Evolutionary Changes of the Target Sites of Two MicroRNAs Encoded in the Hox Gene Cluster of *Drosophila* and Other Insect Species

Sayaka Miura*, Masafumi Nozawa, and Masatoshi Nei

Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University

*Corresponding author: E-mail: sum20@psu.edu.

Abstract

MicroRNAs (miRs) are noncoding RNAs that regulate gene expression at the post-transcriptional level. In animals, the target sites of a miR are generally located in the 3′ untranslated regions (UTRs) of messenger RNAs. However, how the target sites change during evolution is largely unknown. MiR-iab-4 and miR-iab-4as are known to regulate the expression of two Hox genes, *Abd-A* and *Ubx*, in *Drosophila melanogaster*. We have therefore studied the evolutionary changes of these two miR genes and their target sites of the Hox genes in *Drosophila*, other insect species, and *Daphnia*. Our homology search identified a single copy of each miR gene located in the same genomic position of the Hox gene cluster in all species examined. The seed nucleotide sequence was also the same for all species. Searching for the target sites in all Hox genes, we found several target sites of miR-iab-4 and miR-iab-4as in *Antp* in addition to *Abd-A* and *Ubx* in most insect species examined. Our phylogenetic analysis of target sites in *Abd-A*, *Ubx*, and *Antp* showed that the old target sites, which existed before the divergence of the 12 *Drosophila* species, have been well maintained in most species under purifying selection. By contrast, new target sites, which were generated during *Drosophila* evolution, were often lost in some species and mostly located in unalignable regions of the 3′ UTRs. These results indicate that these regions can be a potential source of generating new target sites, which results in multiple target genes for each miR in animals.

Key words: birth-and-death evolution, miR-iab-4, miR-iab-4as, miR-10, small RNA, target gene.

Introduction

In animals, microRNAs (miRs) consist of ~22 nucleotides (nt) and regulate expression of protein-coding genes at the post-transcriptional level (Bartel 2004; Kim and Nam 2006). The seed sequence at the 5′ end of a miR usually binds to the target sites of 3′ untranslated regions (UTRs) of messenger RNA (mRNA) transcripts and represses the expression of the target genes (Bartel 2009; Schnall-Levin et al. 2010). In *Drosophila melanogaster*, 176 miR genes have been identified so far (Griffiths-Jones et al. 2008) but only ~15 miRs have been studied about their target genes (Smibert and Lai 2008). Although experimental studies of target genes are quite limited, a bioinformatic study has suggested that the expression of thousands (at least 15%) of protein-coding genes in the genome may be regulated by miR genes (Grun et al. 2005). Therefore, a single miR appears to control many protein-coding genes (Enright et al. 2003; Lim et al. 2005). Hertel et al. (2006) recently reported that human,

D. melanogaster, and *Caenorhabditis elegans* share ~20 miR genes in their genomes. However, bioinformatic analysis suggested that only five miR genes share the target genes among these species (Chen and Rajewsky 2006). This finding indicates that most miRs have experienced gains and losses of their target genes during evolution. It would therefore be interesting to study the evolutionary changes of miR genes and their target genes.

Hox genes control the body segmentation of metazoan embryos. The expression of Hox genes is generally regulated by other Hox genes during the development (Carroll et al. 2005). In addition, in *D. melanogaster*, the miR genes, *MIR-iab-4* and *MIR-iab-4as*, are known to regulate the expression levels of Hox genes, *Abd-A* and *Ubx* (Ronshaugen et al. 2005; Stark et al. 2008; Tyler et al. 2008). The two miR genes share a locus, where *MIR-iab-4as* is encoded by the antisense strand of *MIR-iab-4* (supplementary fig. S1, Supplementary Material online).
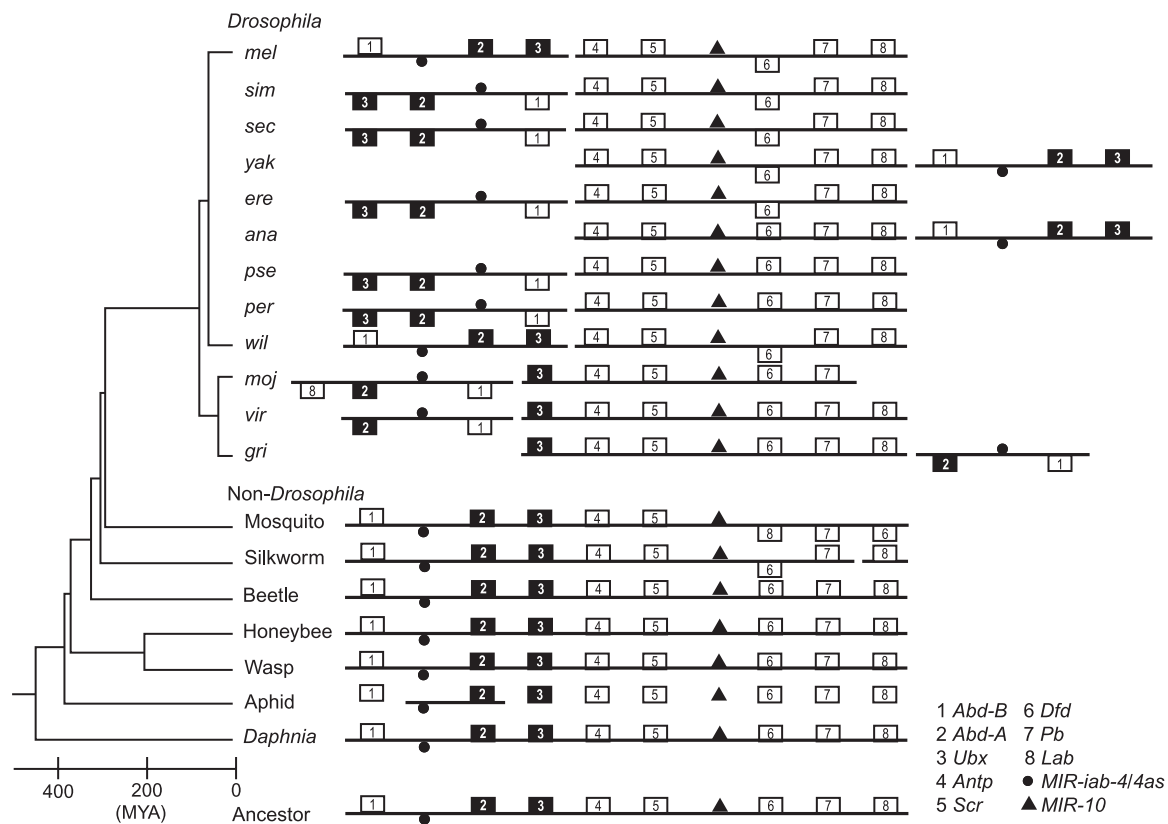
Fig. 1.—Structures of Hox gene clusters and locations of *MIR-iab-4/4as* and *MIR-10*. Rectangles above and below the chromosomes represent Hox genes located on the different strands. The numbers in rectangles correspond to Hox genes, which is given in the right margin. Filled rectangles indicate the experimentally determined target genes (i.e., *Abd-A* and *Ubx*) of miR-iab-4 and miR-iab-4as. The filled circles and triangles indicate *MIR-iab-4/4as* and *MIR-10*, respectively. Note that *MIR-iab-4as* is located on the antisense strand of *MIR-iab-4*. The distance between *Ubx* and *Antp* in *Drosophila melanogaster* is ~10 Mb. The ancestral Hox gene cluster of insects and *Daphnia* is shown at the bottom (Lemons and McGinnis 2006). The divergence times of the species (Hedges et al. 2006) are given at the left side. Abbreviations are as follows: *mel*, *D. melanogaster*; *sim*, *D. simulans*; *yak*, *D. yakuba*; *ere*, *D. erecta*; *ana*, *D. ananassae*; *pse*, *D. pseudoobscura*; *per*, *D. persimilis*; *wil*, *D. willistoni*; *vir*, *D. virilis*; *moj*, *D. mojavensis*; *gri*, *D. grimshawi*; beetle, red flour beetle; wasp, jewel wasp.

These miR genes are located in the Hox gene cluster with another miR gene, *MIR-10*, in *D. melanogaster*, mosquito, and honeybee (Tanzer et al. 2005; fig. 1). However, it is unclear whether other insect species and *Daphnia* contain the miR genes at the same genomic locations and how long their target sites have been maintained during evolution. Since the genomic structures of the Hox gene cluster have been studied in many different species (Von Allmen et al. 1996; Negre et al. 2003; Yasukochi et al. 2004; Negre and Ruiz 2007; Chai et al. 2008), it is possible to examine the evolutionary relationships between the miR genes and Hox genes.

In this study, we first examined the locations of the three miR genes (*MIR-iab-4*, *MIR-iab-4as*, and *MIR-10*) within the Hox gene cluster and studied their evolutionary changes. We then identified the putative target sites of the three miR genes and examined the gains and losses of the target sites of miR-iab-4 and miR-iab-4as in Hox genes in the *Drosophila* species.

Materials and Methods

Identification of the miR and Hox Genes

In this study, we studied the genomes of 12 *Drosophila* species, six other insect species, and *Daphnia*, which are listed in figure 1. The genome sequence of *D. melanogaster* (Release 5) was obtained from the Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/), and the sequences of the other 11 *Drosophila* species (CAF1) were obtained from AAA Wiki (http://rana.lbl.gov/drosophila/). The genome sequences of mosquito (*Anopheles gambiae*) (AgamP3.3), red flour beetle (*Tribolium castaneum*) (Build 2), and honeybee (*Apis mellifera*) (Build 4) were obtained from NCBI (http://www.ncbi.nlm.nih.gov/). The genome sequences of jewel wasp (*Nasonia vitripennis*) (Nvit 2.0) and pea aphid (*Acyrthosiphon pisum*) (Acyr 1.0) were downloaded from the Human Genome Sequencing Center at the Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/). The genome sequence of silkworm (*Bombyx mori*) (version 2.0) was

downloaded from the Silkworm Genome Database (http://silkworm.genomics.org.cn/) and that of *Daphnia* (*Daphnia pulex*) (release 1) was from the wFleaBase (http://wfleabase.org/).

To identify *MIR-iab-4*, *MIR-iab-4as*, and *MIR-10* in these genomes, the nucleotide sequences of the miR genes in *D. melanogaster* were downloaded from the miRBase (release 15) (Griffiths-Jones et al. 2008). Using these sequences as the queries, we conducted a BlastN search (Altschul et al. 1997) against each genome sequence with the cut-off *e* value of 10^{-4}. The hit sequences were extracted and aligned with the *D. melanogaster* sequences by using ClustalW (Thompson et al. 1994). We used the following criteria for miR genes: 1) the sequence shows ≥70% sequence identity with the *D. melanogaster* sequences at the mature region; 2) the free energy of the hairpin structure predicted by the software, mfold (version 3.2) (Mathews et al. 1999; Zuker 2003) is less than or equal to −15 kcal/mol following the previous studies (Lu et al. 2008; Nozawa et al. 2010); 3) the mature sequence is derived from one arm of the hairpin structure; and 4) the number of paired sites between the mature and star (the sequence which form a duplex structure with mature sequence) sequences is >15 following the previous study (Ambros et al. 2003).

To identify the Hox genes, the protein sequences of eight Hox genes (*Abd-B*, *Abd-A*, *Ubx*, *Antp*, *Scr*, *Dfd*, *Pb*, and *Lab*) in *D. melanogaster* were downloaded from NCBI (accession no. NP_996220.1, NP_476693.1, NP_996219.1, NP_996172.1, NP_996164.1, NP_477201.1, NP_996162.1, and NP_476613.1, respectively). Using the sequences as the queries, we conducted a TBlastN search against each genome sequence. The genomic region with the lowest *e* value was considered as a locus of each Hox gene. It should be mentioned that Hox genes of 11 *Drosophila* species were sometimes located on different scaffolds. In this case, we used the chromosomal assemblies constructed by Schaeffer et al. (2008).

Identification of Target Sites in Hox Genes

The target sites of animal miRs are generally located in the 3′ UTRs of the target transcripts. However, 3′ UTRs of Hox genes have not been experimentally determined in any of the species used here except in *D. melanogaster*. Because polyadenylation signals were unclear, we aligned the 3′ downstream sequences of the protein-coding regions with the 3′ UTRs from *D. melanogaster* (accession no. NM_206498.1, NM_169733.2, NM_206497.1, NM_206453.1, NM_206442.1, NM_057853.2, NM_057321.4, and NM_057265.3). In most Hox genes, the 3′ downstream regions were well aligned among *Drosophila* species, and therefore we predicted 3′ UTRs based on the sequence similarity (for the example alignments of the potential 3′ UTRs, see supplementary fig. S2,
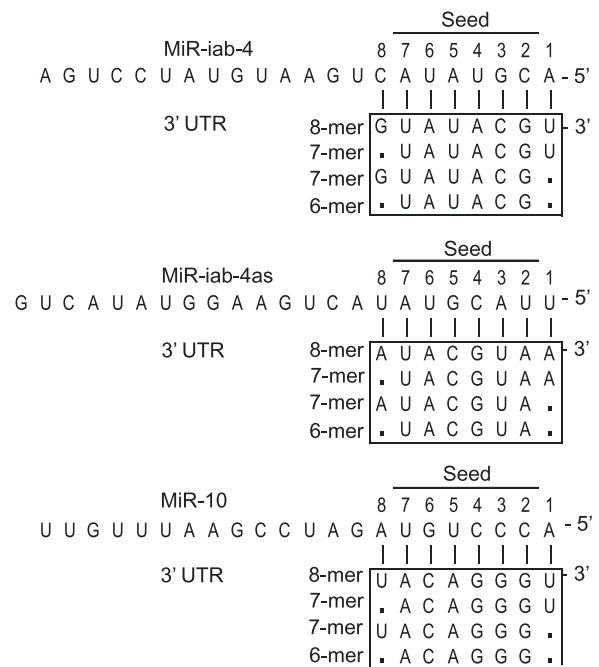
FIG. 2.—Possible target sites of miR-iab-4, miR-iab-4as, and miR-10. Vertical lines indicate Watson-Crick pairings. A 6-mer site pairs with from second nucleotide to seventh nucleotide from the 5′ of the miR, a 7-mer site pairs with from first to seventh nucleotide or from second to eighth, and an 8-mer site pairs with from first to eighth of the miR. Dots can be any nucleotide.

Supplementary Material online). For non-*Drosophila* species and *Drosophila* species in which the 3′ downstream region of the gene was not similar to the *D. melanogaster* 3′ UTR sequence, we assumed that the lengths of 3′ UTRs are essentially the same as those of *D. melanogaster*. Because the 3′ UTRs of the Hox genes in *D. melanogaster* consist of 2,179 nt (*Abd-B*), 2,065 nt (*Abd-A*), 2,396 nt (*Ubx*), 2,035 nt (*Antp*), 2,262 nt (*Scr*), 412 nt (*Pb*), 492 nt (*Dfd*), and 650 nt (*Lab*), we regarded 2,200 nt (*Abd-B*), 2,100 nt (*Abd-A*), 2,400 nt (*Ubx*), 2,100 nt (*Antp*), 2,300 nt (*Scr*), 500 nt (*Pb*), 500 nt (*Dfd*), and 700 nt (*Lab*), respectively, from the stop codon as the approximate 3′ UTRs.

The target sites of miR-iab-4, miR-iab-4as, and miR-10 were inferred in these potential 3′ UTR sequences. The sequences that were complementary to the seed sequence (bases 2–7 of a miR) were regarded as 6-mer target sites (fig. 2). In addition, the sites complementary for bases 2–8, 1–7, and 1–8 of a miR were regarded as 7/8-mer target sites following the previous definition (Bartel 2009). The nucleotide sequences of the target sites are listed in figure 2.

We also determined the orthologous relationships of the target sites based on the alignment of 3′ UTRs of *Drosophila* *Abd-A*, *Ubx*, and *Antp*. We first extracted the target site sequences (or corresponding sequences in the alignment if the target site is absent) as well as 10 nt upstream and downstream sequences including gaps. The sequence identity for

Fig. 3.—Alignment of nucleotide sequences of *MIR-iab-4*, *MIR-iab-4as*, and *MIR-10* of *Drosophila melanogaster (mel)*, mosquito, silkworm, red flour beetle (beetle), honeybee, jewel wasp (wasp), pea aphid (aphid), and *Daphnia*. The mature and star regions are shown with lines on the top of the sequences. Regions that can pair with 6-mer target sites (seed sequences) are represented by boxes. Dots denote the nucleotides that are the same as the nucleotide of *D. melanogaster*, and hyphens represent nucleotide gaps.

all pairs of sequences in the same locations of the alignment was then computed. If at least two of three (upstream, target site [or corresponding site], and downstream) regions showed the sequence identity of ≥60%, the sequences compared were regarded to be orthologous (e.g., supplementary fig. S3, Supplementary Material online). After determining orthologous target sites, we classified the target sites into functional and nonfunctional. Functional target sites are the sites that are perfectly complementary with the seed sequence (i.e., 6-mer and 7/8-mer target sites as defined above), whereas nonfunctional target sites are the sites that contain some mismatches with the seed sequence. Note that because the 3′ downstream sequences of *Abd-A*, *Ubx*, and *Antp* in non-*Drosophila* species were unalignable with the *D. melanogaster* 3′ UTRs as mentioned above, we did not determine the orthologous target sites of these genes in non-*Drosophila* species.

Results

MiR Genes in Hox Gene Clusters

We reconstructed the Hox gene clusters in all the species examined (fig. 1). As in the cases of previous studies (Von Allmen et al. 1996; Negre et al. 2003; Yasukochi et al. 2004; Lemons and McGinnis 2006; Negre and Ruiz 2007; Chai et al. 2008), our results showed several rearrangements of Hox gene clusters during insect evolution (fig. 1). We also determined the numbers and locations of the

miR genes, *MIR-iab-4*, *MIR-iab-4as*, and *MIR-10* in the insect and *Daphnia* genomes. These miR genes were shown to exist only in the Hox gene clusters (fig. 1). The expression of the three miR genes was previously detected in 12 *Drosophila* species and other species used here (Aravin et al. 2003; Tanzer et al. 2005; Ruby et al. 2007; Weaver et al. 2007; Shippy et al. 2008; Singh and Nagaraju 2008; Yu et al. 2008; Ganesh and Ramachandra 2009; Wheeler et al. 2009; Werren et al. 2010), and this indicates that these miR genes are all functional. In addition, the locations of the miR genes are conserved in all the species used, that is, *MIR-iab-4/MIR-iab-4as* is located between *Abd-B* and *Abd-A* and *MIR-10* is between *Scr* and *Dfd*. Note that we previously reported three potential *MIR-10*s in *D. yakuba* but two of them were regarded as non-miR genes if we used more stringent criteria (Nozawa et al. 2010). Therefore, the number and the arrangement of the miR genes appear to have been perfectly conserved during insect and *Daphnia* evolution for ~470 million years (My) (Hedges et al. 2006).

We also compared the nucleotide sequences of the miR genes among the species. The results showed that many substitutions have occurred in the loop region (middle portion of the alignment in fig. 3) and extended stem regions (left and right ends in fig. 3). By contrast, the mature sequence and star sequence, which composes a duplex structure with the mature sequence, have been perfectly conserved except for a single nucleotide change in miR-iab-4 and miR-iab-4as, respectively. Moreover, the seed

Table 1

Numbers of Predicted Target Sites of miR-iab-4, miR-iab-4as, and miR-10 in Hox Genes

Species	4[a]/4as[b]				4/4as/10[c]
	Abd-B	Abd-A	Ubx	Antp	Scr
Drosophila					
D. melanogaster	0/1(1)	1/4*	2*/5*	(2)/2*	0/0/0
D. simulans	0/1(1)	1/4*	2*/5*	(1)/2*	0/0/0
D. sechellia	0/1(1)	1/4*	2*/4*	(1)/2*	0/0/0
D. yakuba	0/1(1)	1/4*	2*/6*	(1)/2*	0/0/0
D. erecta	0/1(1)	1/4*	3(1)*/6*	(1)/2*	0/0/0
D. ananassae	0/0	1/4*	4(1)*/6*	1(1)/3*	0/0/0
D. pseudoobscura	0/0	1/5*	3(1)*/5*	1/3*	0/0/0
D. persimilis	0/0	1/5*	3(1)*/5*	1/3*	0/0/0
D. willstoni	0/0	1/4(1)*	4(1)*/7*	0/3*	(1)/1/(1)
D. mojavensis	1(1)/2*	1/6(1)*	4*/7*	1(1)/4*	1/1/(1)
D. virilis	0/0	1/3*	4*/8*	1(2)*/4*	1/1/(1)
D. grimshawi	0/0	1/4*	3*/9(1)*	(2)/4*	0/0/(1)
Non-*Drosophila*					
Mosquito	0/1	0(1)/1	0(2)/1	0/1	0/1/1
Silkworm	-	3*/4*	0/2(1)*	4*/3*	(1)/1/1
Red flour beetle	1/0	2*/3(1)*	0/0	0/(1)	0/0/1
Honeybee	-	0(1)/0(2)	1/0	0/0	0/2(1)*/0
Jewel wasp	-	1/5(1)*	1(1)/0	2*/1(1)	4(1)*/1/0
Pea aphid	1(1)/(1)	2*/1	3(4)*/1	-	-
Daphnia	0/0	1/3*	2*/2*	1/2*	0/0/0

The numbers in parentheses are the numbers of 6-mer target sites. *Dfd*, *Pb*, and *Lab* mostly do not have target sites of miR-iab-4, miR-iab-4as, or miR-10. All Hox genes except for *Scr* generally do not have target sites of miR-10. Hyphens indicate that the sequences of potential 3′ UTRs include undetermined regions.

[a] The number of target sites of miR-iab-4. The numbers include the number of target sites for both miR-iab-4 and miR-iab-4as.

[b] The number of target sites of miR-iab-4as. The numbers include the number of target sites for both miR-iab-4 and miR-iab-4as.

[c] The number of target sites of miR-10.

* The probability to have the number of target sites by chance is less than 0.05. The probability was calculated by the binomial test (see supplementary method, Supplementary Material online).

sequence was the same for all species examined (for the 12 *Drosophila* species, see supplementary fig. S4, Supplementary Material online). This perfect conservation of the seed sequence implies that the identical nucleotide sequence has been used as their target sites in all the species, if we assume that only the seed sequence is necessary for the target recognition and the target sequence is perfectly complementary to the seed sequence.

Number of Target Sites for miR-iab-4, miR-iab-4as, and miR-10

We predicted target sites of miR-iab-4, miR-iab-4as, and miR-10 in Hox genes of the 12 *Drosophila* species (table 1). We then found that in addition to *Abd-A* and *Ubx*, *Antp* has target sites of miR-iab-4 and miR-iab-4as in most *Drosophila* species. In fact, *Antp* has been predicted as a target gene in 12 *Drosophila* species (Stark et al. 2008), although there is no experimental confirmation. *Abd-B* also has target sites for miR-iab-4as in six *Drosophila* species. In the case of miR-10, *Scr* has been predicted as a target gene in *D. melanogaster* and *D. pseudoobscura* (Enright et al. 2003). Our method also predicted one 6-mer target site for miR-10 in four *Drosophila* species but not in *D. melanogaster* and *D. pseudoobscura*.

Among the *Drosophila* species, the numbers of target sites of miR-iab-4 and miR-iab-4as in *Abd-A*, *Ubx*, and *Antp* are nearly the same (table 1), although some differences exist (e.g., the number of target sites for miR-iab-4as in *Ubx* of *D. grimshawi* is 10, which is greater than those [four to eight] of other species). With some exceptions in *Antp*, most of these target sites were 7/8-mer, which are known to be more efficient to be regulated by miRs than 6-mer sites (Brennecke et al. 2005; Nielsen et al. 2007).

We conducted the binomial test (see supplementary method, Supplementary Material online) to examine whether the observed number of target sites can be expected by chance. The results showed that the observed number of target sites for miR-iab-4as is significantly greater than the expected by chance in *Abd-A*, *Ubx*, and *Antp* (table 1), whereas the observed number for miR-iab-4 can be explained by chance except for *Ubx*. Note that due to the sequence similarity between miR-iab-4 and miR-iab-4as (Stark et al. 2008; Tyler et al. 2008), the target site for miR-iab-4as can also be a target site for miR-iab-4, if the site is perfectly complementary to bases 2–9 of miR-iab-4as (supplementary fig. S1, Supplementary Material online). In this study, we regarded these overlapping sites as the target sites for both miR-iab-4 and miR-iab-4as. Although the

observed number of target sites for miR-iab-4 is significantly greater than the expected by chance in *Ubx*, most target sites for miR-iab-4 also appear to be the target sites for miR-iab-4as (supplementary table S1, Supplementary Material online). All target sites for miR-iab-4 in *Abd-A* were overlapping ones except for *D. virilis*. Considering the fact that in *D. melanogaster* miR-iab-4as represses the expression of both *Abd-A* and *Ubx* efficiently, whereas miR-iab-4 only weakly represses *Ubx* and has no effect on *Abd-A* (Tyler et al. 2008), the overlapping target sites may be mainly for miR-iab-4as.

Target sites for miR-iab-4 and miR-iab-4as were also found in *Abd-A*, *Ubx*, and *Antp* of many non-*Drosophila* species examined. Yet, the numbers of target sites of miR-iab-4 and miR-iab-4as in these Hox genes vary with the species (table 1). In addition, the numbers of target sites for miR-iab-4as in *Abd-A* and *Ubx* are not always greater than those for miR-iab4. This is in contrast to the situation in *Drosophila* species. As extreme cases, red flour beetle does not have any target sites for the miRs in *Ubx*. Of course, we cannot completely rule out the possibility that actual 3′ UTRs in non-*Drosophila* species are longer than the putative 3′ UTRs defined in this study and therefore we could not find all the target sites. However, even when we analyzed the 3 kb downstream regions after the stop codon rather than 2.4 kb, there is no target site in *Ubx* of red flour beetle. In addition to these three genes, mosquito, red flour beetle, and pea aphid also have the target sites in *Abd-B*, but the observed numbers of the target sites are expected by chance. Honeybee and jewel wasp also have at least three target sites of miR-iab-4 and miR-iab-4as in *Scr*, respectively.

Note that other Hox genes (*Dfd*, *Pb*, and *Lab*) do not have target sites for miR-iab-4 and miR-iab-4as in most species examined. It should also be noted that there was no target site for miR-10 in most insect Hox genes examined except for *Scr*. In *Abd-A*, *Ubx*, and *Antp*, by contrast, the target sites for miR-iab-4 and miR-iab-4as exist in the most insect species examined and *Daphnia*. Therefore, it appears that these three genes have been target genes of miR-iab-4 and miR-iab-4as before the divergence of the insect species and *Daphnia*, and some species have gained and lost their target sites during the evolution.

Gains and Losses of Target Sites during the Evolution of 12 *Drosophila* Species

To understand the evolutionary changes of the target sites in more details, we determined the orthologous relationships of the target genes (for details, see Materials and Methods). Because only *Abd-A*, *Ubx*, and *Antp* had target sites of miR-iab-4 or miR-iab-4as in all the *Drosophila* species, we studied the target sites for miR-iab-4 and miR-iab-4as in these three Hox genes. Our analysis identified 9, 20, and 15 groups of target sites at different locations in *Abd-A*, *Ubx*, and *Antp*,

Table 2

Orthologous Relationship of Target Sites for miR-iab-4 and miR-iab-4as in *Abd-A* and their potential functionality in 12 *Drosophila* species

Species	Position of Target Sites								
	1[a]/A[b]	2/A	3/A	4/A	5/A	6/A	7/SA	8/A	9/SA
mel	F	N	-	N	F	F	FF	N	NN
sim	F	N	-	N	F	F	FF	N	NN
sec	F	N	-	N	F	F	FF	N	NN
yak	F	N	-	N	F	F	FF	N	NN
ere	F	N	-	N	F	F	FF	N	NN
ana	F	N	-	N	F	F	FF	N	NN
pse	F	F	-	N	F	F	FF	N	NN
per	F	F	-	N	F	F	FF	N	NN
wil	F	-	(F)	-	F	F	FF	N	NN
moj	F	-	-	F	F	F	FF	(F)	NF
vir	F	-	-	N	F	F	NN	N	FN
gri	F	-	-	N	F	F	FF	N	-
E[c]	383	188	-	104	383	383	383	-	37
C[c]	383*	66	-	43	383*	383*	364*	-	37

The 6-mer functional target sites are shown in parentheses. F, functional target site; N, nonfunctional target site; -, no orthologous site.

[a] Target site name in number. Underlined target sites are the sites that were generated before the divergence of 12 *Drosophila* species.

[b] MiR that binds to the target site: S, miR-iab-4; A, miR-iab-4as; SA, both miR-iab-4 and miR-iab-4as.

[c] E, evolutionary time which is the summation of all the branches of the species tree after the gain of a target site; C, conserved time which is the total time of the conservation of a target site; -, a species-specific target site.

* The probability of conservation by chance is less than 0.05.

respectively (table 2, supplementary tables S2A and S2B, Supplementary Material online). Among them, three functional target sites in *Abd-A* and *Ubx* are conserved in all 12 *Drosophila* species, whereas *Antp* have only one conserved site in the species. These target sites are all 7/8-mer target sites for miR-iab-4as or for both miR-iab-4 and miR-iab-4as. All remaining target sites have been affected by gains and/or losses in some lineages. Therefore, birth-and-death evolution has certainly occurred in the target sites for these miR genes.

To obtain more accurate picture of the birth-and-death evolution of target sites, we reconstructed the ancestral sequence of each target site by using the maximum likelihood method (Yang et al. 1995) and estimated the numbers of potential target sites in ancestral species and gains and losses of the target sites. The timing of each gain was assigned to the lineage in which the sequence became a potentially functional target site, and the timing of loss was assigned to the lineage in which the 6-mer region experienced at least one substitution or one indel. A predicted target site that only exists in a particular species was assumed to be gained during the evolution of the species. The results show that the ancestor of 12 *Drosophila* species had four target sites in *Abd-A* and *Antp* and six sites in *Ubx* (fig. 4). Although the number of target sites has been roughly constant during the evolution, several lineage or species-specific gains and losses were observed. For an extreme example, *Ubx* of
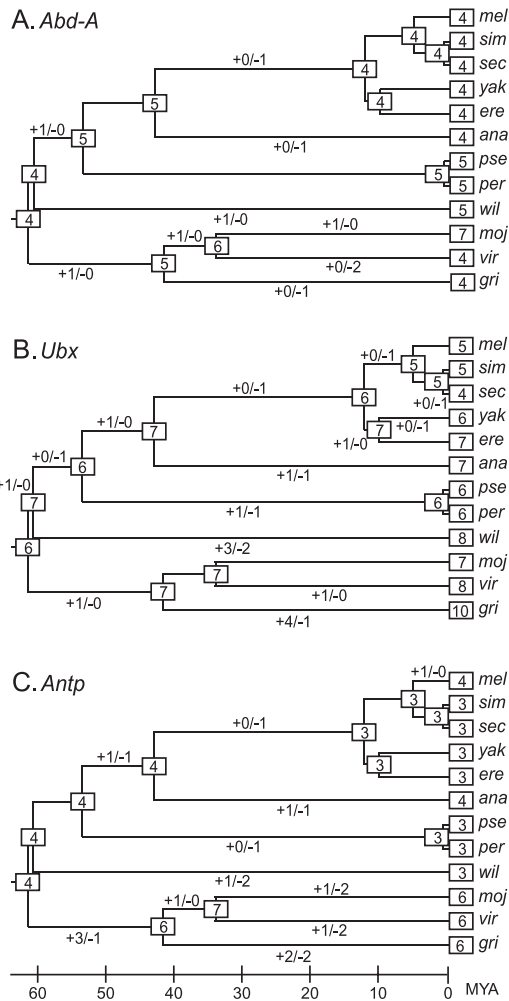
Fig. 4.—The numbers of target sites for miR-iab-4 and miR-iab-4as in *Abd-A* (A), *Ubx* (B), and *Antp* (C) and the numbers of gains and losses of the target sites during the evolution of 12 *Drosophila* species. The numbers of target sites in extant and ancestral species are represented in boxes. The numbers at the branch are the numbers of gains (+) and losses (-) of target sites. Divergence times shown below the trees are from Tamura et al. (2004).

Table 3

Numbers of Gains and Losses of Target Sites by Nucleotide Substitutions and/or Indels

	Gain			Loss		
Target	Substitution	Indel	Both	Substitution	Indel	Both
Abd-A	2	1	2	1	0	2
Ubx	3	1	10	5	1	0
Antp	5	1	5	6	1	8
Total	10	3	17	12	2	10

than half of gains grouped as "both" were classified in this category. Similarly, 22 of 24 losses of target sites have occurred by substitutions or both substitutions and indels.

In the previous section, we found that a large number of target sites for miR-iab-4as in *Abd-A*, *Ubx*, and *Antp* are likely under functional constraints in *Drosophila* species. Here, to examine whether each orthologous target site has been under functional constraints, we computed the expected time of conservation of a target site under neutral evolution. We first estimated the substitution rate in 3' UTRs of *Abd-A*, *Ubx*, and *Antp* for each pair of 12 *Drosophila* species. The average substitution rate was ~0.002 per site per My for all *Abd-A*, *Ubx*, and *Antp*. The probability that a 6-mer target site is conserved for T My by chance is therefore calculated as $(1 - 0.002T)^6$. Similarly, the probabilities of conservation of 7-mer and 8-mer sites by chance are $(1 - 0.002T)^7$ and $(1 - 0.002T)^8$, respectively. Therefore, the probabilities become less than 5% after ~160, ~180, and ~200 My for 8-mer, 7-mer, and 6-mer target sites, respectively (fig. 5). For example, suppose that there is a 7-mer orthologous target site conserved in two species whose divergence time is 100 My. Since the evolution occurs independently in the two species after the divergence, the total conserved time of the target site should be at least 200 My (100 + 100 My). The conserved time (200 My) is greater than 180 My, which is the maximum conserved time by chance for 7-mer at the 5% significance level. Therefore,

D. grimshawi has acquired five target sites after the divergence of 12 *Drosophila* species. Four of them have been generated after the divergence from *D. mojavensis* and *D. virilis* around 43 million years ago (Tamura et al. 2004).

We next examined how nucleotide substitutions and indels have affected the gains and losses of the target sites in *Abd-A*, *Ubx*, and *Antp* (table 3). The results show that 27 target sites of 30 sites have been gained by substitution or both substitutions and indels. It should be mentioned that a target site was regarded to be gained by both substitutions and indels if the ancestral sequence before it became a target site could not be estimated due to the absence of orthologous target sites (irrespective of functional or nonfunctional) in the outgroup *Drosophila* species; more

Fig. 5.—Probabilities of conservations of target sites by chance. A dashed line is the probability of 0.05.

the conservation of the target site is unlikely to be expected under neutral evolution but seems to be due to functional constraints. Note that this analysis is based on the assumption that the substitution rate is the same across the 3′ UTRs in all evolutionary lineages, which is probably unrealistic. Nevertheless, this rough estimation would still be useful to obtain a general idea about functional constraints on each orthologous target site.

We applied this calculation to the actual data. Among 29 groups of target sites identified in *Abd-A* and *Ubx* (table 2 and supplementary table S2*A*, Supplementary Material online), all ten target sites that existed before the divergence of 12 *Drosophila* species (old target sites, see fig. 4) are 7/8-mer and conserved more than 200 My. Of course, it is also possible that these old target sites were located in regions that have other functions and are under purifying selection for reasons other than for miR regulation. However, six target sites of ten conserved old sites are shared in all the species, and it is quite unlikely that all the six sites overlap with the sites for other functions. In the case of *Antp*, however, two of four old target sites have not been conserved more than expected by chance (supplementary table S2*B*, Supplementary Material online). Yet, note that these two nonconserved target sites are 6-mers at the ancestor of 12 *Drosophila* species so that the recognition by the miRs may not be efficient, and consequently, the functional constraints on these sites may not be stringent.

All the remaining target sites (new target sites) in the three Hox genes that were generated during the evolution of *Drosophila* species show the conserved times of less than 160 My. Note that most new target sites were generated very recently, and the evolutionary time, which is the sum of all the branch lengths after the gain of a target site are still less than 160 My (table 2 and supplementary table S2, Supplementary Material online). Therefore, we cannot determine whether these sites are under functional constraints or not. Nevertheless, two target sites (no. 2 in *Abd-A* and no. 5 in *Ubx*, see table 2 and supplementary table S2*A*, Supplementary Material online, respectively) showed evolutionary times of more than 160 My and conserved times of less than 160 My. Therefore, these two new target sites may be under less functional constraints compared with most old target sites.

Discussion

In this study, we have examined the evolution of miR genes and their target sites in the insect and *Daphnia* Hox gene clusters. We found that the miR genes, *MIR-iab-4*, *MIR-iab-4as*, and *MIR-10*, are highly conserved with respect to the copy numbers and nucleotide sequences. The identical seed sequences among the species examined indicate that same sequence motifs have been used for the target sites of these miRs during insect and *Daphnia* evolution. Yet, the

number of target sites in *Abd-A*, *Ubx*, and *Antp* seems to vary considerably among insect species and *Daphnia*. We have also found that the observed number of target sites for miR-iab-4as in these three Hox genes appears to have been maintained by functional constraints in all the *Drosophila* species. However, this does not necessarily mean that all target sites are equivalent for regulation by the miRs. We found that the old target sites that existed before the divergence of 12 *Drosophila* species tend to have been highly conserved and under purifying selection, whereas some new target sites may not. It is therefore possible that the old target sites may be more efficient than others.

It has been suggested that the location of the target sites in the secondary structures of 3′ UTRs affects the repression efficiency by miRs (Robins et al. 2005). More specifically, if the target sites are located at the loop region of the secondary structures of 3′ UTRs, the target genes may be repressed more efficiently. We therefore examined whether the old target sites tend to be located in the loop region of the 3′ UTRs compared with the new target sites in the *D. melanogaster Abd-A*, *Ubx*, and *Antp*. For this analysis, we used the secondary structures of 3′ UTRs with the lowest free energy predicted by the software, RNAfold (Gruber et al. 2008). The results showed that the proportions of the target sites in the loop region were not significantly different between old and new target sites ($3/10 = 30\%$ and $1/3 = 33\%$, respectively). However, the number of target sites analyzed was small. In addition, several different structures of 3′ UTRs with similar free energies were predicted, so that it is very difficult to determine the real structure. Moreover, if we used entire mRNA for the prediction, secondary structures of the 3′ UTR portion became quite different. In fact, two target sites in *Abd-A* were predicted to be located in the loop regions when only 3′ UTR was used for the prediction, but the locations of these sites changed to the stem regions when entire mRNA was considered (supplementary table S3, Supplementary Material online). Therefore, at present our conclusion about the different efficiencies between old and new target sites is tentative. It is important to verify the functionality of these putative target sites experimentally and examine whether there are any functional differences between the old and new target sites in the future.

We found that several new target sites that have been generated during *Drosophila* evolution are located in unalignable regions of 3′ UTRs. These regions have accumulated more mutations than other regions and therefore appear to have a potential to generate target sites of miRs. Also, we found several losses of target sites. Target sites in animals generally range from only 6 to 8 nt long and seem easy to be generated or lost by mutations. Therefore, existence of the 3′ UTRs with frequent mutations may explain why animal miRs tend to have many target genes than plant miRs and why target genes for each miR are often different among species (Chen and Rajewsky 2006).

In this study, we have primarily focused on the target sites of miR-iab-4, miR-iab-4as, and miR-10 in the Hox genes. However, it is possible that these miRs also regulate the expression of other protein-coding genes. Therefore, we examined the number of 6-mer putative target sites of miR-iab-4, miR-iab-4as, and miR-10 in each of the 3′ UTRs of 12,079 *D. melanogaster* genes (data from the FlyBase; http://flybase.org/). The results showed that quite a few 3′ UTRs (1,030/12,079 = 8.5% for miR-iab-4 and miR-iab-4as; 294/12,079 = 2.4% for miR-10) have at least one putative target sites (supplementary table S4, Supplementary Material online). Among them, *Ubx* and *Abd-A* are the genes with the largest numbers of target sites (five and four, respectively) for miR-iab-4as. By contrast, the numbers of target sites for miR-iab-4 in *Ubx* and *Abd-A* are only two and one, respectively, and many genes have the same or larger numbers of target sites for miR-iab-4 than *Ubx* or *Abd-A*. There is no target sites for miR-10 in *Ubx* and *Abd-A*, although this miR is known to regulate several Hox genes (*HoxA1*, *HoxA3*, and *HoxD10*) in vertebrates (Lund 2010). Nevertheless, there are nearly 300 genes that have at least one target site for miR-10, and in the most extreme case, the gene, *Vmat*, has as many as nine target sites for miR-10. Taken together, it is quite possible that these miRs also regulate non-Hox genes.

In vertebrates, there are usually four Hox gene clusters (HoxA, HoxB, HoxC, and HoxD). These gene clusters except for HoxD encode *MIR-196* at the same positions as that of *MIR-iab4/4as* in insects and *Daphnia*, although there is no sequence similarity between them (Tanzer et al. 2005). In addition, it is known that its target genes, HoxA7, HoxB8, HoxC8, and HoxD8, are orthologs of *Ubx*, which is a target gene of miR-iab-4/4as in insects and *Daphnia* (Mansfield et al. 2004; Yekta et al. 2004; Hornstein et al. 2005). Many predicted target sites for miR-196 existed before the divergence of mammals (Yekta et al. 2008), indicating that these target sites for miR-196 have apparently been under purifying selection. Therefore, the miR-based regulation of the Hox genes appears to be shared between vertebrates and invertebrates.

Recent experimental studies have demonstrated that regulatory elements such as *cis*-regulatory elements have played pivotal roles in morphological evolution (Stern 1998; Liubicich et al. 2009; Pavlopoulos et al. 2009). MiRs are also regulatory elements and indeed involved in many developmental processes, including cell proliferation, apoptosis, organ growth, and cell differentiation (Jaubert et al. 2007). In animals, the expression patterns of the Hox genes are particularly important in the developmental process. Therefore, if the changes of the number of target sites have affected the expression pattern of the Hox genes, evolution of the target sites for miR-iab-4 and miR-iab-4as may have caused morphological evolution. Of course, a gene is generally regulated by multiple

systems (Hobert 2004; Chen and Rajewsky 2007), and actually, the expression of *Ubx* is also downregulated by transcription factors, Abd-A, in addition to miR-iab-4 and miR-iab-4as in *D. melanogaster* (Bender 2008). Therefore, it is also possible that the effect on expression patterns by the evolution of target sites may be compensated by other regulatory systems. Yet, note that in rice, the differences in height, tiller number, and panicle morphology between *japonica* and *indica* are indeed caused by a single nucleotide substitution in the target site of miR-156 (Jiao et al. 2010; Miura et al. 2010).

In this study, we have made full use of bioinformatic techniques to investigate the evolution of target sites for miRs. However, there are some caveats about our analysis. First, we assumed that lengths of 3′ UTRs in *Drosophila* and non-*Drosophila* species are essentially the same, but this assumption may be unrealistic. Therefore, the number of target sites in non-*Drosophila* species may not be so reliable. However, our analysis is mostly based on *Drosophila* species, in which potential 3′ UTRs are well aligned. Therefore, the ambiguity of the number of target sites in the non-*Drosophila* Hox genes should not affect our conclusion. Second, the target sites predicted in this study were detected only based on the perfect complementarity to the seed sequences as mentioned above, and it is necessary to confirm the functionality of these target sites experimentally. In fact, some potential target sites that are perfectly complimentary to the seed sequences of miRs are actually suggested to be nonfunctional (Didiano and Hobert 2006, 2008). Also, some functional target sites do not have a complete complementarity to the seed sequences of miRs (Slack et al. 2000), although this situation seems to be very rare (only 1% of mammalian conserved target sites) (Bartel 2009). For example, a target site in mouse *HoxB8* lacks perfect complementarity to the seed sequence of miR-196 (Yekta et al. 2004; Hornstein et al. 2005; Yekta et al. 2008). In addition, some researchers have considered wobble pairings between the seed sequences and 3′ UTRs in predicting target sites (Enright et al. 2003; Stark et al. 2003). Therefore, we estimated the target sites by allowing one wobble pairing with seed sequences. However, the number of predicted target sites for the miRs in the *Drosophila* Hox genes increased by at most two in most cases (supplementary table S5, Supplementary Material online, see also table 1). In addition, the number of target sites for miR-iab-4as remained greater than the number for miR-iab-4. Note also that the target sites including wobble pairs are much less effective to repress gene expression of the target genes (Brennecke et al. 2005). Indeed, our results based on the perfect complementarity between seed sequence and targets are consistent with the experimental study in *D. melanogaster* (Tyler et al. 2008). Therefore, our conclusion appears to be robust even though some atypical target sites exist in the Hox genes.

Supplementary Material

Supplementary tables S1–S5, figures S1–S4, and supplementary method are available at *Genome Biology and Evolution online* (http://gbe.oxfordjournals.org/).

Acknowledgments

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.

Ambros V, et al. 2003. A uniform system for microRNA annotation. RNA 9:277–279.

Aravin AA, et al. 2003. The small RNA profile during *Drosophila* melanogaster development. Dev Cell. 5:337–350.

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.

Bender W. 2008. MicroRNAs in the *Drosophila* bithorax complex. Genes Dev. 22:14–19.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. PLoS Biol. 3:e85.

Carroll S, Grenier J, Weatherbee S. 2005. From DNA to diversity: molecular genetics and the evolution of animal design. Malden (MA): Wiley-Blackwell.

Chai C-L, et al. 2008. A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, Bombyx mori. Insect Biochem Mol Biol. 38:1111–1120.

Chen K, Rajewsky N. 2006. Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes. Cold Spring Harb Symp Quant Biol. 71:149–156.

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet. 8:93–103.

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

Didiano D, Hobert O. 2008. Molecular architecture of a miRNA-regulated 3' UTR. RNA 14:1297–1317.

Enright AJ, et al. 2003. MicroRNA targets in *Drosophila*. Genome Biol. 5:R1.

Ganesh S, Ramachandra SN. 2009. Computational identification of microRNA homologs from Acrythosiphon pisum (Pea Aphid). J Comp Intell Bioinf. 2:109–119.

Griffiths-Jones S, Saini H, van Dongen S, Enright A. 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36:D154–D158.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. Nucleic Acids Res. 36:W70–W74.

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. PLoS Comput Biol. 1:e13.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971–2972.

Hertel J, et al. 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics. 7:25.

Hobert O. 2004. Common logic of transcription factor and microRNA action. Trends Biochem Sci. 29:462–468.

Hornstein E, et al. 2005. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. Nature 438:671–674.

Jaubert S, Mereau A, Antoniewski C, Tagu D. 2007. MicroRNAs in *Drosophila*: the magic wand to enter the Chamber of Secrets? Biochimie 89:1211–1220.

Jiao Y, et al. 2010. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nat Genet. 42:541–544.

Kim V, Nam J. 2006. Genomics of microRNA. Trends Genet. 22:165–173.

Lemons D, McGinnis W. 2006. Genomic evolution of Hox gene clusters. Science 313:1918–1922.

Lim LP, et al. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 433:769–773.

Liubicich DM, et al. 2009. Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. Proc Natl Acad Sci U S A. 106:13892–13896.

Lu J, et al. 2008. The birth and death of microRNA genes in *Drosophila*. Nat Genet. 40:351–355.

Lund AH. 2010. miR-10 in development and cancer. Cell Death Differ. 17:209–214.

Mansfield JH, et al. 2004. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet. 36:1079–1083.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 288:911–940.

Miura K, et al. 2010. OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat Genet. 42:545–549.

Negre B, Ranz JM, Casals F, Caceres M, Ruiz A. 2003. A new split of the Hox gene complex in *Drosophila*: relocation and evolution of the gene labial. Mol Biol Evol. 20:2042–2054.

Negre B, Ruiz A. 2007. HOM-C evolution in *Drosophila*: is there a need for Hox gene clustering? Trends Genet. 23:55–59.

Nielsen CB, et al. 2007. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA 13:1894–1910.

Nozawa M, Miura S, Nei M. 2010. Origins and evolution of microRNA genes in *Drosophila* species. Genome Biol Evol. 2:180–189.

Pavlopoulos A, et al. 2009. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. Proc Natl Acad Sci U S A. 106:13897–13902.

Robins H, Li Y, Padgett RW. 2005. Incorporating structure to predict microRNA targets. Proc Natl Acad Sci U S A. 102:4006–4009.

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, et al. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. Genome Res. 17:1850–1864.

Schaeffer SW, et al. 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics 179:1601–1655.

Schnall-Levin M, Zhao Y, Perrimon N, Berger B. 2010. Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3′UTRs. Proc Natl Acad Sci U S A. 107:15751–15756.

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.

Slack FJ, et al. 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell. 5:659–669.

Smibert P, Lai EC. 2008. Lessons from microRNA mutants in worms, flies and mice. Cell Cycle. 7:2500–2508.

Stark A, Brennecke J, Russell RB, Cohen SM. 2003. Identification of *Drosophila* MicroRNA targets. PLoS Biol. 1:E60.

Stark A, et al. 2008. A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands. Genes Dev. 22:8–13.

Stern DL. 1998. A role of Ultrabithorax in morphological differences between *Drosophila* species. Nature 396:463–466.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol Biol Evol. 21:36–44.

Tanzer A, Amemiya CT, Kim CB, Stadler PF. 2005. Evolution of microRNAs located within Hox gene clusters. J Exp Zoolog B Mol Dev Evol. 304:75–85.

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.

Von Allmen G, et al. 1996. Splits in fruitfly Hox gene complexes. Nature 380:116.

Weaver DB, et al. 2007. Computational and transcriptional evidence for microRNAs in the honey bee genome. Genome Biol. 8:R97.

Werren JH, et al. 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 327: 343–348.

Wheeler BM, et al. 2009. The deep evolution of metazoan microRNAs. Evol Dev. 11:50–68.

Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.

Yasukochi Y, et al. 2004. Organization of the Hox gene cluster of the silkworm, Bombyx mori: a split of the Hox cluster in a non-*Drosophila* insect. Dev Genes Evol. 214:606–614.

Yekta S, Shih IH, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. Science 304:594–596.

Yekta S, Tabin CJ, Bartel DP. 2008. MicroRNAs in the Hox network: an apparent link to posterior prevalence. Nat Rev Genet. 9:789–796.

Yu X, et al. 2008. The silkworm (Bombyx mori) microRNAs and their expressions in multiple developmental stages. PLoS One. 3:e2997.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.

Associate editor: Eugene Koonin