Insect Molecular Biology (2008) **17**(4), 427–436

In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*)

J. Singh and J. Nagaraju

Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Nacharam, Hyderabad, India

Abstract

MicroRNAs (miRNAs), are endogenous, ~22-nucleotide-long RNA molecules. They bind to the complementary sites on target mRNAs and regulate protein production of the target transcript by unknown mechanisms. Since the discovery of first miRNA in *Caenorhabditis elegans,* different approaches have been pursued for the prediction of miRNAs and their target(s). Because of many difficulties and limitations involved in the experimental identification of spatially and temporally expressed miRNAs, many computational approaches have been successfully employed for prediction of miRNAs and their target(s). In the present study, we demonstrate a genome-wide computational approach to predict miRNAs and their target(s) in the red flour beetle, *Tribolium castaneum*. We have predicted and characterized 45 miRNAs by genome-wide homology search against all the reported miRNAs. These miRNAs were further validated by statistical and phylogenetic analyses. In addition, we have also attempted to predict the putative targets of these miRNAs, by making use of 3′ untranslated regions of mRNAs from *T. castaneum*. These miRNAs and their targets in *T. castaneum* will serve as useful resources for initiating studies on their experimental validation and functional analyses of miRNA-regulated phenotypes in *T. castaneum* through gene knockdown and transgenesis.

Keywords: *Tribolium castaneum*, miRNAs, red flour beetle, *in silico* prediction.

Received 5 October 2007; accepted after revision 23 April 2008. Correspondence: J. Nagaraju, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad 500076, India. Tel.: +91 40 2717 1427; fax: +91 40 2715 5610; e-mail: jnagaraju@cdfd.org.in

Introduction

MicroRNAs (miRNAs) are ~22-nucleotide (nt) noncoding RNA molecules generated from stem-loop hairpin structures called miRNA precursors (premiRNAs) of ~80 nt. These premiRNAs are first transcribed as longer RNAs (Lee *et al.*, 2002) and are then processed by the RNAase III enzyme, Drosha (Lee *et al.*, 2003). Another RNAase III enzyme, Dicer (Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001), cuts these ~80-nt premiRNAs to release ~22-nt mature miRNA.

miRNAs play a regulatory role in many cellular and developmental processes such as cell division (Leaman *et al.*, 2005), cell death (Stark *et al.*, 2003), hormone secretion (Poy *et al.*, 2004) and neural development (Jin *et al.*, 2004). Recent evidence also suggests that miRNAs function as tumour suppressors (Calin *et al.*, 2002) and oncogenes (Calin *et al.*, 2005). Since the first discovery of the miRNAs *lin-4* (Lee *et al.*, 1993) and *let-7* (Reinhart *et al.*, 2000) in *Caenorhabditis elegans*, miRNA abundance in the genomes of various organisms (Lau *et al.*, 2001; Lee & Ambros, 2001) including insects (Lagos-Quintana *et al.*, 2001), plants (Bartel & Bartel, 2003), viruses (Cullen, 2006) and higher vertebrates (Lim *et al.*, 2003) has been reported. miRNAs are generally conserved in closely related species and to some degree in distantly related species as well (Pasquinelli *et al.*, 2000); eg about 10% of miRNAs identified in invertebrates are also conserved in mammals and other higher animals, suggesting cross-species conservation of their regulatory functions (Weber, 2005).

In insects, miRNAs from *Drosophila melanogaster* and *Anopheles gambiae* of order Diptera, *Apis mellifera* of order Hymenoptera and recently from *Bombyx mori* of order Lepidoptera have been predicted and submitted to the miRNA registry miRBase (Griffiths-Jones, 2004). However, no miRNAs have been reported so far from the very diverse and large order Coleoptera, which includes many devastating agricultural pests. With the recent availability of whole genome sequence data, linkage groups, expressed sequence tags (ESTs) and various genetic markers, the red flour beetle, *Tribolium castaneum* has emerged as a genetic model system of Coleoptera (Richards *et al.*, 2008). By making use of these genomic resources, in the present

study we identified and characterized miRNAs, predicted their target(s), and elucidated their phylogenetic relationship with those reported from other insects.

Experimental identification of miRNAs and their targets is difficult and cumbersome. Consequently, several computational methods have been developed and employed for reliable and rapid identification of temporally, spatially and meagrely expressed miRNA genes. Among many approaches that are being used for the prediction of miRNAs, the one based on phylogenetic conservation of sequences across multiple species is reported to provide reliable prediction of functional miRNAs (Stark *et al.*, 2007a,b).

Several computational tools based on comparative genomics have been successfully employed in the identification of miRNAs, eg programs like miRscan (Lim *et al.*, 2003) and miRseeker (Lai *et al.*, 2003). These programs extract conserved intergenic regions between two closely related species and then check for premiRNA secondary structure, which may result in exclusion of many potential miRNAs as mature miRNA sequences are more conserved than their precursor sequences (Lau *et al.*, 2001). The miRFinder program (Bonnet *et al.*, 2004a), however, predicts miRNAs based on the mature miRNA sequences but uses alignment of closely related species and hence is not sensitive enough to identify conserved miRNA homologues from distantly related species. Keeping these shortcomings under consideration, we carried out a whole genome homology search of *T. castaneum* against all the miRNAs reported in the miRNA registry. We intentionally used the phylogenetic conservation approach considering that each predicted miRNA could be compared with its experimentally validated miRNA homologue.

In the present study, because we have considered all of the reported miRNAs of different species for the analysis, we believe that the approach employed here is sensitive enough to predict the conserved miRNA homologues from distantly related species as well. Besides, stringent screening based on statistical parameters ensures relatively few false positives thus suggesting the specificity of the approach. As a result, we identified 45 miRNAs, which were further validated by a randomization test for the robustness of their precursor secondary structures and phylogenetic analyses of their precursor sequences. In addition, we have also predicted 132 putative targets based on partially complementary sites of the predicted miRNAs in 3′UTRs of mRNAs involved in several key functions.

Results and discussion

Prediction of miRNAs

The different steps involved in miRNA prediction are shown in Fig. 1. A BLASTn search (Altschul *et al.*, 1990) of all the known mature miRNAs (miRbase Release: 9.2) against the whole genome sequences of *T. castaneum* resulted in 62 hits. These hits were subsequently scanned for their precursor sequences by taking a sliding window of about 100 nt (moving in increments of approximately 10 nt) from the region ~80 nt upstream of the beginning of the mature miRNA to ~80 nt downstream of the miRNA. The characteristic secondary structures of all of the 62 miRNA precursors were determined by the MFold program (Zuker & Stiegler, 1981), which computes the minimum free energy (MFE) contribution for various possible secondary structures. miRNA precursor structures having an MFE more than –18 kcal/mol or a bulge size more than 7 bp or mature miRNA located on the loop region were excluded. Finally, the screening resulted in 45 potential miRNAs (Table 1) and their premiRNA sequences (see Table S1 of Supplementary Material File S1), which can fold back and make typical hairpin-like secondary structures (see Supplementary Material File S2 for the 45 predicted premiRNA secondary structures).

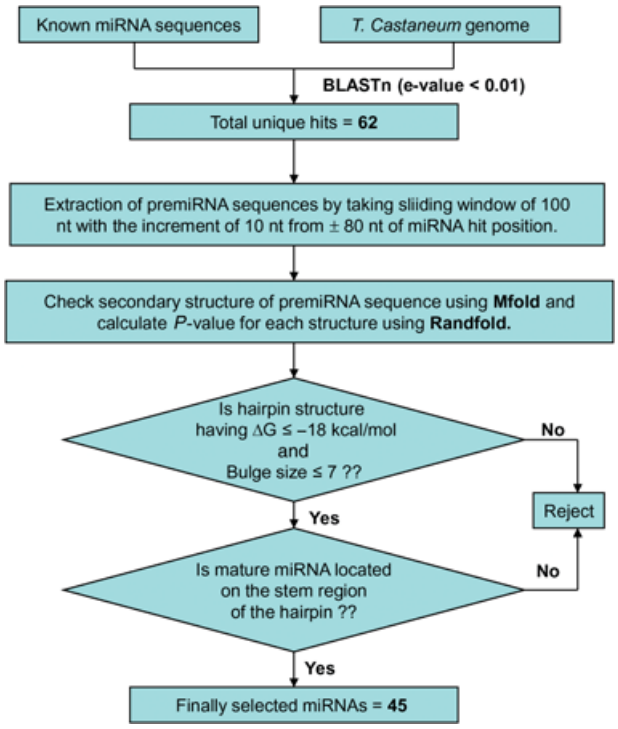


Figure 1. An overview of different steps involved in microRNA prediction (see text for details).

While the data were being analysed, a new release of miRBase (Release 10.0) became available and we extended our prediction to these newly added 489 miRNAs. Out of these, the homology search predicted three new miRNAs but none of them could pass through the secondary structure filter. The new release of miRBase has also introduced a new class of miRNAs called mirtrons (Ruby *et al.*, 2007). Mirtrons are de-branched introns that mimic

Table 1. Details of genomic locations and sequences of 45 predicted microRNAs (miRNAs) from *Tribolium castaneum*

miRNA	Linkage group	Strand	Starting position	Ending P position	Match extent	Mature miRNA sequence
tca-miR-1	LG_5	(–)	11731445	11731424	22/22	UGGAAUGUAAAGAAGUAUGGAG
tca-miR-2	LG_3	(+)	560862	560884	23/23	UAUCACAGCCAGCUUUGAUGAGC
tca-miR-2-1	LG_3	(+)	561230	561252	23/23	UAUCACAGCCAGCUUUGAUGAGC
tca-miR-2-2	LG_3	(+)	561381	561403	23/23	UAUCACAGCCAGCUUUGAUGAGC
tca-let-7	LG_6	(+)	7248807	7248827	21/21	UGAGGUAGUAGGUUGUAUAGU
tca-miR-7	LG_7	(+)	2378814	2378836	23/23	UGGAAGACUAGUGAUUUUGUUGU
tca-miR-8	LG_9	(–)	7581536	7581514	23/23	UAAUACUGUCAGGUAAAGAUGUC
tca-miR-10	LG_2	(+)	2953620	2953641	22/22	ACCCUGUAGAUCCGAAUUUGU
tca-miR-12	LG_3	(–)	20454819	20454797	23/23	UGAGUAUUACAUCAGGUACUGGU
tca-miR-13a	LG_3	(+)	560987	561018	22/22	UAUCACAGCCAUUUUGAUGAGU
tca-miR-13b	LG_3	(+)	561098	561119	22/22	UAUCACAGCCAUUUUGACGAGU
tca-miR-14	LG_3	(+)	3935242	3935262	21/21	UCAGUCUUUUUCUCUCUCCUA
tca-miR-31a	LG_5	(–)	13878846	13878825	22/22	GGCAAGAUGUCGGCAUAGCUGA
tca-miR-33	LG_4	(–)	9369401	9369383	19/19	GUGCAUUGUAGUUGCAUUG
tca-miR-34	LG_7	(–)	14720926	14720905	22/22	UGGCAGUGUUGUUAGCUGGUUG
tca-miR-71	LG_3	(+)	560709	560730	22/22	UGAAAGACAUGGGUAGUGAGAU
tca-miR-87	LG_6	(+)	2916492	2916512	21/21	GUGAGCAAAGUUUCAGGUGUG
tca-miR-92a	LG_8	(+)	1116108	1116127	20/21	AUUGCACUUGUCCCGGCCUG
tca-miR-92b	LG_8	(+)	1115554	1115573	19/20	AUUGCACUAGUCCCGGCCUA
tca-miR-100	LG_6	(+)	7248675	7248696	22/22	AACCCGUAGAUCCGAACUUGUG
tca-miR-124	LG_5	(–)	6938517	169384957	23/23	UAAGGCACGCGGUGAAUGCCAAG
tca-miR-125	LG_6	(+)	7248999	249020	22/22	UCCCUGAGACCCUAACUUGUGA
tca-miR-133	LG_5	(–)	11705004	11704983	19/20	UUGGUCCCCUUCAACCAGCUGU
tca-miR-137	LG_7	(+)	5152298	5152319	22/22	UUAUUGCUUGAGAAUACACGUA
tca-miR-184	LG_7	(–)	9858702	9858681	22/22	UGGACGGAGAACUGAUAAGGGC
tca-miR-190	LG_7	(–)	1058064	1058040	25/25	AGAUAUGUUUGAUAUUCUUGGUUGUU
tca-miR-210	LG_4	(–)	8182078	8182058	21/21	UUGUGCGUGUGACAGCGGCUA
tca-miR-219	LG_3	(+)	6722720	6722742	23/23	UGAUUGUCCAAACGCAAUUCUUG
tca-miR-263b	LG_3	(+)	5921190	5921209	19/20	UUGGCACUGGAAGAAUUCAC
tca-miR-275	LG_9	(–)	9307503	9307481	23/23	UCAGGUACCUGAAGUAGCGCGCG
tca-miR-276	LG_3	(–)	13466831	13466810	22/22	UAGGAACUUCAUACCGUGCUCU
tca-miR-276*	LG_3	(–)	13466867	13466846	22/22	GCGAGGUAUAGAGUUCCUAC
tca-miR-277	LG_7	(–)	14725559	14725537	23/23	UAAAUGCACUAUCUGGUACGACA
tca-miR-279	LG_7	(–)	8332745	8332724	22/22	UGACUAGAUCCACACUCAUUAA
tca-miR-281	LG_7	(–)	1799849	1799829	21/21	UGUCAUGGAGUUGCUCUCUUU
tca-miR-281*	LG_7	(–)	1799884	1799863	22/22	AAGAGAGCUAUCCGUCGACAGU
tca-miR-282	LG_7	(+)	10121107	10121130	24/24	UAGCCUCUCCUAGGCUUUGUCUGU
tca-miR-283	LG_3	(–)	204592189	20459199	20/20	AAAUAUCAGCUGGUAAUUCU
tca-miR-305	LG_9	(–)	307403	9307381	23/23	AUUGUACUUCAUCAGGUGCUCUG
tca-miR-307	LG_8	(+)	14748960	14748979	20/20	UCACAACCUCCUUGAGUGAG
tca-miR-317	LG_7	(–)	14733806	14733782	25/25	UGAACACAGCUGGUGGUAUCUCAGU
tca-miR-927	LG_9	(+)	10250933	10250954	22/22	UUUAGAAUUCCUACGCUUUACC
tca-miR-932	LG_5	(–)	17260944	17260923	22/22	UCAAUUCCGUAGUGCAUUGCAG
tca-miR-iab-4-3p	LG_2	(–)	2465584	2465561	24/24	CGGUAUACCUUCAGUAUACGUAAC
tca-miR-iab-4-5p	LG_2	(–)	2465618	2465597	22/22	ACGUAUACUGAAUGUAUCCUGA

the structural features of premiRNAs to enter the miRNA-processing pathway without the Drosha-mediated cleavage. So far, 15 mirtrons from *D. melanogaster* and three mirtrons from *C. elegans* have been identified, but we failed to predict any of the conserved mirtrons in *T. castaneum*. One reason for this may be the paucity of premiRNA-sized introns in *T. castaneum*, unlike in flies and nematodes.

Forty-five miRNAs predicted in the present study included three identical copies of *tca-mir-2* and two each of *tca-miR-13* and *tca-miR-92* with one or two mismatches, and less predominant forms of miR-281 and miR-276, represented as miR-281* and miR-276*, respectively. Of these 45 predicted miRNAs, 14 have homologues only in insects, seven in both insects and worms, whereas 17 have been reported in both lower and higher organisms (Table S2 of Supplementary Material File S1). In addition, our predicted miRNAs included miR-927 and miR-932, which are reported to be present only in *A. mellifera* (Weaver *et al.*, 2007). Similarly, our miRNA list also included two miRNAs, miR-137 and miR-190, previously reported only in vertebrates and not in any other insect species except *A. mellifera*. Most of our predicted miRNA sequences have compositions beginning with uridine at the 5′ end as is the case with miRNAs reported from other organisms (Lagos-Quintana *et al.*, 2002; Reinhart & Bartel, 2002).

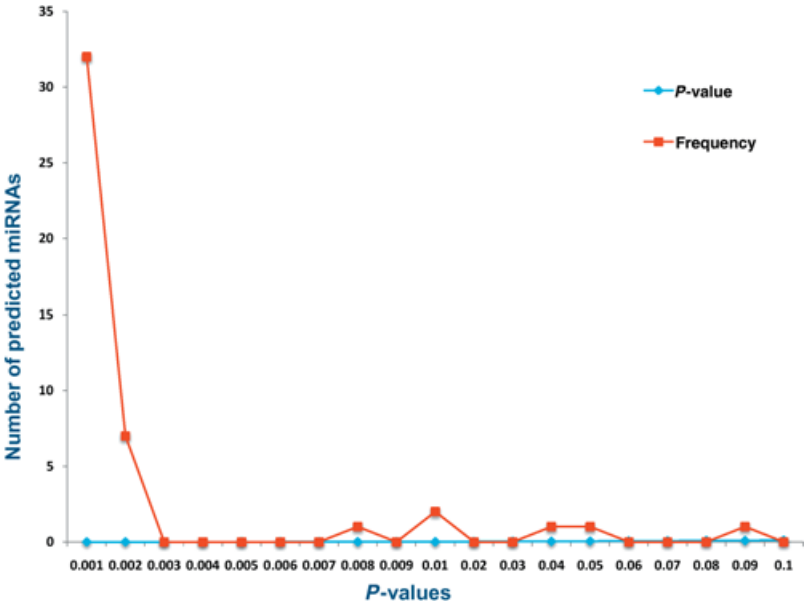


Figure 2. Distribution of *P*-values for the 45 predicted microRNA precursor (premiRNA) sequences. premiRNAs having *P*-values > 0.01 were tca-mir-33 (0.08), tca-mir-317 (0.05), mir-932(0.04) and tca-mir-124 (0.02).

Randomization test

To assess the robustness of the 45 predicted premiRNA secondary structures, we carried out a randomization test using the RandFold program (Bonnet *et al*., 2004b), in which the predicted MFE of folding was compared with the values obtained for structures inferred from randomly shuffling the original sequence. We used dinucleotide shuffling of bases as MFE can be calculated by considering the minimum energy values obtained by complementary base pairs, which decrease by the stacking energy of successive base pairs or increase by the destabilizing energy associated with noncomplementary bases (Zuker & Stiegler, 1981).

It has been found that the probability (*P*-value) of randomized sequences having MFE values less or equal to the original value is less than 0.01 for more than 90% of the known premiRNAs (Bonnet *et al*., 2004b). The results obtained in the present study are consistent with this observation as 41 of the 45 predicted premiRNAs had *P*-values < 0.01 (see Fig. 2). These results suggest that MFE values of the predicted premiRNA structures are significantly different from that of random sequences. In order to further validate the results, we carried out phylogenetic analyses of the predicted miRNA precursor sequences.

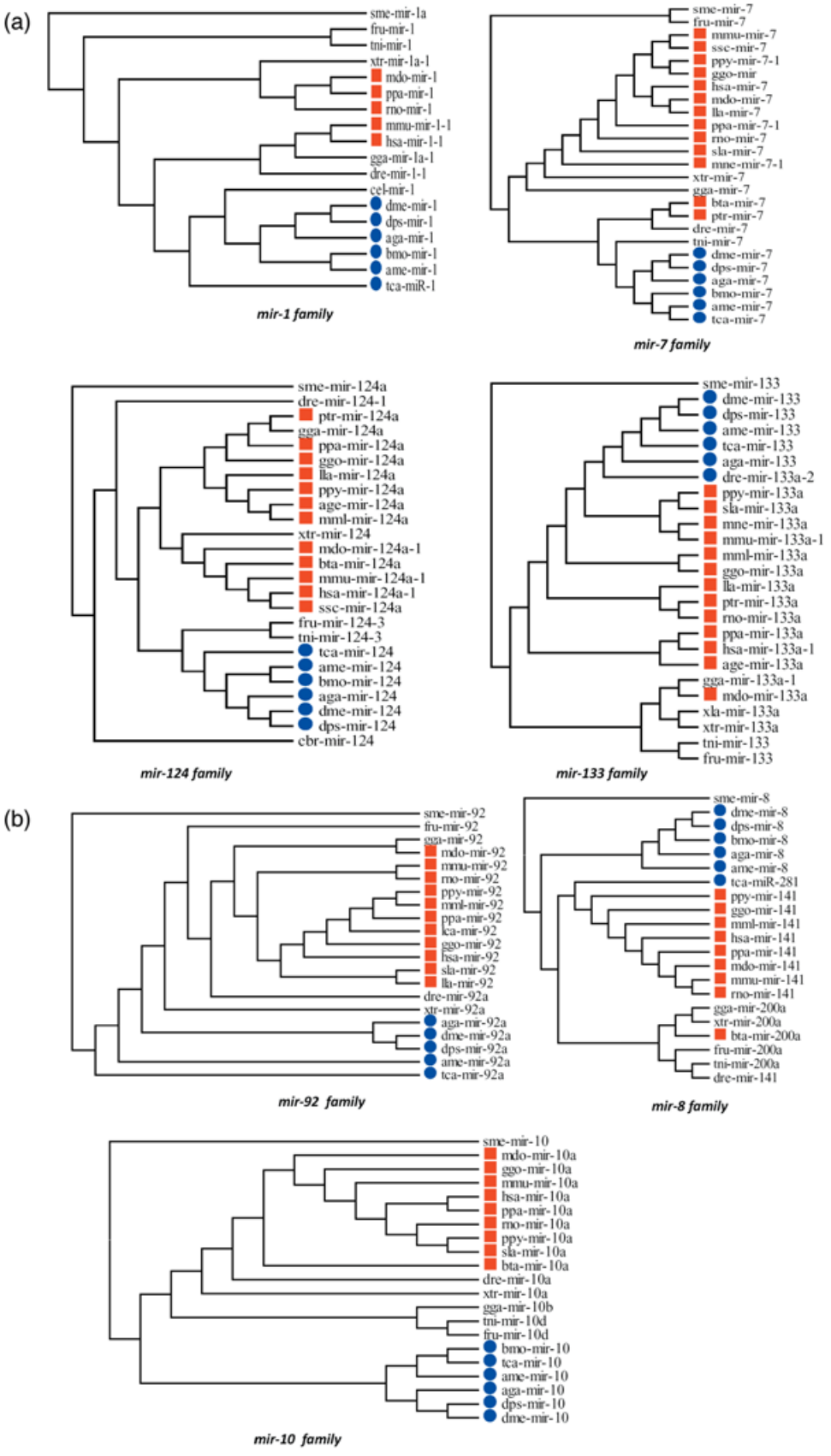
Phylogenetic analyses

Forty-five predicted miRNAs could be classified into different families based on their premiRNA sequences, which included 12 known RNA families of Rfam, version 8.0 (Griffiths-Jones *et al*., 2005), (see the distribution in Table S3 of Supplementary Material File S1). From these 12 families, seven large families were selected and Kimura-2 parameter-based neighbour-joining trees were constructed using MEGA-4.0 (Tamura *et al*., 2007) with *Schmidtea mediterranea*, a freshwater planaria, as a root to illustrate the evolutionary relationships among the members of the family (see Fig. 3). The phylogenetic trees of the seven families clearly show grouping of *T. castaneum* miRNAs with those of other insects.

Target prediction

In animals, miRNAs primarily target the 3′UTRs of target mRNA(s) (Brennecke *et al*., 2003; Lin *et al*., 2003). There are very few recent reports demonstrating that target mRNAs are also repressed by miRNA-binding sites on 5′UTR or coding regions as efficiently as in the 3′UTR (Rhoades *et al*., 2002; Lytle *et al*., 2007). Hence, we confined our target search only to the 3′UTRs of mRNAs. The gene models generated as part of the Tribolium genome project include mostly coding sequences (CDS) and not

Figure 3. Phylogenetic analysis of seven large families of microRNA precursors (premiRNAs). Trees were constructed by the Kimura-2 parameter-based neighbour-joining method using MEGA-4 and other parameters including bootstrap values are taken as default. Each family consists of a premiRNA sequence of different species including those of *Triboliumcastaneum*. Insect and mammalian species are indicated as blue circles and red squares, respectively. Abbreviations: sme, *Schmidtea mediterranea*; cel, *Caenorhabditis elegans*; cbr, *Caenorhabditis briggsae*; dps, *Drosophila pseudoobscura*; dme, *Drosophila melanogaster*; bmo, *Bombyx mori*; aga, *Anopheles gambiae*; ame, *Apis mellifera*; xtr, *Xenopus tropicalis*; tni, *Tetraodon nigroviridis*, fru, *Fugu rubripes*; dre, *Danio rerio*; gga, *Gallus gallus*; has, *Homo sapiens*; mmu, *Mus musculus*; rno, *Rattus norvegicus*; ggo, *Gorilla gorilla*; mdo, *Monodelphis domestica*; ppa, *Pan paniscus*; ptr, *Pan troglodytes*; ppy, *Pongo pygmaeus*; lla, *Lagothrix lagotricha*; bta, *Bos taurus*; mml, *Macaca mulatta*.



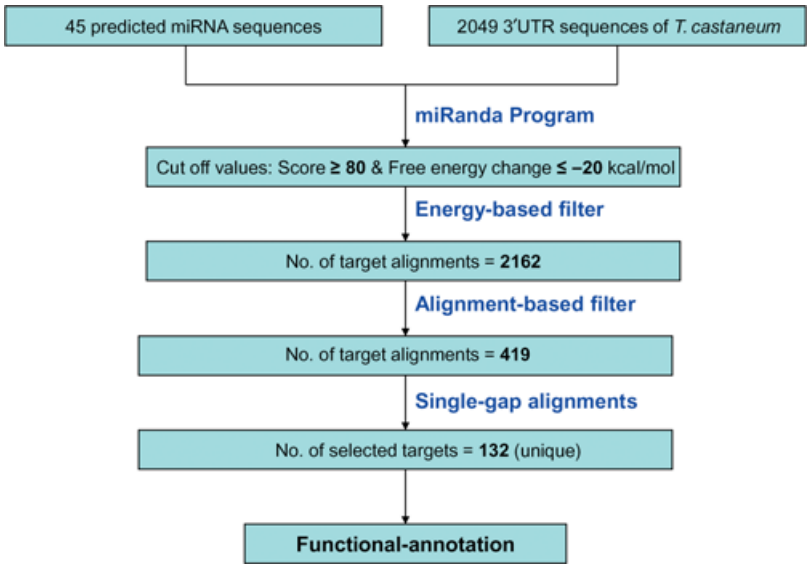


Figure 4. An overview of different steps involved in microRNA target(s) prediction in *Tribolium castaneum*.

full-length mRNAs. Therefore, using the 9013 UniGene data set of *T. castaneum* http://www.ncbi.nlm.nih.gov/UniGene/, we extracted the 3′UTR sequences (see Experimental procedures for more details).

In plants, miRNAs bind their targets by complete or nearly complete complementarity (Jones-Rhoades & Bartel, 2004). By contrast, animal miRNAs are partially complementary to their target mRNAs. This has rendered computational approaches for target identification based merely on reverse complementary searches quite challenging in animals (Vella *et al.*, 2004; Rajewsky, 2006). We employed the miRanda program (Enright *et al.*, 2003), which utilizes thermodynamics and dynamic-programming alignments, along with statistical parameters, for target prediction in *T. castaneum*. We also considered stringent filters for screening targets to minimize the background matches, thus ensuring the least false positives. The different steps involved in target prediction are shown in Fig. 4.

Primary screening of 45 predicted miRNAs target site(s) in 3′UTR sequences of *T. castaneum* yielded 2162 miRNA::mRNA alignments. These alignments were then scanned for the strong *seed region* complementarity, that is consecutive Watson–Crick matches on positions 2 to 7 at the 5′ end of a miRNA (Krek *et al.*, 2005; Lewis *et al.*, 2005; Lim *et al.*, 2005). G : U wobbles are less common in the 5′ end of the miRNA::mRNA duplex but are tolerable at various positions in the *seed region* (Didiano & Hobert, 2006) and hence we allowed one G : U pairing in the *seed region*. In order to obtain functionally valid complementary 5′ region of miRNA on the target mRNA, mismatches were preferred to gaps and hence not more than one gap was allowed. Finally, the screening resulted in the prediction of 132 potential miRNA target genes in *T. castaneum.* Out of these 132 target genes, 64 showed functional similarities with *Drosophila* genes (summarized in Table 2). The target sites of these 64 genes from *T. castaneum* were then compared with known targets of *Drosophila* reported in miRBase but none of the target sites recovered known miRNA binding sites of *Drosophila*.

The functions of the predicted miRNA targets included catalytic activity, nucleic acid/protein binding and various transcription regulatory factors. These diverse targets provide evidence for the multiple levels of gene regulation by miRNAs in very broad cellular and molecular processes.

Target multiplicity and cooperativity

Many of the predicted miRNAs revealed target multiplicity; *tca-miR-34* was found to have a maximum of 15 targets followed by *tca-miR-133* and *tca-miR*-124 with 11 and 10 targets, respectively (for distribution of targets per miRNA, refer to Table S4 of Supplementary Material File S1). Some of the target genes had more than one miRNA hits suggesting cooperative regulation of their expression eg the *AY800247* gene was targeted by as many as three miRNAs, miR-317, miR-71 and miR-92 (see Table S5 of Supplementary Material File S1 for more details). In animals, cooperative binding of one or several distinct miRNAs on a single target gene is reported to be important for the functionality of miRNA-mediated gene regulation (Enright *et al.*, 2003; John *et al.*, 2004). Because of the unfinished annotation of the *Tribolium* genome, the predicted miRNA targets reported in the present study constitute only a subset of the total targets in *Tribolium*.

Conclusions

The predicted miRNAs and their targets reported in the present study will serve as potential resources to initiate

Table 2. Details of microRNA targets and their putative functions in *Tribolium castaneum*. Molecular functions of these target genes were obtained from FlyBase based on Gene Ontology

Target accession no.	miRNA(s) hits	Target site(s)	miRNA co-ordinates on 3′UTR	Annotation ID	Protein annotation
XM_963434	miR-263b, miR-12	1, 1	(202–222), (233–255)	CG16973-PC	Protein serine/threonine kinase activity
XM_970019	miR-133	1	(11–32)	CG7595-PB	Actin-dependent ATPase activity
XM_964617	miR-92	1	(72–91)	CG4733-PA	Protein phosphatase type 2A regulator activity
XM_962335	miR-12	1	(165–188)	CG1059-PA	Protein transmembrane transporter activity
XM_965475	miR-133	1	(684–706)	CG31367-PA	General RNA polymerase II transcription factor activity
XM_961297	miR-282	1	(192–215)	CG3488-PA	Lipase activity
XM_965360	miR-184	1	(70–92)	CG16903-PA	Cyclin-dependent protein kinase regulator activity
XM_963643	miR-125	1	(71–93)	CG12004-PA	Molecular function not known
XM_969563	miR-87	1	(91–112)	CG13340-PA	Leucyl aminopeptidase activity
XM_965982	miR-1	1	(192–214)	CG6131-PA	Structural constituent of chitin-based cuticle
XM_967014	miR-305	1	(186–208)	CG9413-PA	Amino acid transmembrane transporter activity
XM_966578	miR-92	1	(232–252)	CG14307-PB	RNA polymerase II transcription factor activity
XM_963784	miR-282	1	(12–36)	CG6384-PA	Microtubule binding
XM_970812	miR-276*	1	(140–160)	CG11982-PA	Zinc ion binding
XM_970506	miR-276*	1	(74–94)	CG3373-PA	Strictosidine synthase activity
XM_967023	miR-137	1	(390–411)	CG8625-PA	ATP-dependent helicase activity
XM_966419	miR-87	1	(21–42)	CG31638	Molecular function not known
XM_968390	miR-100, miR-99a	1, 1	(57–79), (57–79)	CG2182-PA	Molecular function not known
XM_967804	miR-13a, miR-2	1, 1	(98–119), (98–119)	CG4758-PA	Protein transporter activity
XM_964554	miR-275	1	(192–215)	CG10954-PA	Structural constituent of cytoskeleton
XM_963463	miR-34	1	(12–34)	CG10338-PA	Molecular function not known
XM_963974	miR-133	1	(18–40)	CG5930-PA	Phosphopantetheine binding
XM_962144	miR-92	1	(53–72)	CG2471-PA	Protein binding
XM_968744	miR-219	1	(681–703)	CG5059-PA	Molecular function not known
XM_962587	miR-276, miR-34	1, 1	(433–453), (381–403)	CG6772-PA	Protein kinase activity
XM_966451	miR-133	1	(140–160)	CG31523-PA	Acyltransferase activity
XM_964999	miR-307	1	(155–174)	CG4778-PA	Chitin binding
XM_961262	let-7	1	(4–24)	CG1873-PA	Translation elongation factor activity
XM_962851	miR-276*	1	(398–416)	CG31094-PB	Low-density lipoprotein receptor activity
XM_967411	miR-124	1	(1–23)	CG13926-PA	Molecular function not known
XM_966343	miR-14	1	(58–78)	CG5629-PA	Molecular function not known
XM_961254	miR-282	1	(29–52)	CG4143-PA	Methyl-CpG binding
XM_968221	miR-99a	1	(9–31)	CG14998-PE	Molecular function not known
XM_966736	miR-1	1	(93–115)	CG5576-PA	Antibacterial humoral response, protein binding
XM_965116	miR-33	1	(824–842)	CG17342-PA	Receptor signaling protein serine/threonine kinase activity
XM_969476	miR-2	1	(193–214)	CG10435-PA	Molecular function not known
XM_962838	miR-34	1	(30–52)	CG5730-PC	Calcium-dependent phospholipid binding
XM_963782	miR-210	1	(52–71)	CG7896-PA	Receptor activity
XM_963895	miR-124	1	(11–34)	CG9779-PA	Protein transport
XM_968991	miR-210	1	(51–70)	CG6210-PB	Molecular function not known
XM_968738	miR-281*	1	(215–236)	CG10231-PA	Cyclic nucleotide metabolic process
XM_970872	miR-277	1	(7–28)	CG3625-PB	Molecular function not known
XM_964037	miR-34	1	(4–26)	CG30421-PA	Ubiquitin thiolesterase activity
XM_965239	miR-13a	1	(103–123)	CG10574-PA	Protein phosphatase inhibitor activity
XM_968247	miR-34	1	(251–271)	CG13431-PA	Acetylglucosaminyltransferase activity
XM_962256	miR-281	1	(23–42)	CG10175-PC	Carboxylesterase activity
XM_967341	miR-263b	1	(129–148)	CG6831-PA	Actin binding
XM_967702	miR-317	1	(117–140)	CG6944-PA	Structural constituent of cytoskeleton
XM_964688	miR-317	1	(85–108)	CG4562-PA	ATPase activity, coupled to transmembrane movement of substances
XM_961718	miR-279	1	(80–102)	CG6724-PA	Molecular function not known
XM_969210	miR-282	1	(104–128)	CG14234-PA	Molecular function not known
XM_962040	miR-31a	1	(181–203)	CG6416-PF	Protein binding
XM_970453	miR-124	1	(57–80)	CG5362-PA	L-malate dehydrogenase activity
XM_969655	miR-99a	1	(231–252)	CG9038-PA	Protein modification process
XM_965417	miR-279	1	(130–151)	CG17723-PA	Zinc ion transmembrane transporter activity
XM_964448	miR-71	1	(399–421)	CG7957-PA	RNA polymerase II transcription mediator activity
XM_965276	miR-305	1	(100–122)	CG4059-PB	Ligand-dependent nuclear receptor activity
XM_962794	miR-13a, miR-13b	1, 1	(287–307), (286–307)	CG33144-PA	Zinc ion binding
XM_969710	miR-133	1	(271–291)	CG5783-PA	N-acetyltransferase activity
XM_968625	miR-184	1	(44–65)	CG31195-PA	Molecular function not known
XM_966579	miR-34	1	(1–21)	CG1172-PA	Protein modification process
XM_967685	miR-133	1	(470–491)	CG1120-PA	Molecular function not known
XM_968123	miR-2	1	(308–330)	CG15793-PA	MAP kinase kinase activity
XM_961264	miR-282	1	(67–91)	CG6092-PA	Cytidylate kinase activity

their experimental validation. The outcome of this experimental validation should give valuable information to carry out functional analysis of miRNA-regulated phenotypes relevant to basic and applied biology of insects in general and coleopterans in particular.

Experimental procedures

Prediction of miRNAs

A total of 5395 known miRNA sequences, including 473 star (*) miRNAs were downloaded from miRBase (Release 10.0), http://microrna.sanger.ac.uk/sequences/. The Human Genome Sequencing Center (HGSC) has sequenced the *T. castaneum* genome with sevenfold coverage using a whole genome shotgun approach. The genome sequence was downloaded from the ftp site of HGSC at Baylor College of Medicine ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Tcastaneum/Tcas2.0/linearScaffolds/.

Homology search

A BLASTn search of all the 5395 miRNA sequences in the whole genome sequence of *T. castaneum* was carried out with the e-value < 0.01 and the default parameters were used, including low complexity filter. The two criteria used for screening the BLAST results were: (1) more than 80% identity between each potential *T. castaneum* miRNA and the corresponding miRNA in the reference set (known miRNA homologue); (2) the length difference between each potential *T. castaneum* miRNA and the corresponding miRNA in the reference set is not more than three bases.

Naming convention

Naming of the miRNAs predicted in the present study is in accordance with miRbase (Griffiths-Jones *et al.*, 2006). The mature sequences are designated 'miR', and the precursor hairpins are labelled as 'mir' with the prefix '*tca*' for *T. castaneum*. In the cases where distinct precursor sequences have identical miRNAs, they were named as *tca-mir-2-1*, *tca-mir-2-2* and so on. Closely related miRNAs with one or two mismatches were denoted by suffixing letters eg *tca-mir-13a* and *tca-mir-13b*. If two miRNAs had the same precursor, an asterisk (*) was used to denote the less predominant form eg *tca-miR-181** is the less predominant form of *tca-miR-181*, having the same precursor *tca-mir-181*.

Secondary structure validation

premiRNA sequences were extracted using a sliding window of about 100 nt in size (moving in increments of approximately 10 nt) from the region ~80 nt upstream of the beginning of the mature miRNA to ~80 nt downstream of the miRNA. Extracted miRNA precursor sequences were then submitted to Mfold (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi) for checking of the fold-back secondary structure. The three criteria used for selecting premiRNA structures were: (1) structure should have free energy change (ΔG) less than or equal to –18kcal/mole; (2) the bulge size (un-pairing) should not be more than 7 bp; (3) mature miRNA hit should be on the stem region of hairpin structure.

By applying these three criteria to scan the precursor secondary structure, we could tap 92% of the total known miRNAs reported in miRBase (release 10).

$$\text{Sensitivity} = \frac{\text{Number of predicted true positives}}{\text{Total number of true positives}} = \frac{4689}{5071} = 92\%$$

Randomization test

Robustness of each of the predicted premiRNA structures was assessed by a randomization test using the Randfold program downloaded from http://bioinformatics.psb.ugent.be/supplementary_data/erbon/nov2003/. The Vienna RNA package required for Randfold was downloaded from http://www.tbi.univie.ac.at/RNA/. The parameters assigned for calculating the *P*-value of the predicted premiRNA secondary structures were: (1) number of randomization equal to 1000; (2) the type of shuffling was dinucleotide.

Extraction of 3′UTR sequences

An overview of steps involved in the 3′UTR extraction is shown in Fig. S1 of Supplementary File S1. A total of 9013 Unigene sequences of *T. castaneum* was downloaded from NCBI (ftp://ftp.ncbi.nih.gov/respository/UniGene) and conceptual translation of each of these sequences was performed using Genscan (Burge & Karlin, 1997). Genscan results were then parsed and 2418 genes with poly(A) information were selected. For each of the selected genes, the region from the end of the stop codon to the beginning of poly(A) was assigned as 3′UTR and was extracted using in-house perl script. Finally, 2049 3′UTR sequences having lengths over 20 bases were selected.

Target prediction using miRanda program

miRanda inputs were 45 predicted miRNAs and 2049 extracted 3′UTR sequences of *T. castaneum*. The parameters assigned for miRanda were: Smith–Waterman hybridization default alignment score (Smith & Waterman, 1981) greater than or equal to 80, MFE of miRNA::mRNA duplex (Wuchty *et al.*, 1999) less than or equal to –20kcal/mol and the other parameters were kept as default.

Scanning of hits

Screening of miRNA-target alignments was based on the following three empirical rules (position count starting from 5′ end of the miRNA): (1) no mismatches at positions 2 to 7 (seed region); (2) not more than one G : U pairing allowed in the seed region; and (3) not more than one gap allowed in the alignment.

Functional annotation

Out of 132 predicted target genes, molecular functions of 64 genes showing functional similarities with Drosophila were obtained from FlyBase (Crosby *et al.*, 2007) based on Gene Ontology (Ashburner *et al.*, 2000). For the other 68 target genes, functions were assigned using BLASTn (see Table S5 of Supplementary Material File S1).

Acknowledgements

We thank and acknowledge the Human Genome Sequencing Center at Baylor College of Medicine for providing the sequence data of *T. castaneum.* JN acknowledges financial support from the Department of Biotechnology, Government of India under Centre of Excellence for

Genetics and Genomics of Silkmoths. JS is the recipient of a fellowship from the Department of Biotechnology, Government of India.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.

Bartel, B. and Bartel, D.P. (2003) MicroRNAs: at the root of plant development? *Plant Physiol* **132**: 709–717.

Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. (2004a) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc Natl Acad Sci. USA* **101**: 11511–11516.

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

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

Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78–94.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E. *et al*. (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* **99**: 15524–15529.

Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E. *et al*. (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* **353**: 1793–1801.

Crosby, M.A., Goodman, J.L., Strelets, V.B., Zhang, P. and Gelbart, W.M. (2007) FlyBase: genomes by the dozen. *Nucleic Acids Res* **35**: D486–D491.

Cullen, B.R. (2006) Viruses and microRNAs. *Nat Genet* **38** Suppl: S25–30.

Didiano, D. and Hobert, O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* **13**: 849–851.

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

Griffiths-Jones, S. (2004) The microRNA registry. *Nucleic Acids Res* **32**: D109–111.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R. and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121–124.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140–144.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I. *et al*. (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.

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

Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T.A. *et al*. (2004) Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci* **7**: 113–117.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol* **2**: e363.

Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* **14**: 787–799

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

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J. *et al*. (2005) Combinatorial microRNA target predictions. *Nat Genet* **37**: 495–500.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735–739.

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

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

Leaman, D., Chen, P.Y., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U. *et al*. (2005) Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. *Cell* **121**: 1097–1108.

Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862–864.

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

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

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

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

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

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J. *et al*. (2005) Microarray analyses shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769–773.

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

Lytle, J.R., Yario, T.A. and Steitz, J.A. (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc Natl Acad Sci USA* **104**: 9667–9672.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B. *et al*. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86–89.

Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E. *et al*. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**: 226–230.

Rajewsky, N. (2006) microNA target predictions in animals. *Nat Genet* **38** Suppl: S8–13.

Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E. *et al*. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–906.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell* **110**: 513–520.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W. *et al*. (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.

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

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

Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) Identification of *Drosophila* MicroRNA targets. *PLoS Biol* **1**: E60.

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

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W. *et al*. (2007b) Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**: 219–232.

Tamura, K., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. and Slack, F.J. (2004) The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev* **18**: 132–137.

Weaver, D.B., Anzola, J.M., Evans, J.D., Reid, J.G., Reese, J.T., Childs, K.L. *et al*. (2007) Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol* **8**: R97.

Weber, M.J. (2005) New human and mouse microRNA genes found by homology search. *Febs J* **272**: 59–73.

Wuchty, S., Fontana, W., Hofacker, I.L. and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**: 145–165.

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

Supplementary material

The following supplementary material is available for this article:

Supplementary Material File S1 includes five tables and one figure:

Table S1. Details of genomic location, sequence, free energy change and *P*-value of all the 45 predicted microRNA precursors from *Tribolium castaneum*

Table S2. Categorization of predicted microRNAs (miRNAs) into different groups based on known miRNAs reported in miRBase (Release 10.0)

Table S3. List of 12 known Rfam families and their member(s) from predicted microRNA precursors

Table S4. List of number of predicted targets per microRNA in *Tribolium castaneum*

Table S5. Details of miRNA targets and their putative functions in *Tribolium castaneum*. Molecular functions of these target genes were obtained from a BLASTn search

Figure S1. Overview of steps followed for extracting 3′UTR sequences in *Tribolium castaneum*.

Supplementary Material File S2 contains foldback secondary structures of all the 45 predicted premiRNAs.
This material is available as part of the online article from:
http://www.blackwell-synergy.com
doi:10.1111/j.1365-2583.2008.00816.x
(This link will take you to the article abstract).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.