

# Functional Shifts in Insect microRNA Evolution

Antonio Marco, Jerome H. L. Hui, Matthew Ronshaugen\*, and Sam Griffiths-Jones\*

Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom

\*Corresponding author: E-mail: matthew.ronshaugen@manchester.ac.uk; sam.griffiths-jones@manchester.ac.uk.

Accepted: 27 August 2010

## Abstract

MicroRNAs (miRNAs) are short endogenous RNA molecules that regulate gene expression at the posttranscriptional level and have been shown to play critical roles during animal development. The identification and comparison of miRNAs in metazoan species are therefore paramount for our understanding of the evolution of body plans. We have characterized 203 miRNAs from the red flour beetle *Tribolium castaneum* by deep sequencing of small RNA libraries. We can conclude, from a single study, that the *Tribolium* miRNA set is at least 15% larger than that in the model insect *Drosophila melanogaster* (despite tens of high-throughput sequencing experiments in the latter). The rate of birth and death of miRNAs is high in insects. Only one-third of the *Tribolium* miRNA sequences are conserved in *D. melanogaster*, and at least 18 *Tribolium* miRNAs are conserved in vertebrates but lost in *Drosophila*. More than one-fifth of miRNAs that are conserved between *Tribolium* and *Drosophila* exhibit changes in the transcription, genomic organization, and processing patterns that lead to predicted functional shifts. For example, 13% of conserved miRNAs exhibit seed shifting, and we describe arm-switching events in 11% of orthologous pairs. These shifts fundamentally change the predicted targets and therefore function of orthologous miRNAs. In general, *Tribolium* miRNAs are more representative of the insect ancestor than *Drosophila* miRNAs and are more conserved in vertebrates.

**Key words:** miRNAs, *Tribolium*, deep sequencing, arm switching, embryonic development.

## Introduction

The discovery of microRNAs (miRNAs) has brought the importance of posttranscriptional regulation of gene expression to the forefront of biology. miRNAs are short endogenous RNA sequences (~22 nt) that mediate translational repression of target messenger RNAs (mRNAs) (Bartel 2004). During the last decade, miRNAs have been found to play major roles in virtually every biological process: from development and signaling to viral infections (reviewed in Kloosterman and Plasterk 2006). Moreover, miRNAs are ubiquitous in multicellular animals and plants (Griffiths-Jones et al. 2008) but almost absent in single-celled organisms, suggesting a pivotal role in the emergence of multicellularity. In animals, miRNAs regulate many aspects of development: from body patterning (Yekta et al. 2004; Ronshaugen et al. 2005) to cell differentiation (e.g., Makeyev et al. 2007), so changes in their functional features are likely associated with the evolution of body plans as well as phenotypic variation within related species (reviewed in Niwa and Slack 2007). These findings have driven an increasing interest in understanding the evolution of miRNA function.

Typically, animal miRNAs are processed from longer transcripts in the nucleus by an RNase III enzyme called Drosha,

producing a hairpin structure (called the miRNA precursor or pre-miRNA; Lee et al. 2002). The pre-miRNA is transported to the cytoplasm, where it is further cleaved by Dicer, another RNase III enzyme, giving rise to an RNA duplex approximately 22 nt long (Lee et al. 2003). One of the strands of this duplex (the so-called mature miRNA) associates with the RNA-induced silencing complex and binds to complementary sequences in the 3' untranslated region of target mRNAs, leading to repression of translation or transcript degradation (reviewed in Bartel 2009). In high-throughput sequencing experiments, the other strand (often called the star sequence or miR\*) is often detected at lower levels and is assumed to be degraded. However, in many cases, both arms of the pre-miRNA hairpin produce functional miRNA products (Glazov et al. 2008; Okamura et al. 2008).

Intergenic miRNAs tend to be clustered in the genome, probably because they are processed from the same primary transcript (Altuvia et al. 2005; Saini et al. 2007). A substantial proportion of miRNAs (30–50% in animals) are embedded into introns of protein-coding genes, suggesting cotranscription of the miRNA and its host gene (Rodriguez et al. 2004; Baskerville and Bartel 2005). However, instances of independent transcriptional regulation of intronic

miRNAs have also been reported (e.g., Tang and Maxwell 2008; Bell et al. 2010). A number of key questions regarding miRNA function and evolution can only be answered by comparing the features of miRNAs in multiple species. For example, which miRNAs are conserved in arthropods and which are lineage specific? Are clusters of miRNAs conserved throughout evolution? Do clustered miRNAs have common features? Is the choice of arm from which to process the mature miRNA conserved in animals?

*Drosophila melanogaster* is the prototype for genetic analysis of arthropods (Ashburner et al. 2004). Much of our knowledge of animal miRNA biology comes from this species (see Behura 2007 and references therein), as well as from other non-insect model species (such as *Caenorhabditis elegans* and *Mus musculus*). The comparison of genomic and functional properties of miRNAs—such as conservation of sequence, arm choice, and cotranscription—between *Drosophila* and mammals is vital to our understanding of the role of miRNAs in animal evolution. However, *Drosophila* is only one representative of the diversity of insects. It is crucial to understand the conservation of miRNAs in a broader set of insects in order to understand miRNA function and evolution in animals. Here we describe insights gained from sequencing the miRNA complement of the red flour beetle *Tribolium castaneum*.

*Tribolium* has a relatively short generation time, is easy to rear in the laboratory, and is amenable to sophisticated genetic manipulations (Brown et al. 2003). A sequenced and assembled red flour beetle genome is also available (Richards et al. 2008). Moreover, studies based on protein sequence conservation indicate that dipterans (including *Drosophila* and mosquitoes) are fast evolving and that *Tribolium* is a more appropriate model to compare gene evolution between invertebrates and vertebrates (Savard et al. 2006; Richards et al. 2008). *Tribolium castaneum* has genetic and developmental features similar to the last common ancestor of all arthropods; for example, short-germ embryonic development and segmentation (Davis and Patel 2002) and the genomic structure of the Hox gene cluster (Shippy et al. 2008). In contrast, the long-germ mode of embryogenesis in *Drosophila* is a derived state (Davis and Patel 2002). Although a small number of miRNAs have been detected in *Tribolium* using computational homology approaches (Luo et al. 2008; Singh and Nagaraju 2008), experimental confirmation of only one (*iab-4*) has been provided (Shippy et al. 2008). Here we use deep sequencing to obtain an extensive collection of transcribed short RNAs in *Tribolium* and reconstruct the miRNA catalog of the flour beetle. Our comprehensive set of *Tribolium* miRNAs allows us to analyze patterns of expression and processing and thereby study in detail the functional shifts during miRNA evolution in insects and other animals.

## Materials and Methods

### Short RNA Libraries Generation and Sequencing

Wild type beetles were cultured at 28 °C. RNA was extracted separately from 0- to 5-day embryos and adults with the miRVana miRNA isolation kit (Ambion). Molecules shorter than 40 nucleotides were selected with the flashPAGE fractionator (Ambion) and purified with the flashPAGE reaction cleanup kit (Ambion). Two different libraries containing different embryonic stages were constructed with different barcodes with the SOLiD Small RNA Expression Kit (Ambion). Size-selected small RNAs were ligated to the sequencing adaptors as described by the manufacturer. Reverse transcription was then carried out, followed by RNaseH digestion to make the cDNA libraries. In order to meet the sample quantity for SOLiD sequencing, the cDNA libraries were then further amplified using supplied primer sets containing different barcodes by 15 or 18 cycles of polymerase chain reaction. The final products ranging from ~105 to 150 bp were purified. SOLiD sequencing was performed at the Center for Genomic Research at the University of Liverpool.

### Detection of Transcribed miRNAs

Reads for the two SOLiD sequencing runs were 50 nucleotides long. Thus, we expect that putative miRNA mature sequences (~22 nt) detected in a SOLiD run must contain fragments of the linker used during the sequencing process. Thus, we first trimmed all 3' ends back to 26 nt sequences. All reads were first mapped to annotated *Tribolium* ribosomal RNAs (rRNAs) (<http://www.arb-silva.de/>) and predicted transfer RNAs (tRNAs) (tRNAscan-SE; Lowe and Eddy 1997) using Bowtie (Langmead et al. 2009), allowing one mismatch between the read and the sequence. Reads mapped to these sequences were discarded. After this filtering step, the remaining reads were mapped to the *Tribolium* reference genome version 3.0 ([ftp://ftp.ncbi.nih.gov/genomes/Tribolium\\_castaneum/](ftp://ftp.ncbi.nih.gov/genomes/Tribolium_castaneum/)), again with Bowtie, allowing one mismatch and mapping to up to five positions. The terminal 3' nucleotide was removed from the unmapped reads and the reads mapped again, repeating the process sequentially to a minimum of 19 nucleotides in length. Similar sequential trimming approaches have been recently used, although in a different context (Cloonan et al. 2009).

Overlapping reads were grouped and flanking regions (−50 to +100 and −100 to +50) retrieved from the *T. castaneum* genome assembly (version 3.0). These fragments were scanned for hairpins with RNAfold (Hofacker et al. 1994). In order to detect potential miRNAs, we applied several filters to the resulting hairpins: 1) hairpins should contain at least 10 reads mapped to the putative arms; 2) the hairpin folding energy must be −15 kcal/mol or lower; 3) at least 50% of the nucleotides of one arm of the hairpin must

pair with bases from the other arm; 4) the 5' end of the mature miRNA should be accurately processed, that is, 80% of the mapped reads from at least one arm should have the same 5' end; 5) both arms of the hairpin should have associated reads and the most abundant read from each arm should be paired, overlapping by more than 70% in the hairpin structure. For potential miRNAs detected in contigs not associated to any of the ten assembled *Tribolium* chromosomes, we additionally filtered out those miRNAs not supported by reads that map uniquely and exactly. To minimize the effect of cross-mapping during the detection of paralogous miRNAs, we discarded putative miRNAs only supported by reads mapping to multiple positions that are not also annotated as miRNAs.

### miRNA Families

miRNAs were assigned to known families using two independent approaches. First, all sequences were systematically searched against the miRBase database (release 15; Griffiths-Jones et al. 2008) using basic alignment search tool (Blast) (word size = 4; match reward = 5; mismatch penalty = -4) (Altschul et al. 1997). Second, flanking regions of each miRNA (500 nt centered on the miRNA) were used to search against the Rfam 10.0 library of covariance models (Griffiths-Jones et al. 2005) using Infernal 1.0 (Nawrocki et al. 2009) and Rfam-curated thresholds. At this step, we additionally discarded three miRNAs that hit the Rfam tRNA model but were not filtered out at the read preprocessing step. We detected potential homologs sequences with MapMi (Guerra-Assunção and Enright 2010), setting the minimum score threshold to 20 and up to three mismatches (see Results for the list of genomes scanned). Briefly, MapMi maps mature sequences to a genome, and then it folds the region looking for hairpins. We additionally included homology relationships described in miRBase. Specific examples described in this work were aligned (using ClustalW; Thompson et al. 1994) and manually inspected (using RALEE; Griffiths-Jones 2005). *Tribolium* miRNAs were grouped into families by all-against-all Blast searches of their precursor hairpins, filtering out hits with E values above 0.001, and then assignments were hand curated. The genome assembly versions used were as

follows: *D. melanogaster* (release 5.0), *Anopheles gambiae* (2.1), *Apis mellifera* (4.0), *Bombyx mori* (2.0), *Acyrtosiphon pisum* (1.0), *Daphnia pulex* (1.0), *Capitella teleta* (1.0), *Schistosoma mansoni* (3.1), *C. elegans* (7.1), *Branchiostoma floridae* (2.0), *Gallus gallus* (2.1), and *Homo sapiens* (37.1).

### Relative Arm Usage

To quantify the relative amount of mature miRNA products from each arm of the same miRNA hairpin precursor, we define here the “relative arm usage” measure. This quantity is defined as follows:

$$\log_2(N_{5'}/N_{3'}),$$

where  $N_{5'}$  is the number of reads mapped to the 5' arm of the hairpin precursor and  $N_{3'}$  the number of reads from the 3' arm. Relative arm usage units are bits. Positive values indicate a bias toward 5' arm usage and negative values a bias toward 3'. Zero means that mature sequences are produced at equal levels from both arms.

## Results

### A Comprehensive Catalog of *Tribolium* miRNAs

In order to detect processed miRNAs in *Tribolium*, we constructed small RNA libraries from two different populations. The first population was composed of both male and female adults, including fecund females. We additionally sequenced small RNAs from a population of early embryonic stages (0–5 days) to further explore the differences in miRNA expression during early development. Both libraries were sequenced using the ABI SOLiD platform (see Materials and methods), yielding a total of ~120 million sequence reads. Around 10% of the mapped reads were removed as potential rRNA or tRNA contaminants (table 1). For the remaining sequences from the first small RNA library, we successfully mapped about 12% of the reads to the reference genome. This proportion is comparable to a recent SOLiD-based miRNA sequencing study in the silkworm *B. mori* (Cai et al. 2010). However, only 3% of the reads generated from the early development library mapped to the genome. The proportion of mapped reads derived from tRNA and rRNA sequences is similar in both experiments. In adults,

**Table 1**  
Reads Sequenced in Small RNA Libraries of *Tribolium*

	Late Development and Adults	Early Development
Total number of reads	67,070,132	52,620,004
Filtered out (tRNAs/rRNAs)	953,849 (1.4%)	142,175 (0.3%)
Mapped to the genome	7,993,536 (11.9%)	1,733,263 (3.3%)
Reads in miRNAs	1,753,289 (21.9%)	49,606 (2.9%)
Homologs of known miRNAs	1,648,508	45,388
Newly detected miRNAs	104,781	4,218



we associated 22% of the mapped reads to known or predicted miRNAs. Less than 3% of the reads mapped from the early development library were ascribed to miRNAs (table 1). This suggests that the total quantity of small RNAs is significantly lower in the early embryo. Other factors that may contribute to the low mapping rates include the strict parameters used for mapping. Moreover, sequencing errors from SOLiD technology are likely to lead to unmapped reads rather than single base errors. Only mapped reads were subsequently used to detect putative miRNAs using a pipeline developed in-house (see Materials and Methods).

Collectively, our RNA libraries support the existence of 203 miRNAs (supplementary file 1, Supplementary Material online). We provided transcriptional evidence for 51 predicted *Tribolium* miRNAs cataloged in miRBase (version 15; Griffiths-Jones et al. 2008). Moreover, we detected 33 miRNAs not yet described for *Tribolium* but homologs to described miRNAs in other species (see below). The remaining 119 are newly described miRNAs with no obvious homology to known miRNAs. It is important to note that our annotation strategy requires that reads support mature miRNAs from both arms of the precursor (so-called miR and miR\* sequences). This is the most conservative high-throughput strategy so far described, in common with a recent annotation of mouse miRNAs (Chiang et al. 2010). We find the presence of reads supporting the miR\* sequence to be the most useful single criterion for miRNA annotation. As sequencing depth increases, the likelihood of detecting low abundance miR\* sequences also increases. We suggest that requiring support for both arms provides an optimal balance of sensitivity and specificity at the coverage we have seen. However, we expect that a small number of bona fide but low abundance *Tribolium* miRNAs may be missed by our strategy, where the miR\* sequence falls below the detectable limit imposed by the sequencing coverage.

To further explore the performance of our strategy for miRNA detection, we used our pipeline to reanalyze a third-party data set of *D. melanogaster* small RNA reads (Ruby, Stark, et al. 2007). We repeated the procedure used for our *Tribolium* data sets, and we detected 118 potential miRNAs, covering approximately 70% of the miRBase catalog for *D. melanogaster*. We characterized almost all miRNAs newly described in the original paper (Ruby, Stark, et al. 2007), with the exception of 18 sequences that did not pass our strict filtering procedure (which is more appropriate for our larger sequencing data set). Our strategy additionally detected three mirtrons described elsewhere (Ruby, Jan, and Bartel 2007) and mir-2498: a miRNA that escaped

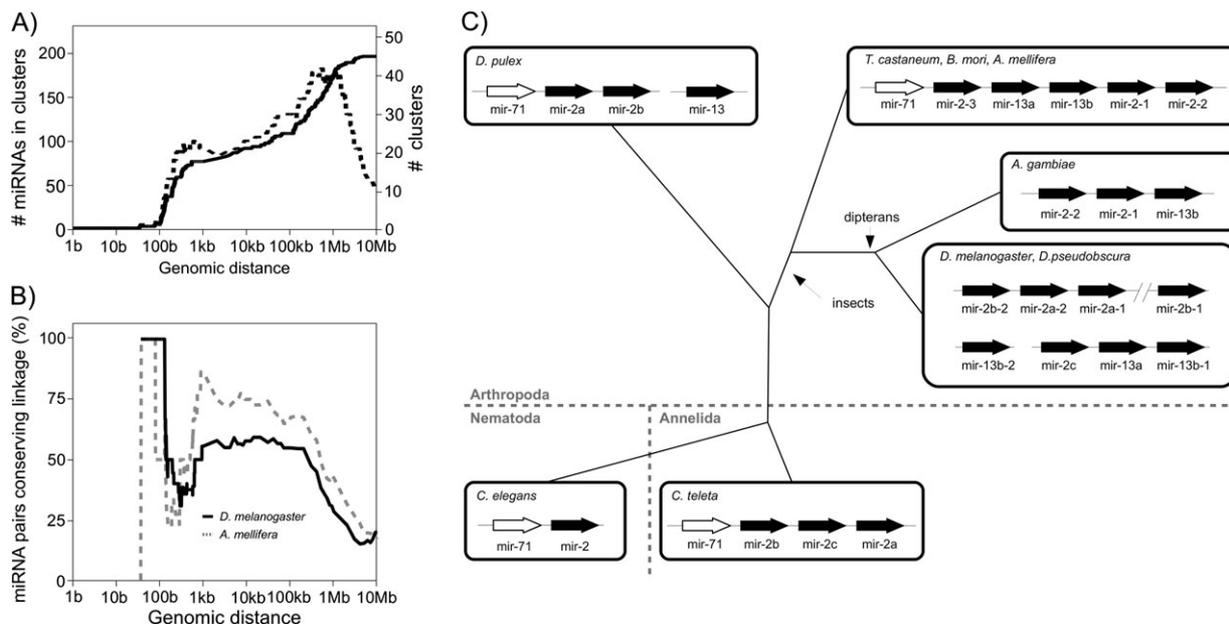
the original analysis and has been recently detected based on massive sequencing experiments (Berezikov et al. 2010).

### Expanding the Set of Insect miRNAs

In order to determine which of the miRNAs described here are also present in other insects and other invertebrate species, we performed systematic searches with MapMi (Guerra-Assunção and Enright 2010; see Materials and Methods for details) to identify homologs in the complete genome assemblies of *A. gambiae*, *A. pisum*, *A. mellifera*, *B. mori*, *C. teleta*, *C. elegans*, *D. pulex*, *D. melanogaster*, and *S. mansoni*. No identifiable insect homolog could be found for 62 *Tribolium* miRNAs. These miRNAs include not only singleton sequences but also *Tribolium*-specific expanded families, like the mir-3806/mir-3808/mir-3811 cluster in the sex chromosome (fig. 1D). All 51 previously predicted miRNAs in *Tribolium* were present in other arthropods. This was expected as they were identified by comparative sequence analyses. Within this set of 51, we observed that only mir-71 was not detected in dipterans. Indeed, a *Drosophila* mir-71 homolog has not been reported in miRBase, whereas homologs in multiple invertebrates are known. In total, 44 of the 141 (~31%) newly detected *Tribolium* miRNAs are present in other arthropods but not present in dipterans (fig. 1A); half of these (23) are conserved in other invertebrates. The data suggest that *Tribolium* may conserve more common miRNAs with other insects than with *Drosophila* and may therefore represent a more ancestral model of conserved miRNA function.

To further explore whether the *Tribolium* miRNA catalog is more representative of insects than that of *Drosophila*, we compared the percentage of known *Drosophila* (miRBase, version 15) and *Tribolium* (this study) miRNAs with detectable homologs in other insects using MapMi. In figure 1B, we observe that *Tribolium* miRNAs are less likely to be conserved in *Drosophila* than in other insects. On average, 40% of *Tribolium* miRNAs have homologs in *Apis* or *Bombyx*, whereas ~35% of the *Drosophila* set have detectable homologs in these two species. This indicates the presence of multiple miRNAs broadly conserved in insects that have been lost in *Drosophila*. For instance, mir-2796, previously thought to be silkworm specific (Liu et al. 2010), is absent from *Drosophila* but detected in our *Tribolium* sequences, and putative homologs can be found in *Anopheles* and *Apis* (fig. 1C). Using MapMi, we analyzed three chordate genomes (*H. sapiens*, *G. gallus*, and *B. floridae*) for putative homologs of both *Tribolium* and *Drosophila* miRNAs. We identify 18 miRNAs from our *Tribolium* set that are conserved in all three chordates but not in *Drosophila*.

sequences (RALEE; Griffiths-Jones 2005), colored according to structural conservation. The five sequences with the highest number of reads for each arm in our *Tribolium* experiments are shown on top of the alignment. (D) Genomic location of *Tribolium*-specific mir-3806/mir-3808/mir3811 family members and folded precursor hairpins. Mature products are indicated in uppercase.



**FIG. 2.**—Genomic clustering of insect miRNAs. (A) Number of miRNAs in clusters (solid line) and number of clusters (dashed line) for different genomic distances in *Tribolium*. (B) Number of miRNAs in clusters in *Tribolium* that conserve the clustering in *Drosophila* (solid line) or in *Apis mellifera* (gray dashed line). (C) Conservation of mir-71/mir-2/mir-13 clusters in multiple invertebrate species.

**Genomic Clusters of Insect miRNAs**

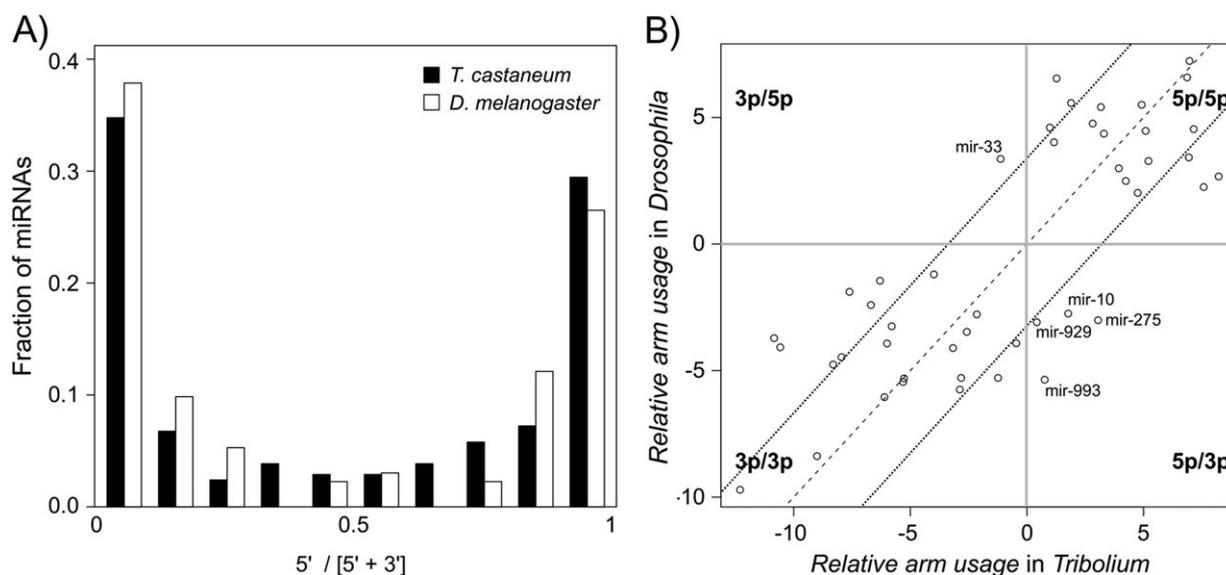
Around 46% (93) of *Tribolium* miRNAs overlap predicted protein-coding genes in the BeetleBase annotation (Wang et al. 2007), whereas the remainder (110) are in intergenic regions. This proportion is comparable to that found in mammals (Rodriguez et al. 2004). Almost all nonintergenic miRNAs are located in introns, and two-thirds of the intronic miRNAs are on the coding strand, indicating that their expression is likely to be regulated by the host gene transcription regulatory sequences. Only 8 miRNAs are inside predicted coding sequences; a closer inspection of the host genes revealed that these exons code for predicted proteins that are not present in any other species. Moreover, these 8 miRNAs are evolutionarily conserved and expressed in other species. We believe this is a consequence of dubious protein-coding gene annotation. We did not exclude protein-coding regions from our analysis; yet, none of our miRNA set overlaps confidently annotated protein-coding exons (i.e., exons encoding for proteins conserved in other species).

miRNA sequences are often clustered in the genome (Griffiths-Jones et al. 2008), probably because they are processed from a single transcript (Lee et al. 2002; Saini et al. 2008). We observed that almost 40% of *Tribolium* miRNAs are within 1 kb of another miRNA (fig. 2A). With an inter-miRNA distance of 10 kb, the proportion of clustered miRNAs is 47%, and 50% of miRNAs are linked at a distance of 27 kb or less (fig. 2A). The mean number of miRNAs per cluster is approximately three for 1 kb clusters and almost four

for larger groups (10–30 kb). From these results, we interpret that approximately half of all miRNAs in *Tribolium* are expressed from polycistronic transcripts, which vary in length up to 20 kb.

To investigate whether this clustering is evolutionarily conserved in insects, we calculated the proportion of miRNAs clustered in *Tribolium* that are also clustered in either *A. mellifera* or *D. melanogaster*. Although the birth and death rates in insect miRNAs are high, those elements conserved between two species tend to maintain their clustering features (fig. 2B). We observed that *Tribolium* clusters are better conserved in *A. mellifera* than in *D. melanogaster*. At 1-kb clustering distance, 86% of conserved miRNA pairs clustered in *Tribolium* maintain their linkage in *Apis*, whereas only 56% are linked in *Drosophila*. At 10 kb, these proportions are 75% and 58% for *Apis* and *Drosophila*, respectively (fig. 2B), and 21 miRNAs, organized in 8 clusters, conserve their linkage between *Tribolium* and *Apis* (supplementary file 2, Supplementary Material online). By comparing clusters of miRNAs in multiple species and accounting for their orientation, we can identify potential transcriptional units. For example, the mir-100/let-7/mir-125 cluster is known to be highly conserved in animals and is likely transcribed as a single unit. Other examples of clusters predicted to be single transcripts are as follows: mir-277/mir-34, mir-275/mir-305, mir-12/mir-283 (supplementary file 2, Supplementary Material online), and the cluster formed by mir-71 and multiple mir-2 family members (fig. 2C). The latter cluster is a good example of high conservation of organization

Downloaded from gbe.oxfordjournals.org at Institute for Development Policy and management, University of Manchester on October 6, 2010



**FIG. 3.**—Arm usage bias in insect miRNAs. (A) Proportion of reads detected in the 5' arm of miRNAs with respect to the total number of reads from the miRNA in *Tribolium*. (B) Comparison of the “relative arm usage” (see Materials and methods for definition) between *Tribolium* and *Drosophila*. miRNAs within the 3'/5' and 5'/3' quadrants show a switch in their arm usage. The dashed line indicates the theoretical expectation for conserved arm usage between the two species. Dotted lines limit the boundaries of the dashed line to less than 10-fold differences in arm usage.

within insects and other invertebrates, whereas in *Drosophila*, the cluster is fragmented and lacks mir-71 (fig. 2C). In summary, these results show a high conservation of clustering in insect miRNAs and suggest a higher level of cluster fragmentation in the *Drosophila* lineage.

We identified five loci in the *Tribolium* genome where mature miRNAs are processed from both sense and antisense strands: *iab-4*, *mir-307*, *mir-1233*, *mir-3867*, and *mir-3817*. With the exception of *iab-4* (Tyler et al. 2008) and *mir-307* (Stark et al. 2007), for which antisense products were also reported for *Drosophila*, no bidirectional transcription is conserved in two different insects (Stark et al. 2007; Tyler et al. 2008; Jagadeeswaran et al. 2010; Liu et al. 2010).

### Functional Shifts in Insect miRNAs

A dominant mature miRNA can be processed from the 5' or 3' arm of the hairpin precursor. Across all *Drosophila* miRNAs, there is a slight bias toward 5' arm usage, whereas in *Tribolium*, equal numbers of miRNAs prefer the 5' and 3' arms (fig. 3A). The *Drosophila* and *Tribolium* distributions are not, however, statistically different ( $P = 0.41$ ; Kolmogorov–Smirnov test). Nevertheless, these minor differences in arm usage indicate that some miRNAs may have switched during evolution the arm from which the dominant mature miRNA is produced. In figure 3B, we plot the relative arm usage for *Tribolium* and *Drosophila* (see Materials and methods). Deviations from zero of this measure indicate that there is a bias toward 5' (positive values) or 3' (negative values) arm usage of a given miRNA. The data clearly show that

five miRNAs have a switch in arm preferences during insect evolution. For example, the 5' arm of mir-33 produces the dominant product in *Drosophila*, whereas the 3' arm dominates in *Tribolium*. In silkworm, the dominant arm in mir-33 is 3' (Jagadeeswaran et al. 2010), suggesting that the *Drosophila* arm usage is a derived character. Switches are also observed for mir-10, mir-993, mir-929, and mir-275. Although only five miRNAs showed a complete switch in their arm usage, it is striking that 20 miRNAs (~44% of *Tribolium* and *Drosophila* 1:1 ortholog pairs) exhibit an arm usage bias 10 times greater in one species with respect to the other (fig. 3B). This shows that significant changes in the proportions of mature sequences produced from 5' and 3' arms are common in insect miRNA evolution and by extension probably in other animals. Mature miRNA sequences produced from 5' and 3' arms of the same precursor hairpin are not similar. Significant shifts in arm usage are therefore predicted to change the target profile and therefore function of a given miRNA.

The previous analysis accounts for only 1:1 orthologous miRNA pairs between *Drosophila* and *Tribolium*. We also investigated eight homologs groups with one-to-many paralog associations, accounting for 20 *Tribolium* miRNAs. We may expect that the existence of multiple copies of the same miRNA after gene duplication may facilitate the arm-switching process (de Wit et al. 2009). Surprisingly, after discarding possible cross-mapped reads, we only observed two potential arm-switching events in paralogous families: one in mir-9a and another in mir-87.

We examined the arm usage of clustered miRNAs at different clustering distances. The data show that clustered

**Table 2**

Pairs of Clustered miRNAs Producing the Dominant miRNA from the Same Arm

Clustered miRNAs	Distance (kb)	Linked miRNA Pairs		Proportion	P value <sup>b</sup>
		miR from the Same Arm <sup>a</sup>	All		
All pairs	<1	74	119	0.62	0.010
	<5	170	310	0.55	0.040
	<10	227	428	0.53	0.065
Same family pairs	<1	29	48	0.60	0.065
	<5	59	115	0.51	0.197
	<10	62	125	0.50	0.317
Unrelated pairs	<1	45	71	0.63	0.022
	<5	111	195	0.57	0.026
	<10	165	303	0.54	0.036

<sup>a</sup> Pairs of clustered miRNAs in which the most abundant mature sequence is processed from the same arm.<sup>b</sup> P values for deviations from nonassociation between clustering and production of miRNAs from the same arm, calculated as described in Sokal and Rohlf (1995, p. 813).

miRNAs tend to have the same dominant arm (see table 2). A variation of a permutation test (Sokal and Rohlf 1995, p. 813) shows that this result is statistically significant at lower clustering distances (table 2). As described above, paralogs tend to produce functional miRNAs from the same arm, so we further corrected for this effect by removing any pair of miRNAs from the same family (see Materials and methods). Notably, the statistical association between clustered miRNAs and same arm usage is stronger in this case for bigger clusters (table 2). The data suggest that common motifs in the primary transcript may concurrently affect the arm choice of multiple miRNAs in the cluster.

The 5' end of the mature sequence is known to be relatively well defined. Changes in the 5' end lead to a phenomenon called seed shifting (Wheeler et al. 2009), which is likely to significantly alter the targets of the mature miRNA. By comparing the mature products of orthologous miRNAs between *Tribolium* and *Drosophila*, we characterized a total of 6 seed-shifting events out of 46 total miRNA orthologous pairs (mir-283, mir-263a, mir-137, mir-282, mir-33, and mir-10). Two of these events were detected in miRNAs that also exhibited arm switches (mir-33 and mir-10). We conclude that seed shifting is as frequent as arm switching during evolution. Together, the two phenomena affect one-fifth of all miRNAs conserved between *Drosophila* and *Tribolium*.

### miRNAs in Early Development

Comparing the relative abundance of miRNAs in both RNA libraries highlights miRNAs involved in early development. In table 3, we show miRNAs that are 10 times more abundant in early embryos than in the adult library and comprise at least 0.1% of the reads mapped to miRNAs. The data implicate 14 miRNAs predicted to be involved in early development in *Tribolium*, including three miRNAs of the same family: mir-309a, mir-309b, and mir-309c, all probably processed from the same transcript. The sole mir-309 homolog

in *Drosophila* is only expressed during the first 2–4 h of development (Aravin et al. 2003), suggesting a conserved role for the mir-309 family during early development in insects. mir-124 accounts for more than 12% of the reads in the early development library. This miRNA is highly conserved in animals and functions in neural system formation both in vertebrates and in invertebrates (Cheng et al. 2009; Clark et al. 2010).

### Discussion

We describe a catalog of 203 miRNAs from the first small RNA deep-sequencing experiments in *T. castaneum*. Tens of high-throughput sequencing experiments have been performed in *Drosophila* giving a total number of 171 miRNAs (miRBase, version 15). Our annotation strategy in *Tribolium*

**Table 3**

miRNAs Highly Expressed in Early Development

miRNA	Reads in Early Embryos <sup>a</sup>	Reads in Adults <sup>a</sup>	Fold Increase <sup>b</sup>
mir-1233	112 (0.22%)	13 (0.00%)	302.7
mir-309b	292 (0.58%)	42 (0.00%)	244.3
mir-9d	116 (0.23%)	17 (0.00%)	239.8
mir-3830	60 (0.12%)	12 (0.00%)	175.7
mir-309c	204 (0.4%)	57 (0.00%)	125.8
mir-309a	1,142 (2.26%)	713 (0.04%)	56.3
mir-3889	322 (0.64%)	308 (0.02%)	36.7
mir-92b	660 (1.3%)	1,525 (0.09%)	15.2
mir-2944a	88 (0.17%)	208 (0.01%)	14.9
mir-3892	60 (0.12%)	154 (0.01%)	13.7
mir-3840	61 (0.12%)	185 (0.01%)	11.6
mir-993	662 (1.31%)	2,057 (0.12%)	11.3
mir-124	6,191 (12.23%)	19,345 (1.09%)	11.2
mir-3845	147 (0.29%)	487 (0.03%)	10.6

<sup>a</sup> Percentages refer to the proportion of reads in a given miRNA with respect to the total number of reads mapped to any miRNA.<sup>b</sup> Fold increase is the ratio between the percentage of reads in early embryos and the percentage of reads in adults.

is intentionally conservative, but we nonetheless conclude that the *Tribolium* miRNA complement is at least 15% greater than that of *Drosophila*. Around 68% of the *Tribolium* miRNAs have detectable homologs in other arthropods, although the proportion of homologs between two insect species is generally below 40% (fig. 1B), suggesting a relatively high rate of turnover during miRNA evolution in insects. We also characterize 47 miRNAs from the *Tribolium* catalog that are present in at least one other arthropod, but not in other invertebrates. This arthropod-specific set includes miRNAs that are expressed during early development in *Drosophila* such as mir-11, mir-309, mir-14, mir-305, and mir-275 (Aravin et al. 2003). *Tribolium* mir-309 paralogs are also present in early developmental stages (table 3). *iab-4* is also arthropod specific, and it functions to modulate Hox gene activity during development (Ronshaugen et al. 2005). Two of the arthropod-specific miRNAs newly detected in this work were also overrepresented in early embryos (mir-3840 and mir-3830 in table 3).

Despite some existing controversy, the net gain of miRNAs in the drosophilid lineage is currently estimated between 0.3 and 1.0 gain per My (Lu et al. 2008; Berezikov et al. 2010). In our *Tribolium* data set, we detect 62 miRNAs not present in any other studied species. Assuming an approximate divergence time of 350 My for holometabolous insects (Wiegmann et al. 2009), the net gain of miRNAs along the *Tribolium* branch is roughly 0.18 per My. The primary sources of error on this number are due to our conservative annotation approach (causing the rate of gain to be underestimated) and missed homologs in other species (leading to an overestimate). Nonetheless, our data support a higher overall net rate of miRNA emergence in the *Drosophila* lineage than other insects.

Approximately half of *Tribolium* miRNAs are clustered in the genome (fig. 2A). We show that this clustering is evolutionarily conserved in insects (fig. 2B). The linkage between miRNAs is more conserved between *Tribolium* and *A. mellifera* than between *Tribolium* and *Drosophila*, suggesting some rearrangement in *Drosophila* clusters. An illustrative example is the mir-71/mir-2/mir-13 cluster (mir-2 and mir-13 are themselves paralogs). In invertebrates, this cluster is composed of mir-71 and one or more mir-2/mir-13 sequences. In insects, the cluster is highly conserved, with mir-71 and five mir-2/mir-13 elements in tandem. However, in dipterans, mir-71 has been lost, and in *Drosophila*, the mir-2 family is fragmented into four loci (fig. 2C). We detected sense and antisense transcription in 5 miRNA loci, but only 2 conserve this bidirectional transcription in *Drosophila* (mir-307 and *iab-4*). Indeed, no other bidirectional miRNA in any insect conserves this feature in another species. Bidirectional transcription of miRNAs is therefore not a common stable feature.

Shifts in the processing pattern of miRNA precursors lead to changes in their mature sequences. These changes alter

the predicted targeting preferences and therefore function of a miRNA. In total, we have shown that one in five miRNAs conserved between *Tribolium* and *Drosophila* have undergone functional shifts. Around 13% of conserved miRNAs between *Drosophila* and *Tribolium* exhibit seed shifting, and we describe arm-switching events in 11% of the orthologous pairs. Arm switching has been previously overlooked but is an important source of evolutionary novelty. Additionally, more than 40% of miRNA loci exhibit 10-fold differences in the proportions of mature sequences processed from 5' and 3' arms. In a significant number of cases, it may therefore be misleading to transfer annotation between orthologous miRNAs based exclusively on conservation. It should be noted that the miRNAs exhibiting shifts between *Drosophila* and *Tribolium* have been conserved over ~350 My and are highly expressed. We can therefore confidently assume that these sequences are functional. Current models of miRNA evolution stress the importance of changes in target sites, whereas miRNA function remains highly conserved (Chen and Rajewsky 2007). However, the relatively high proportion of functional shifts described here also underscores the importance of changes at the miRNA level during the evolution of gene regulatory networks.

*Tribolium castaneum* has many developmental features conserved from the last common insect ancestor (Tautz et al. 1994; Marques-Souza et al. 2008; Shippey et al. 2008), and it is emerging as an alternate and complementary model for insect biology (Brown et al. 2003; Richards et al. 2008; Peel 2009). Our results clearly suggest that *Tribolium* is a good model to study insect miRNA function. First, *Tribolium* miRNAs are more likely to be conserved in other insects than *Drosophila* miRNAs. In fact, there are at least 18 miRNAs shared with chordates that can be studied in *Tribolium* but not in *Drosophila*. Second, clustering patterns of miRNAs are better conserved between *Tribolium* and honeybee than between *Tribolium* and *Drosophila*. Further investigation of the *Tribolium* miRNA complement will increase our knowledge of the evolution of posttranscriptional regulation in animals and, ultimately, help us to understand the origin of extant body plans.

## Supplementary Material

Supplementary files 1 and 2 are available at *Genome Biology and Evolution* Online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/G011346/1), the University of Manchester (fellowships to S.G.J. and M.R.), and the Wellcome Trust VIP scheme. We thank the Center for Genomic Research at the University of Liverpool for sequencing and Ana Kozomara for technical assistance with miRNA expression data sets.

## Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Altuvia Y, et al. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* 33:2697–2706.
- Aravin AA, et al. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell.* 5:337–350.
- Ashburner M, Golic KG, Hawley RS. 2004. *Drosophila: a laboratory handbook*. 2nd ed. New York: Cold Spring Harbor Laboratory Press.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 116:281–297.
- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell.* 136:215–233.
- Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA.* 11:241–247.
- Behura SK. 2007. Insect microRNAs: structure, function and evolution. *Insect Biochem Mol Biol.* 37:3–9.
- Bell ML, Buvoli M, Leinwand LA. 2010. Uncoupling of expression of an intronic microRNA and its myosin host gene by exon skipping. *Mol Cell Biol.* 30:1937–1945.
- Berezikov E, et al. 2010. Evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*. *Nat Genet.* 42:6–9.
- Brown SJ, Denell RE, Beeman RW. 2003. Beetling around the genome. *Genet Res.* 82:155–161.
- Cai Y, et al. 2010. Novel microRNAs in silkworm (*Bombyx mori*). *Funct Integr Genomics.* doi:10.1007/s10142-010-0162-7
- Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet.* 8:93–103.
- Cheng L, Pastrana E, Tavazoie M, Doetsch F. 2009. miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nat Neurosci.* 12:399–408.
- Chiang HR, et al. 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* 24:992–1009.
- Clark AM, et al. 2010. The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Res.* doi:10.1093/nar/gkq083.
- Cloonan N, et al. 2009. RNA-MATE: a recursive mapping strategy for high-throughput RNA-sequencing data. *Bioinformatics.* 25:2615.
- Davis GK, Patel NH. 2002. Short, long, and beyond: molecular and embryological approaches to insect segmentation. *Annu Rev Entomol.* 47:669–699.
- de Wit E, Linsen S, Cuppen E, Berezikov E. 2009. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res.* 19:2064–2074.
- Glazov EA, et al. 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res.* 18:957–964.
- Griffiths-Jones S. 2005. RALEE—RNA Alignment editor in Emacs. *Bioinformatics.* 21:257–259.
- Griffiths-Jones S, et al. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33:D121–D124.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36:D154–D158.
- Guerra-Assunção JA, Enright A. 2010. MapMi: automated mapping of microRNA loci. *BMC Bioinformatics.* 11:133.
- Hofacker IL, et al. 1994. Fast folding and comparison of RNA secondary structures. *Monatshfte Chemie/Chem Mon.* 125:167–188.
- Jagadeeswaran G, et al. 2010. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC Genomics.* 11:52.
- Kloosterman WP, Plasterk RHA. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell.* 11:441–450.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Lee Y, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature.* 425:415–419.
- Lee Y, Jeon K, Lee J, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* 21:4663–4670.
- Liu S, et al. 2010. MicroRNAs show diverse and dynamic expression patterns in multiple tissues of *Bombyx mori*. *BMC Genomics.* 11:85.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Lu J, et al. 2008. The birth and death of microRNA genes in *Drosophila*. *Nat Genet.* 40:351–355.
- Luo Q, et al. 2008. Genome-wide mapping of conserved microRNAs and their host transcripts in *Tribolium castaneum*. *J Genet Genomics.* 35:349–355.
- Makeyev EV, Zhang J, Carrasco MA, Maniatis T. 2007. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell.* 27:435–448.
- Marques-Souza H, Aranda M, Tautz D. 2008. Delimiting the conserved features of hunchback function for the trunk organization of insects. *Development.* 135:881–888.
- Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics.* 25:1335–1337.
- Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. *Curr Opin Genet Dev.* 17:145–150.
- Okamura K, et al. 2008. The regulatory activity of microRNA\* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol.* 15:354–363.
- Peel AD. 2009. Forward genetics in *Tribolium castaneum*: opening new avenues of research in arthropod biology. *J Biol.* 8:106.
- Richards S, et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature.* 452:949–955.
- Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res.* 14:1902–1910.
- 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.
- Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature.* 448:83–86.
- Ruby JG, et al. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* 17:1850–1864.
- Saini HK, Enright AJ, Griffiths-Jones S. 2008. Annotation of mammalian primary microRNAs. *BMC Genomics.* 9:564.
- Saini HK, Griffiths-Jones S, Enright AJ. 2007. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci USA.* 104:17719–17724.
- Savard J, Tautz D, Lercher MJ. 2006. Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evol Biol.* 6:7.
- Shippy TD, et al. 2008. Analysis of the *Tribolium* homeotic complex: insights into mechanisms constraining insect Hox clusters. *Dev Genes Evol.* 218:127–139.
- Singh J, Nagaraju J. 2008. In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). *Insect Mol Biol.* 17:427–436.
- Sokal RR, Rohlf FJ. 1995. *Biometry: the principles and practice of statistics in biological research*. New York: WH Freeman and Company.

- Stark A, et al. 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* 17: 1865–1879.
- Tang G, Maxwell ES. 2008. *Xenopus* microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Res.* 18: 104–112.
- Tautz D, Friedrich M, Schröder R. 1994. Insect embryogenesis—what is ancestral and what is derived? *Development.* 120(Suppl): 193–199.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Tyler DM, et al. 2008. Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. *Genes Dev.* 22:26–36.
- Wang L, Wang S, Li Y, Paradesi MSR, Brown SJ. 2007. BeetleBase: the model organism database for *Tribolium castaneum*. *Nucleic Acids Res.* 35:D476–D479.
- Wheeler B, et al. 2009. The deep evolution of metazoan microRNAs. *Evol Dev.* 11:50–68.
- Wiegmann BM, Kim J, Trautwein MD. 2009. Holometabola. In: Hedges SB, Kumar S, editors. *The timetree of life*. New York: Oxford University Press US. pp. 260–263.
- Yekta S, Shih I, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science.* 304:594–596.

**Associate editor:** Greg Elgar