RNA Biology 9:3, 1–7; March 2012; © 2012 Landes Bioscience

Evolution and function of the extended miR-2 microRNA family

Antonio Marco,* Katarzyna B. Hooks and Sam Griffiths-Jones*
Faculty of Life Sciences, University of Manchester; Manchester, UK

MicroRNAs are essential post-transcriptional regulators. Many animal microRNAs are clustered in the genome, and it has been shown that clustered microRNAs may be transcribed as a single transcript. Polycistronic microRNAs are often members of the same family, suggesting a role of tandem duplication in the emergence of clusters. The mir-2 microRNA family is the largest in *Drosophila melanogaster*, with 8 members that are mostly clustered in the genome. Previous studies suggest that the copy number and genomic distribution of mir-2 family members has been subject to significant change during evolution. The effects of such changes on their function are still unknown. Here we study the evolution of function in the mir-2 family. Our analyses show that, in spite of the change in number and organization among invertebrates, most mir-2 loci produce very similar mature microRNA products. Multiple mature miR-2 sequences are predicted to target genes involved in neural development in Drosophila. These targeting properties are conserved in the distant species *Caenorhabditis elegans*. Duplication followed by functional diversification is frequent during protein-coding gene evolution. However, our results suggest that the production of microRNA clusters by gene duplication rarely involves functional changes. This pattern of functional redundancy among clustered paralogous microRNAs reflects birth-and-death evolutionary dynamics. However, we identified a small number of mir-2 sequences in Drosophila that may have undergone functional shifts associated with genomic rearrangements. Therefore, redundancy in microRNA families may facilitate the acquisition of novel functional features.

Keywords: MicroRNA family, miRNA, polycistronic, evolution, microRNA targets, genomics

Submitted: 11/03/11

Revised: 12/22/11

Accepted: 12/23/11

http://dx.doi.org/10.4161/rna.9.3.19160

*Correspondence to: Antonio Marco and Sam Griffiths-Jones;
Email: antonio.marco@manchester.ac.uk and sam.griffiths-jones@manchester.ac.uk

Introduction

MicroRNAs, crucial regulators of gene expression at the post-transcriptional level, are often clustered in the genome.[1] According to miRBase,[2] more than a quarter of both Drosophila and human microRNAs are less than 10 kb away from other microRNAs. These clustered microRNAs are often co-expressed, suggesting that they are produced from a single transcript.[3-6] The majority of microRNA clusters contain members of the same family, indicating a major role of tandem duplication in cluster formation.[7-9] In the case of protein-coding genes, duplication is acknowledged as the main source of functional innovation, since duplicates are free to diversify in their functions.[10] Similarly, duplicated microRNAs may acquire new targets and therefore novel functions. However, microRNAs processed from the same transcript are linked by their expression pattern, imposing a functional constraint on their evolutionary diversification. Whether microRNA tandem duplications facilitate the emergence of new functions or generate redundant products remains to be explored.

Mir-2 is the largest microRNA family in *Drosophila melanogaster* and one of the first to be discovered.[11-13] The mir-2 family has 8 members in the *D. melanogaster* genome (*mir-2a-1*, *mir-2a-2*, *mir-2b-1*, *mir-2b-2*, *mir-2c*, *mir-13a*, *mir-13b-1* and *mir-13b-2*), six of which are organized in two clusters.[14] In most other studied insects, there are five mir-2 sequences encoded by a single transcript (see ref. 15 and references therein). *Caenorhabditis elegans* has only one mir-2 sequence.[12,13]

Here we study the mir-2 family to investigate the impact of microRNA family expansions on functional diversification.

We combine comparative genomics with expression data analyses and functional annotation of predicted targets to compare the functional features of mir-2 sequences. Our results will help us to understand the role of tandem microRNA duplications in the evolution of gene regulation.

Results

Mir-2 is a conserved microRNA family in invertebrates. In order to characterize mir-2 family members, we performed comprehensive sequence similarity searches against multiple sequenced organisms (see Materials and Methods). We detected mir-2 hairpin precursor sequences in many invertebrates (**Fig. 1A**; **File S1**) but none in vertebrate species. The 3' arm of the hairpin is highly conserved, although the many changes in the 5' arm are fully consistent with the precursor hairpin structure (**Fig. 1A**). The gene copy number is highly variable among species (from one in *C. elegans* to eight in *D. melanogaster*) suggesting that the mir-2 content of each lineage is the product of multiple birth-and-death events.

Since mir-2 sequences are short and very similar, their genomic contexts can improve our ability to annotate and explore their evolutionary origins. The genomic organization of mir-2 family members across phyla (**Fig. 2**) suggests that the ancestral mir-2 microRNA was clustered with mir-71, an evolutionarily unrelated microRNA. Mir-71 itself is found in protostomes, but also in cephalochordates, hemichordates and echinoderms.[16,17] The origin of mir-71 therefore pre-dates the split of protostomes and deuterostomes, although it has been lost in chordates. Mir-2 arose later, most likely before the last common ancestor of protostomes. Although mir-71 and mir-2 are still linked in most species, mir-71 has been lost independently in two dipteran lineages. The expansion of the mir-2 family by tandem duplication and deletion has generated mir-2 clusters of different lengths in different species. The mir-13 subfamily has a conserved characteristic one-nucleotide deletion in its 3' arm (**Fig. 1A**), indicating that these sequences originated from duplicated mir-2 locus in the common ancestor of insects. Combined analysis of sequence conservation and cluster structure (**Figs. 1A and 2**) suggests that the ancestral insect cluster split in two in the Drosophila lineage, with subsequent additional duplications. As a consequence, different mir-2 copies in Drosophila are under the transcriptional control of different regulatory sequences.

Functional conservation and redundancy of mir-2 products. The pattern of sequence conservation in the mir-2 family sequences shown in **Figure 1** suggests that the dominant mature microRNA is produced from the 3' arm of mir-2 precursors. Our re-analysis of deep-sequencing data from *D. melanogaster*, *Tribolium castaneum* and *C. elegans* confirms that the 3' arm is highly expressed compared with the 5' arm in most mir-2 family members (**File S2**). Deep-sequencing analyses from honeybee and silkworm also reveal the

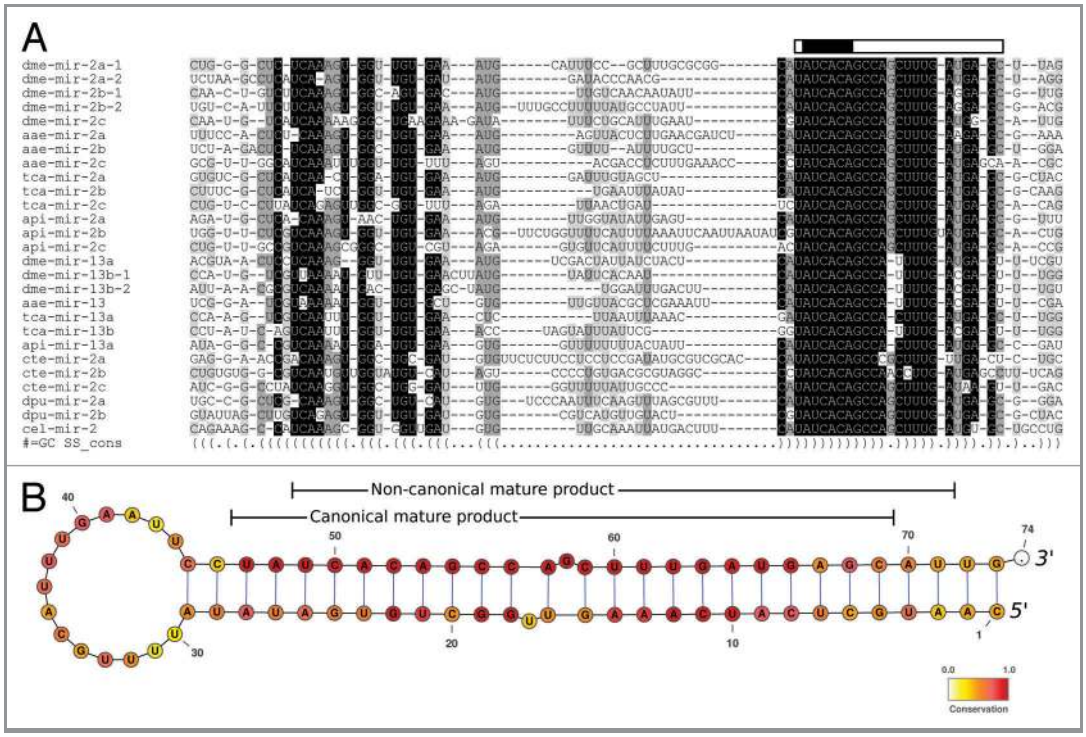

Figure 1. Sequence conservation in the mir-2 family. (A) The alignment of mir-2 precursor sequences in representative genomes, shadowed by sequence conservation (visualized using Ralee[42]), where darker tones reflect higher conservation. Structure of the consensus sequence is shown below the alignment in dot-bracket annotation. The open white box over the alignment indicates the canonical mature product, with the seed sequence highlighted (black). (B) Consensus structure of the mir-2 precursor in invertebrates, colored with VARNA[43] according to sequence conservation. The canonical and non-canonical mature products produced by some mir-2 precursors are also indicated.

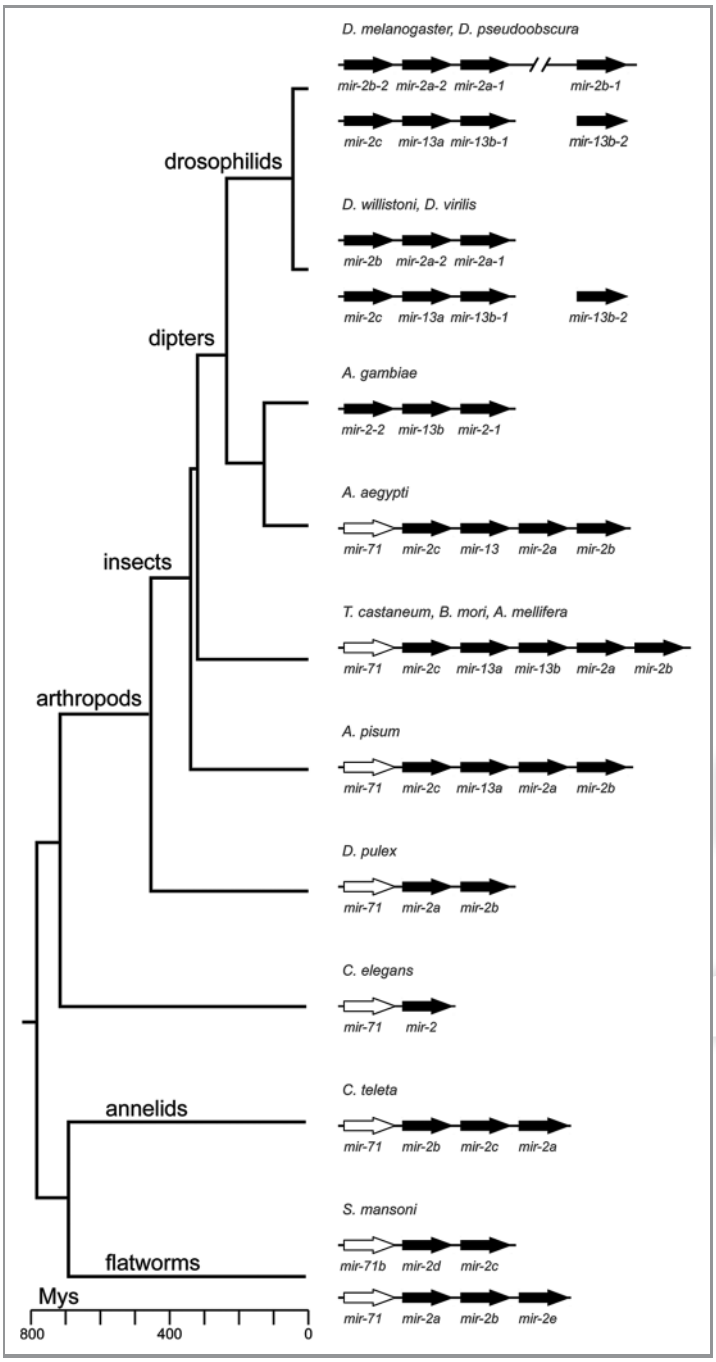

Figure 2. Copy distribution of mir-2 sequences. Phylogenetic tree of invertebrate species and genomic organization of mir-2 sequences. Divergence times were extracted from ref. 47. Black arrows depict mir-2 family members, and white arrow mir-71 sequences. Arrows linked by the same straight line indicate microRNAs linked in the genome by less than 10 kb.

same expression pattern.[18,19] We observe that most mir-2 sequences conserve the location of the Drosha and Dicer cleavage sites. This position determines the first nucleotides of the microRNA, and hence the seed sequence. The seed is defined as nucleotides 2 to 7 of a mature microRNA, and it is crucial for transcript targeting.[20] Since sequence conservation is very high in the 3' arms, seed sequences are the same for all mir-2 family products in which the Dicer cleavage site is conserved (**Fig. 1**).

Functional shifts in mir-2 products. Deep sequencing data from Drosophila suggest that the 3' arm of mir-2a produces two alternative mature products, in contrast to the majority of mir-2 family members. Each accounts for a significant proportion of the reads produced by mir-2a loci (47% and 28%), and they are offset from one another by 2 nucleotides. The first of these products (the 5'-most) is processed identically to the conserved mature sequence produced from the majority of mir-2 family members, termed the 'canonical' product here (**Fig. 1B**). The second is offset by 2 nucleotides in the 3' direction, and is termed the 'non-canonical' product (**Fig. 1B**). Both of these products map exactly to two alternative hairpin precursors called mir-2a-1 and mir-2a-2, suggesting that both products could potentially be made from either locus. However, the 5' arms of these two hairpins are not identical in sequence, and therefore reads mapping to the 5' arms can be assigned to one or other hairpin. It has been previously reported that the characteristic pattern of two nucleotide overhang at the 3-prime ends of mature microRNA duplexes allows the assignment of reads from the 3' arm to one or other hairpin, even though the 3' arm sequences are identical.[14,21,22] This approach predicts that the non-canonical mature sequence, offset by 2 nucleotides, is produced overwhelmingly from the mir-2a-2 locus, whereas mir-2a-1 is processed identically to the other mir-2 family members. Analysis of deep sequencing data from an RNA immunoprecipitation (RIP-seq) of Argonaute proteins shows that both canonical and non-canonical mature products are loaded into the RNA induced silencing complex (RISC),[23] and are therefore likely to be functional. The seed sequences of canonical and non-canonical mature microRNAs are offset, and hence differ in sequence, suggesting that they regulate different targets. Drosophila mir-2c also produces an offset, non-canonical, mature product. However this microRNA is expressed at a very low level and is not found in the AGO RIP-seq data set.[22,23] Our data show that a significant fraction of non-canonical mir-2 products are also expressed from mir-2 loci in *T. castaneum* (**File S2**) and in the honeybee *Apis mellifera* (data not shown).

However, the strategy described above cannot be applied to assign reads to a single locus.

Unlike other mir-2 members, the mir-2a-2 precursor produces approximately equal amounts of mature sequences from each arm of the hairpin.[21,22] Nevertheless, mature sequences derived from the 5' arm are not observed in AGO RIP-seq experiments[22,23] and are not, therefore, predicted to be loaded into the RISC complex. This further supports a dominant role of the mature sequence from the 3' arm across the mir-2 family.

Mir-2 products are likely to target neural genes. We have shown that mature products from mir-2 loci are highly conserved and are likely to have the same targeting properties. Do mir-2 sequences therefore conserve their targets throughout evolution? We address this question by comparing the targets of *D. melanogaster* and *C. elegans* miR-2 mature sequences. We used the canonical seed method[20] to predict transcripts whose 3'UTR are targeted by all miR-2 family members in *D. melanogaster* and the only miR-2 sequence in *C. elegans* (see Methods). All but two miR-2 sequences in Drosophila have identical seeds and therefore identical predicted target sets (**Fig. 1A**).The two microRNAs with different targets were miR-2a-2 and miR-2c, which are offset with respect to the canonical mir-2 products (**Fig. 1B**).

We mapped Gene Ontology terms to the predicted targets of miR-2 family members, and analyzed the set of terms that were statistically enriched in the targeted gene set (see Materials and Methods). We focused on terms within the 'Developmental process' category, which is particularly informative for development and tissue specificity.[24] We detected 675 genes targeted by Drosophila the miR-2 canonical sequence, and 979 for the functional *Caenorhabditis* miR-2 product. For both Drosophila and *Caenorhabditis,* we observed an enrichment in genes involved in neural development (**Table 1**). We therefore predict a role for mir-2 in neural function. Indeed, expression data from deep-sequencing analyses in Drosophila indicate that mir-2 products are highly expressed in adult heads.[14] We also studied the targets of the non-canonical products from miR-2a-2 and miR-2c in Drosophila. Both miR-2a-2 and miR-2c are predicted to target 286 genes. In these cases, we did not find any significantly enriched functional classes (not shown).

The seed model for microRNA targets predicts that offset mature products from

Table 1. Top 20 enriched GO terms in the developmental process category

Species	Enriched GO term	# genes[1]	q-value[2]
Drosophila	multicellular organismal development	121	0.0000
	nervous system development	67	0.0000
	central nervous system development	25	0.0000
	sensory organ development	37	0.0000
	anatomical structure morphogenesis	91	0.0000
	organ morphogenesis	48	0.0000
	neurogenesis	54	0.0000
	cell differentiation	84	0.0000
	neuron differentiation	46	0.0000
	developmental process	129	0.0000
	cell fate commitment	31	0.0000
	organ development	79	0.0000
	generation of neurons	53	0.0000
	system development	106	0.0000
	anatomical structure development	123	0.0000
	cellular developmental process	85	0.0000
	brain development	16	0.0011
	eye development	30	0.0022
	neuron development	38	0.0023
	regionalization	37	0.0023
Caenorhabditis	cellular component morphogenesis	38	0.0047
	anatomical structure morphogenesis	112	0.0062
	neurogenesis	22	0.0164
	generation of neurons	22	0.0164
	neuron development	20	0.0165
	cell morphogenesis	22	0.0167
	neuron differentiation	21	0.0167
	muscle structure development	23	0.0187
	muscle organ development	6	0.0191
	nervous system development	22	0.0204
	neuron projection morphogenesis	18	0.0209
	organ morphogenesis	10	0.0226
	axonal fasciculation	11	0.0232
	neuron projection development	18	0.0233
	anatomical structure formation involved in morphogenesis	25	0.0238
	cell projection morphogenesis	19	0.0245
	syncytium formation by plasma membrane fusion	3	0.0334
	syncytium formation	3	0.0334
	cell part morphogenesis	19	0.0347
	neuron recognition	11	0.0356

[1]Number of genes with predicted canonical seed targets (see Methods) annotated to a GO term; [2]q-value is the p-value corrected for a false discovery rate of 0.05 (ref. 46)

the mir-2a-1 and mir-2a-2 loci will target different sites. However, it is well established that sequence complementarity outside the seed motif is important (and perhaps even sometimes sufficient) for target recognition (reviewed in ref. 25). To explore whether offset microRNAs with the same nucleotide sequence may have different targeting properties, we predicted targets with a different tool, miRanda, which places less weight on the microRNA seed and accounts more fully for the hybridization energy between the microRNA and the target.[26] In this particular case, Drosophila miR-2a-1 is predicted to target 553 transcripts, and miR-2a-2 putatively binds to 788, with 368 targeted genes common to both. The overlap of target genes is greater than expected by chance ($p = 0.008$, see Materials and Methods). This suggests that, although the seed shifting between miR-2a-1 and miR-2a-2 may induce functional changes, the two microRNAs likely conserve partially redundant targeting properties.

Discussion

The evolutionary history of microRNA families is characterized by frequent duplications, losses and rearrangements.[7,8,27,28] Here we describe the evolution of the largest conserved insect microRNA family: mir-2. We showed that this family is widely represented in invertebrates, and the copy number and genomic distribution varies greatly between species. Deep-sequencing data reveal that all mir-2 family members produce their dominant mature microRNAs from the 3' arm, whose sequence is highly conserved (**Fig. 1**). Moreover, most mir-2 precursors have the same Dicer cleavage site, thus producing functional mature miR-2 sequences with the same seed region and predicted targets. According to the available deep-sequencing data, most mir-2 loci within the same species produce redundant products. In Drosophila, antisense-mediated inactivation of mir-2 sequences shows that multiple mir-2 loci have similar (if not identical) functions.[29,30]

It is well-established that pairs of protein-coding loci resulting from gene duplication rapidly diverge in their sequence and/or expression pattern, since functional redundancy is generally a transient situation.[10] Duplication has been also proposed as a mechanism of microRNA functional diversification,[14] although there is no direct evidence of this pattern so far. The mir-2 family suggests that microRNA families may tolerate a situation of functional redundancy in the longer term, as multiple almost identical copies are present in each invertebrate genome. One possible explanation is that mir-2 products are required at high levels and local tandem duplications produce a net increase in the expression level. This is supported by a previous observation that increased expression levels are associated with an increase in microRNA copy number.[31] On the other hand, the presence of redundant mir-2 paralogs might reflect essentiality (see discussion in ref. 29). Functional redundancy in clustered paralogous microRNAs has been previously reported, and may simply reflect high turnover and birth-and-death evolutionary dynamics.[8,27,28] These processes will generate clusters of very similar sequences, and account for the copy number differences between different species.[32] The data strongly suggest that mir-2 family evolution is dominated by high turnover and birth-and-death dynamics mostly driven by random drift.

Clustered paralogous microRNAs are evolutionarily constrained since their expression pattern is linked. However, mir-2 family members in the Drosophila genus are located in two clusters and two single loci. This decoupling of their regulatory sequences may have facilitated functional changes. Indeed, we observe that the identical 3' arms of the mir-2a-1 and mir-2a-2 hairpin precursors produce different offset mature sequences, which we call here canonical and non-canonical miR-2 products (**Fig. 1B**). This phenomenon is called "seed shifting," and has been described to induce functional changes between orthologous microRNAs.[15,17] Experiments in Drosophila suggest that mir-2 products are expressed in brain and have (at least partially) redundant functions.[29,30] However, in situ hybridizations show that the three clusters mir-2b-2~mir-2a-2~mir-2a-1, mir-2c~mir-13a~mir-13b-1 and mir-13b-2 have different spatial expression patterns during early development.[33] We suggest that genomic reorganizations breaking the linkage between mir-2 loci in Drosophila triggered a sub-functionalization event.[34] Interestingly, in the flatworm *Schistosoma mansoni* we observe a duplication of the entire ancestral mir-2 cluster (**Fig. 2** and ref. 35). The functional analysis of the mir-2 family in this parasitic species might shed light on the evolutionary dynamics of clustered microRNAs.

Mir-2 loci are highly expressed in adult heads in Drosophila[22] and in neurons in *Caenorhabditis*.[36] We show that the predicted targets of mir-2 microRNAs in both Drosophila and *Caenorhabditis* are significantly enriched for transcripts with neural development functions (**Table 1**). Mir-2 has also been found to be highly expressed in heads of *Bombyx mori*.[37] Antisense-mediated inactivation of mir-2 in Drosophila produces embryos with defects in head and posterior abdominal segments.[30] Mir-2 has been shown to specifically target the pro-apoptotic genes *rpr*, *grim* and *skl*.[29] Strikingly, these three genes are involved in the selective death by apoptosis of neuroblasts during the normal development of the nervous system.[38] By targeting these pro-apoptotic genes, mir-2 can act as an anti-apoptotic factor in neurons. Indeed, the repression of *rpr* and *grim* by ABD-B prevent apoptosis in neural cells.[39] In the light of these data, we speculate that mir-2 microRNAs have a fundamental role in neuron survival during development and adulthood.

Finally, we note that early works associate mir-6 and mir-11 sequences with the mir-2 family because they have identical (or very similar) seed sequences (e.g., ref. 29). However, there is no evidence of an evolutionary relationship between these three families. Moreover, mir-6 and mir-11 have a distinct expression pattern from mir-2, so functional overlap among these families is unlikely.[29,30,33] We strongly encourage the use of the family name mir-2 to represent only mir-2/mir-13 sequences.

In summary, the mir-2 family is an invertebrate-specific family of microRNAs probably involved in neural development and maintenance. The number

and genomic organization of mir-2 loci varies greatly between species, although the function of paralogous microRNAs is most often redundant. The retention of redundant sequences may be facilitated by the co-transcription of clustered microRNAs. In Drosophila, the ancestral mir-2 cluster has split into multiple independent transcripts, decoupling the transcriptional regulation among mir-2 loci. In this species we find evidence of potential functional shifts of some mir-2 family members.

Materials and Methods

We retrieved all mir-2 precursor sequences from miRBase[2] (version 17) and used BLAST[40] (w = 4, r = 2, q = -3) to search for homologous sequences in multiple genomes from NCBI (www.ncbi.nlm.nih.gov/genome): *Drosophila melanogaster*, *D. virilis*, *D. willistoni*. *D. pseudoobscura*, *Aedes aegypti*, *Anopheles gambiae*, *Acyrthosiphon pisum*, *Bombyx mori*, *Apis mellifera*, *Tribolium castaneum*, *Capitella teleta*, *Daphnia pulex*, *Caenorhabditis elegans*, *Gallus gallus*, *Mus musculus* and *Homo sapiens*. We aligned the putative microRNA hairpin sequences with CMfinder[41] (n = 5, m = 30, M = 100), chose the output alignment that best reflects the microRNA hairpin pairing, and manually refined the alignment using RALEE.[42] The consensus sequence of the alignment was built by taking the most abundant base for each column. All columns with more than 60% gaps were excluded. VARNA 3.7[43] was used to visualize the consensus microRNA structure.

Small RNA libraries, with accession numbers GSE7448 (*D. melanogaster*), GSE15169 (*C. elegans*) and GSE26036 (*T. castaneum*), were retrieved from the GEO database (www.ncbi.nlm.nih.gov/geo/). Reads were mapped to the reference genomes using a sequential trimming approach[15] with the SeqTrimMap tool[44] using default parameters, and microRNAs were detected as described previously.[15] Briefly, predicted hairpin structures within the genome with reads mapped to both arms were first extracted. We only further considered high-quality predictions after careful visual inspection.

To identify potential targets of mir-2 sequences, we first extracted 3'UTR sequences from ENSEMBL (v. 62) via Biomart[45] for *D. melanogaster* and *C. elegans*. Only the longest isoforms were considered. We then detected canonical seed targets in these 3'UTRs.[20] Additional target predictions were performed with miRanda,[26] using default parameters. The significance of the observed overlap of targets between dme-miR-2a-1 and dme-miR-2a-2 was assessed as follows: first, we calculate the target list overlap between both products (number of genes with common targets divided by the total number of targeted genes); then, we calculated the target list overlap between 1000 pairs of randomly selected microRNA products in *D. melanogaster*; finally, the associated p-value is estimated as the proportion of random overlap measures equal or greater than the actual overlap value. In order to detect functional categories enriched in genes targeted by mir-2 products, we analyzed the annotation of transcripts with putative mir-2 target sites in Gene Ontology with HTGOMiner.[46] We used a false discovery rate of 0.05, and focused specifically on enriched terms inside the "developmental process" GO class.

Disclosure of Potential Conflicts of Interest

We declare that there are no financial, personal, or professional interests that could be construed to have influenced this work.

Acknowledgments

We thank Karol Nowicki-Osuch for useful comments on this manuscript. This work is funded by the Biotechnology and Biological Sciences Research Council (BB/G011346/1) and the University of Manchester (fellowship to SG-J). KBH is funded by the Wellcome Trust.

Supplemental Materials

Supplemental materials can be found at: www.landesbioscience.com/journals/rnabiology/article/19160

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004; 116:281-97; PMID: 14744438; http://dx.doi.org/10.1016/S0092-8674(04)00045-5
2. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 2011; 39(Database issue):D152-7; PMID:21037258; http://dx.doi.org/10.1093/nar/gkq1027
3. Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, et al. Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 2005; 33:2697-706; PMID:15891114; http://dx.doi.org/10.1093/nar/gki567
4. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 2005; 11:241-7; PMID:15701730; http://dx.doi.org/10.1261/rna.7240905
5. Saini HK, Enright AJ, Griffiths-Jones S. Annotation of mammalian primary microRNAs. BMC Genomics 2008; 9:564; PMID:19038026; http://dx.doi.org/10.1186/1471-2164-9-564
6. Ryazansky SS, Gvozdev VA, Berezikov E. Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila. BMC Genomics 2011; 12:371; PMID: 21771325; http://dx.doi.org/10.1186/1471-2164-12-371
7. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, et al & Students of Bioinformatics Computer Labs 2004 and 2005. The expansion of the metazoan microRNA repertoire. BMC Genomics 2006; 7:25; PMID:16480513; http://dx.doi.org/10.1186/1471-2164-7-25
8. Maher C, Stein L, Ware D. Evolution of Arabidopsis microRNA families through duplication events. Genome Res 2006; 16:510-9; PMID:16520461; http://dx.doi.org/10.1101/gr.4680506
9. Zhang R, Peng Y, Wang W, Su B. Rapid evolution of an X-linked microRNA cluster in primates. Genome Res 2007; 17:612-7; PMID:17416744; http://dx.doi.org/10.1101/gr.6146507
10. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol 2003; 18:292-8; http://dx.doi.org/10.1016/S0169-5347(03)00033-8
11. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001; 294:853-8; PMID: 11679670; http://dx.doi.org/10.1126/science.1064921
12. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 2001; 294:858-62; PMID:11679671; http://dx.doi.org/10.1126/science.1065062
13. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science 2001; 294:862-4; PMID:11679672; http://dx.doi.org/10.1126/science.1065329
14. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res 2007; 17:1850-64; PMID:17989254; http://dx.doi.org/10.1101/gr.6597907
15. Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. Functional shifts in insect microRNA evolution. Genome Biol Evol 2010; 2:686-96; PMID:20817720

16. Campo-Paysaa F, Sémon M, Cameron RA, Peterson KJ, Schubert M. microRNA complements in deuterostomes: origin and evolution of microRNAs. Evol Dev 2011; 13:15-27; PMID:21210939; http://dx.doi.org/10.1111/j.1525-142X.2010.00452.x
17. Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, et al. The deep evolution of metazoan microRNAs. Evol Dev 2009; 11:50-68; PMID:19196333; http://dx.doi.org/10.1111/j.1525-142X.2008.00302.x
18. Chen X, Yu X, Cai Y, Zheng H, Yu D, Liu G, et al. Next-generation small RNA sequencing for microRNAs profiling in the honey bee Apis mellifera. Insect Mol Biol 2010; 19:799-805; PMID:20807255; http://dx.doi.org/10.1111/j.1365-2583.2010.01039.x
19. Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang H, Arrese EL, Soulages JL, et al. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC Genomics 2010; 11:52; PMID:20089182; http://dx.doi.org/10.1186/1471-2164-11-52
20. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009; 136:215-33; PMID:19167326; http://dx.doi.org/10.1016/j.cell.2009.01.002
21. Liu N, Okamura K, Tyler DM, Phillips MD, Chung W-J, Lai EC. The evolution and functional diversification of animal microRNA genes. Cell Res 2008; 18:985-96; PMID:18711447; http://dx.doi.org/10.1038/cr.2008.278
22. Wang X, Liu XS. Systematic curation of miRBase annotation using integrated small RNA high-throughput sequencing data for C. elegans and Drosophila. Front Genet 2011; 2:25; PMID:22303321; http://dx.doi.org/10.3389/fgene.2011.00025
23. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. An endogenous small interfering RNA pathway in Drosophila. Nature 2008; 453:798-802; PMID:18463631; http://dx.doi.org/10.1038/nature07007
24. Marco A, Konikoff C, Karr TL, Kumar S. Relationship between gene co-expression and sharing of transcription factor binding sites in Drosophila melanogaster. Bioinformatics 2009; 25:2473-7; PMID:19633094; http://dx.doi.org/10.1093/bioinformatics/btp462
25. Brodersen P, Voinnet O. Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 2009; 10:141-8; PMID:19145236; http://dx.doi.org/10.1038/nrm2619
26. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol 2003; 5:R1; PMID:14709173; http://dx.doi.org/10.1186/gb-2003-5-1-r1
27. Li J, Liu Y, Dong D, Zhang Z. Evolution of an X-linked primate-specific micro RNA cluster. Mol Biol Evol 2010; 27:671-83; PMID:19933172; http://dx.doi.org/10.1093/molbev/msp284
28. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in Drosophila species. Genome Biol Evol 2010; 2:180-9; PMID:20624724; http://dx.doi.org/10.1093/gbe/evq009
29. Leaman D, Chen PY, Fak J, Yalcin A, Pearce M, Unnerstall U, et al. Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. Cell 2005; 121:1097-108; PMID:15989958; http://dx.doi.org/10.1016/j.cell.2005.04.016
30. Boutla A, Delidakis C, Tabler M. Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in Drosophila and the identification of putative target genes. Nucleic Acids Res 2003; 31:4973-80; PMID:12930946; http://dx.doi.org/10.1093/nar/gkg707
31. Shomron N, Golan D, Hornstein E. An evolutionary perspective of animal microRNAs and their targets. J Biomed Biotechnol 2009; 2009:594738; PMID:19759918; http://dx.doi.org/10.1155/2009/594738
32. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 2005; 39:121-52; PMID:16285855; http://dx.doi.org/10.1146/annurev.genet.39.073003.112240
33. Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc Natl Acad Sci U S A 2005; 102:18017-22; PMID:16330759; http://dx.doi.org/10.1073/pnas.0508823102
34. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 1999; 151:1531-45; PMID:10101175
35. de Souza Gomes M, Muniyappa MK, Carvalho SG, Guerra-Sá R, Spillane C. Genome-wide identification of novel microRNAs and their target genes in the human parasite Schistosoma mansoni. Genomics 2011; 98:96-111; PMID:21640815; http://dx.doi.org/10.1016/j.ygeno.2011.05.007
36. Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJM. Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res 2008; 18:2005-15; PMID:18981266; http://dx.doi.org/10.1101/gr.083055.108
37. Liu S, Gao S, Zhang D, Yin J, Xiang Z, Xia Q. MicroRNAs show diverse and dynamic expression patterns in multiple tissues of Bombyx mori. BMC Genomics 2010; 11:85; PMID:20122259; http://dx.doi.org/10.1186/1471-2164-11-85
38. Tan Y, Yamada-Mabuchi M, Arya R, St Pierre S, Tang W, Tosa M, et al. Coordinated expression of cell death genes regulates neuroblast apoptosis. Development 2011; 138:2197-206; PMID:21558369; http://dx.doi.org/10.1242/dev.058826
39. Miguel-Aliaga I, Thor S. Segment-specific prevention of pioneer neuron apoptosis by cell-autonomous, postmitotic Hox gene activity. Development 2004; 131:6093-105; PMID:15537690; http://dx.doi.org/10.1242/dev.01521
40. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:9254694; http://dx.doi.org/10.1093/nar/25.17.3389
41. Yao Z, Weinberg Z, Ruzzo WL. CMfinder–a covariance model based RNA motif finding algorithm. Bioinformatics 2006; 22:445-52; PMID:16357030; http://dx.doi.org/10.1093/bioinformatics/btk008
42. Griffiths-Jones S. RALEE–RNA ALignment editor in Emacs. Bioinformatics 2005; 21:257-9; PMID:15377506; http://dx.doi.org/10.1093/bioinformatics/bth489
43. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 2009; 25:1974-5; PMID:19398448; http://dx.doi.org/10.1093/bioinformatics/btp250
44. Marco A, Griffiths-Jones S. Detection of microRNAs in color-space. Bioinformatics 2012; 28:318-23; PMID:22171334; http://dx.doi.org/10.1093/bioinformatics/btr686
45. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011; 2011:bar030; PMID:21785142; http://dx.doi.org/10.1093/database/bar030
46. Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, et al. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). BMC Bioinformatics 2005; 6:168; PMID:15998470; http://dx.doi.org/10.1186/1471-2105-6-168
47. Hedges SB, Kumar S. The Timetree of life. Oxford University Press, 2009.