). The miRNAs conserved in sequence are often expressed within identical tissues during analogous developmental stages, suggesting that these conserved miRNAs among distantly species might prescribe some fundamental development events in these metazoans [22].

In contrast, homologs of miRNAs of miR-370, miR-1268, miR-561, miR-1638, miR-1804, miR-1175, miR-2444 and miR-183 were identified in other animal species. Interesting, miR-183 family has 25 members and 24 of them existed in non-insect, but remaining miR-183 which was only detected in the silkworm. Phylogenetic tree revealed silkworm miR-183 had closer evolution relationship with *Lottia gigantean* miRNA than the other species analyzed (Fig. S3). It may be due to the high conservation of miRNAs, as organisms that have close phylogeny relationship usually share a number of miRNAs, although their pre-miRNAs might be dramatically different. It showed that miR-183 family has function diversity and conservation among species to some extent and also imply that miRNAs arose after the split of insects and other

Table 1 The 16 novel miRNAs indentified in the silkworm

New miRNAs	Gene source	Mature sequence	Side	NM (nt)	Strand	LP (nt)	A + U (%)	MFEs	MFEIs
370	scaf110	aggagcuggggguggaaccuggu	3′	3	Minus	123	36	49.20	1.11
989	scaf1	ugugaugugacguaguggaag	3′	1	Minus	89	52	36.70	0.792
1000	scaf103	auauuguccugucacagcagu	5′	0	Plus	82	57	53.80	1.15
	scaf25284				Plus	96	17	58.20	3.56
1268	scaf6706	ggggcgugguggugggggg	5′	1	Minus	107	32	59.40	1.73
	scaf25				Minus	107	32	59.40	1.73
561	scaf62	gcaaguuuaagauccuugacug	3′	2	Minus	92	59	27.50	0.506
870	scaf103	cuauuugguguuucuucgggu	5′	0	Minus	112	57	47.70	0.747
1516	scaf24	caaaagaacuuauggcccauc	3′	0	Plus	89	57	35.62	0.702
1638	scaf102	uaguuuuguguuguuuguuu	3′	2	Minus	74	74	23.21	0.42
1804	scaf26	uuuuaaguugugaaagcugcaa	3′	3	Plus	100	68	29.92	0.44
iab-4as	scaf11	uuacguauacugaagguauaccg	5′	0	Minus	88	61	39.34	0.732
993	scaf72	aaagcucgucucuacagguauau	3′	2	Minus	86	61	44.20	0.842
1175	scaf48	ugagauucuacuucuccgacuuaa	3′	2	Minus	64	62	23.00	0.579
316	scaf24	ugucuuuuuccgcuuugcugcug	5′	2	Plus	89	50	31.90	0.716
183	scaf13	aauggcacuggaagaauucacgg	5′	0	Minus	86	52	32.86	0.733
2285	scaf109	caaaucugagugaacuuuuacu	3′	0	Minus	117	56	55.30	0.844
2444	scaf20117	aauguguuguuuuuguuuu	3′	2	Minus	66	52	27.60	0.804

NM number of mismatch, *LM* length of mature miRNAs, *LP* length of precursor, *MFEs* minimal folding free energies, *MFEIs* minimal folding free energy indexes

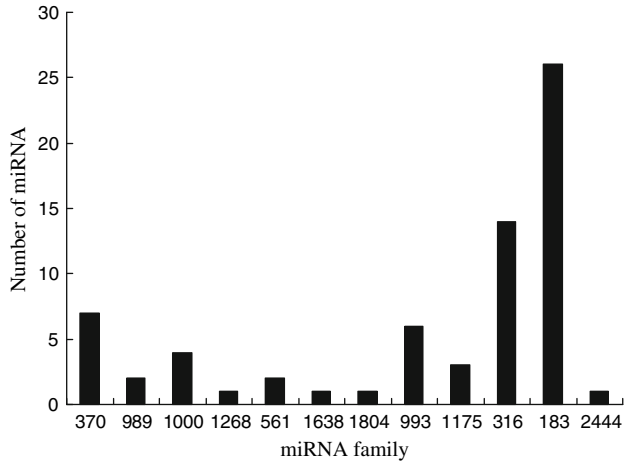

Fig. 2 Size of miRNA families in the silkworm

invertebrates. Phylogenetic analysis demonstrated that miR-183 family appeared after the split of different species orders, suggesting that some miRNAs have undergone dynamic evolutionary changes [22]. The miRNAs are still evolutionarily active and are undergoing a rapid ''birth and death'' within Drosophila [28], and some miRNAs even duplicate within the genome [29]. Phylogenetic conservation and diversity of miRNAs suggest that miRNAs play important roles in animal evolution via driving phenotypic variation during development [30].

Expression analysis of the silkworm miRNAs

To validate the prediction of miRNAs in the silkworm, RT–PCR based assay were performed for 10 miRNAs including miR-370, miR-561, miR-870, miR-989, miR-1804, miR-1000, miR-1268, miR-1638 and miR-iab-4as, which were chosen from the identified miRNAs randomly. The results showed that expression of nine miRNAs were detected, but the miR-1516 had no detectable expression due to the abundance of the miRNAs in tissues or its time- or tissue-specific expression (Fig. 3). Some miRNAs exhibited ubiquitous but their expression levels varied from stages, for instance miR-370 and miR-1000 were highly expressed in the egg but lower in the moth. The miR-1638 showed relatively high expression level in the egg and moth but weakly in the larva and pupa stages. Similarly, the miR-1268 expressed highly in the egg and larva. And miR-989 was also expressed strongly in the egg and uniformly in other three stages (larva, pupa and moth). In addition, miR-1804, miR-561 and miR-870 revealed development-specific expression patterns; miR-561 was expressed only in the larva, miR-870 was expressed in the larva and pupa; while the miR-1804 was expressed in the egg, larva and moth stages but no detectable expression in pupa. The miR-iab-4as was expressed relatively weak in the larva, pupa and moth but highly in the egg stage.

Analysis of miR-iab-4 family related gene expression has been detailed studied. The miR-iab-4-3p exhibited a

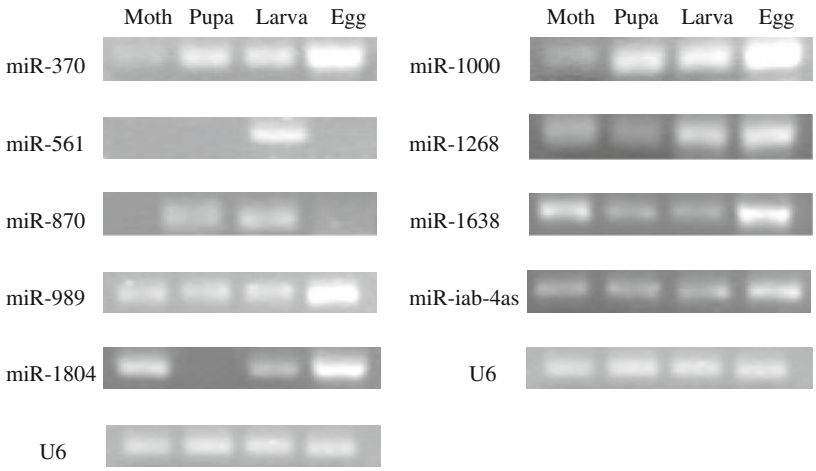

Fig. 3 Expression analysis of the identified miRNAs in the egg, larva, pupae and moth stages of the silkworm. U6 snRNA was used as a positive control

relatively high expression level at the egg and larva stages, implying it may play a crucial role in regulating development-related genes in the silkworm [16]. These stage-biased miRNAs provide insights into the mechanisms in regulation of the developmental stages in the silkworm. Experimental results showed the expression of miR-iab-4as in the egg is consistent with the pattern of miR-iab-4-3p stage, but low in larva. The difference between miR-iab-4-3p and miR-iab-4as may be duo to the functional divergence of the miR-iab-4 family. Previous research indicated silkworm cells are suspended in G2 at DS, and Ser/Thr protein phosphatase, targeted by miR-iab-4-3p, controls the transition of G2 to M, suggesting that miR-iab-4-3p may play a pivotal role in keeping embryonic cells at G2 by suppressing the function of Ser/Thr protein phosphatase [31–33]. The miR-iab-4-5p directly regulated *UBX* protein expression in *Drosophila* and over expressed of miR-iab-4 causes a dominant homeotic transformation of halteres to wings [34]. These results will shed lights on the fine-tuning mechanism of miRNAs in the silkworm.

Prediction of potential targets of novel miRNAs in the silkworm

The potential regulatory targets were identified with less than four mismatches according to the criteria described in the "Materials and methods" section. After removal of repeated candidates, 21 miRNAs target records were produced (Table S2). The target mRNAs can be separated into two groups. One group contains miRNA targets encoding transcription factors. Another group contains miRNA targets encoding a range of different proteins which may play important roles in the aspects of metabolisms, development and signal transduction.

Acknowledgments We gratefully acknowledge the technique assistance of Dr. Y. X. Yu at Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. This research was partly supported by research grants form the National High Tech Development Project of China, the 863 Program (2007AA100504) and the National Basic Research 973 Program (2005CB121004).

References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297
2. Hannon GJ (2002) RNA interference. Nature 418:244–251
3. Avnit-Sagi T, Kantorovich L, Kredo-Russo S, Hornstein E, Walker MD (2009) The promoter of the pri-miR-375 gene directs expression selectively to the endocrine pancreas. PLoS One 4:e5033
4. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854
5. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403:901–906
6. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294:853–858
7. Berezikov E, Cuppen E, Plasterk RH (2006) Approaches to microRNA discovery. Nat Genet 38:S2–S7
8. Griffiths-Jones S (2004) The microRNA registry. Nucleic Acids Res 32:D109–D111
9. Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods Mol Biol 342:129–138
10. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34:D140–D144
11. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158
12. Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol 5:R65
13. Tong CZ, Jin YF, Zhang YZ (2006) Computational prediction of microRNA genes in silkworm genome. J Zhejiang Univ Sci B 7:806–816
14. Liu S, Xia Q, Zhao P, Cheng T, Hong K, Xiang Z (2007) Characterization and expression patterns of let-7 microRNA in the silkworm (*Bombyx mori*). BMC Dev Biol 7:88

15. He PA, Nie Z, Chen J, Lv Z, Sheng Q, Zhou S, Gao X, Kong L, Wu X, Jin Y, Zhang Y (2008) Identification and characteristics of microRNAs from *Bombyx mori*. BMC Genom 9:248
16. Yu X, Zhou Q, Li SC, Luo Q, Cai Y, Lin WC, Chen H, Yang Y, Hu S, Yu J (2008) The silkworm (*Bombyx mori*) microRNAs and their expressions in multiple developmental stages. PLoS One 3:e2997
17. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798
18. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10:1507–1517
19. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ (2005) Real-time quantification of microRNAs by stem–loop RT–PCR. Nucleic Acids Res 33:e179
20. Zhou ZS, Huang SQ, Yang ZM (2008) Bioinformatic identification and expression analysis of new microRNAs from *Medicago truncatula*. Biochem Biophys Res Commun 374:538–542
21. Cao J, Tong C, Wu X, Lv J, Yang Z, Jin Y (2008) Identification of conserved microRNAs in *Bombyx mori* (silkworm) and regulation of fibroin L chain production by microRNAs in heterologous system. Insect Biochem Mol Biol 38:1066–1071
22. Zhang Y, Zhou X, Ge X, Jiang J, Li M, Jia S, Yang X, Kan Y, Miao X, Zhao G, Li F, Huang Y (2009) Insect-specific microRNA involved in the development of the silkworm *Bombyx mori*. PLoS One 4:e4677
23. Zeng Y (2006) Principles of micro-RNA production and maturation. Oncogene 25:6156–6162
24. Guddeti S, Zhang DC, Li AL, Leseberg CH, Kang H, Li XG, Zhai WX, Johns MA, Mao L (2005) Molecular evolution of the rice miR395 gene family. Cell Res 15:631–638
25. Chang DT, Wang CC, Chen JW (2008) Using a kernel density estimation based classifier to predict species-specific microRNA precursors. BMC Bioinform 9(12):S2
26. Xu JH, Li F, Sun QF (2008) Identification of microRNA precursors with support vector machine and string kernel. Genom Proteom Bioinform 6:121–128
27. Zhang B, Pan X, Stellwag EJ (2008) Identification of soybean microRNAs and their targets. Planta 229:161–182
28. Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI (2008) The birth and death of microRNA genes in *Drosophila*. Nat Genet 40:351–355
29. Zhang R, Peng Y, Wang W, Su B (2007) Rapid evolution of an X-linked microRNA cluster in primates. Genome Res 17:612–617
30. Niwa R, Slack FJ (2007) The evolution of animal microRNA function. Curr Opin Genet Dev 17:145–150
31. Iwasaki H, Takahashi M, Niimi T, Yamashita O, Yaginuma T (1997) Cloning of cDNAs encoding *Bombyx* homologues of Cdc2 and Cdc2-related kinase from eggs. Insect Mol Biol 6:131–141
32. Champlin DT, Truman JW (1998) Ecdysteroid control of cell proliferation during optic lobe neurogenesis in the moth *Manduca sexta*. Development 125:269–277
33. Yoshida I, Moto K, Sakurai S, Iwami M (1998) A novel member of the bombyxin gene family: structure and expression of bombyxin G1 gene, an insulin-related peptide gene of the silkmoth *Bombyx mori*. Dev Genes Evol 208:407–410
34. Ronshaugen M, Biemar F, Piel J, Levine M, Lai EC (2005) The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. Genes Dev 19:2947–2952