

Evolutionary transitions of microRNA-target pairs

Masafumi Nozawa^{1,2}, Mai Fujimi¹, Chie Iwamoto¹, Kanako Onizuka¹, Nana Fukuda¹, Kazuho Ikeo^{1,2}, and Takashi Gojobori^{1,3}

¹ Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

² Department of Genetics, SOKENDAI, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

³ Computational Bioscience Research Center, Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, KAUST, 4700 KAUST, Thuwal 23955-6900, Kingdom of Saudi Arabia

* Author for Correspondence: Masafumi Nozawa, Department of Biological Sciences, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan. Telephone, +81-42-677-2576. E-mail, manozawa@tmu.ac.jp

Supplementary Table 1. Summary of the mRNA sequencing data from control flies and transgenic *D. melanogaster* flies constitutively overexpressing one of the miRNAs.

Sample	Stage	Sex	Sample name	Platform	Layout/length (bp)	Accession	# reads
Control	Larva	Female	Cont_LF1	HiSeq 2000	Single end/100	DRR016398	18,802,225
			Cont_LF2	HiSeq 2000	Single end/100	DRR016399	25,277,267
		Male	Cont_LM1	HiSeq 2000	Single end/100	DRR016400	15,457,767
			Cont_LM2	HiSeq 2000	Single end/100	DRR016401	14,833,021
	Pupa	Female	Cont_PF1	HiSeq 2000	Single end/100	DRR016402	12,641,403
			Cont_PF2	HiSeq 2000	Single end/100	DRR016403	11,758,934
		Male	Cont_PM1	HiSeq 2000	Single end/100	DRR016404	12,420,302
			Cont_PM2	HiSeq 2000	Single end/100	DRR016405	12,276,513
	Adult	Female	Cont_AF1	HiSeq 2000	Single end/100	DRR016406	11,521,949
			Cont_AF2	HiSeq 2000	Single end/100	DRR016407	10,188,945
		Male	Cont_AM1	HiSeq 2000	Single end/100	DRR016408	12,349,184
			Cont_AM2	HiSeq 2000	Single end/100	DRR016409	10,301,890
miR277OX*	Larva	Female	miR277_LF1	HiSeq 2000	Single end/100	DRR016410	19,285,423
			miR277_LF2	HiSeq 2000	Single end/100	DRR016411	21,024,063
		Male	miR277_LM1	HiSeq 2000	Single end/100	DRR016412	19,922,513
			miR277_LM2	HiSeq 2000	Single end/100	DRR016413	19,467,702
	Pupa	Female	miR277_PF1	HiSeq 2000	Single end/100	DRR016414	11,534,542
			miR277_PF2	HiSeq 2000	Single end/100	DRR016415	12,418,420
		Male	miR277_PM1	HiSeq 2000	Single end/100	DRR016416	9,895,908
			miR277_PM2	HiSeq 2000	Single end/100	DRR016417	12,601,402
	Adult	Female	miR277_AF1	HiSeq 2000	Single end/100	DRR016418	11,016,414
			miR277_AF2	HiSeq 2000	Single end/100	DRR016419	10,749,504
		Male	miR277_AM1	HiSeq 2000	Single end/100	DRR016420	11,524,337
			miR277_AM2	HiSeq 2000	Single end/100	DRR016421	12,662,696

* OX: overexpression.

Supplementary Table 1. Continued.

Sample	Stage	Sex	Sample name	Platform	Layout/length (bp)	Accession	# reads	
miR982OX	Larva	Female	miR982_LF1	HiSeq 2000	Single end/100	DRR016422	19,783,957	
			miR982_LF2	HiSeq 2000	Single end/100	DRR016423	24,520,488	
	Pupa	Male	miR982_LM1	HiSeq 2000	Single end/100	DRR016424	21,317,084	
			miR982_LM2	HiSeq 2000	Single end/100	DRR016425	20,587,166	
		Female	miR982_PF1	HiSeq 2000	Single end/100	DRR016426	10,875,298	
			miR982_PF2	HiSeq 2000	Single end/100	DRR016427	11,695,660	
		Male	miR982_PM1	HiSeq 2000	Single end/100	DRR016428	10,971,395	
			miR982_PM2	HiSeq 2000	Single end/100	DRR016429	11,654,960	
	Adult	Female	miR982_AF1	HiSeq 2000	Single end/100	DRR016430	10,285,177	
			miR982_AF2	HiSeq 2000	Single end/100	DRR016431	14,094,257	
		Male	miR982_AM1	HiSeq 2000	Single end/100	DRR016432	12,025,403	
			miR982_AM2	HiSeq 2000	Single end/100	DRR016433	11,311,838	
	miR954OX	Larva	Female	miR954_LF1	HiSeq 2000	Single end/100	DRR016434	16,773,132
				miR954_LF2	HiSeq 2000	Single end/100	DRR016435	17,484,476
Male			miR954_LM1	HiSeq 2000	Single end/100	DRR016436	17,923,003	
			miR954_LM2	HiSeq 2000	Single end/100	DRR016437	19,378,113	
Pupa		Female	miR954_PF1	HiSeq 2000	Single end/100	DRR016438	10,396,404	
			miR954_PF2	HiSeq 2000	Single end/100	DRR016439	10,359,912	
		Male	miR954_PM1	HiSeq 2000	Single end/100	DRR016440	11,897,207	
			miR954_PM2	HiSeq 2000	Single end/100	DRR016441	12,498,677	
Adult		Female	miR954_AF1	HiSeq 2000	Single end/100	DRR016442	10,847,577	
			miR954_AF2	HiSeq 2000	Single end/100	DRR016443	17,952,596	
		Male	miR954_AM1	HiSeq 2000	Single end/100	DRR016444	11,363,395	
			miR954_AM2	HiSeq 2000	Single end/100	DRR016445	13,936,028	

Supplementary Table 2. The expression levels of miR277, miR982, and miR954 in *Drosophila* species, as determined using small RNA sequencing data.

Species	Tissue/stage	Accession	RPM*		
			miR277	miR982	miR954
<i>D. melanogaster</i>	Larva	SRR013600	5,163.37	62.71	229.95
	Larva	SRR013602	5,860.11	20.78	166.24
	Pupa	SRR013601	5,450.35	61.03	173.70
	Pupa	SRR013603	5,854.54	0.00	353.85
	Female adult	SRR018039	60,835.12	6.15	86.11
	Male adult	SRR016854	111,574.01	97.81	143.09
<i>D. simulans</i>	Embryo	SRR032434	0.00	0.00	0.00
	Ovary	SRR618934	0.00	8.58	0.00
	Larva	SRR1275487	0.00	0.00	0.00
	Pupa	SRR127483	0.00	68.81	0.00
	Pupa	SRR127485	0.00	0.00	0.00
	Thorax, abdomen	SRR1205790	0.00	32.97	0.00
	Female larva	DRR018570 [†]	0.00	0.00	0.00
	Male larva	DRR018571 [†]	0.00	0.00	0.00
	Female pupa	DRR018572 [†]	0.00	0.00	0.00
	Male pupa	DRR018573 [†]	0.00	0.00	0.00
	Female adult	DRR018574 [†]	0.00	0.00	0.00
	Male adult	DRR018575 [†]	0.00	78.05	0.00
	<i>D. sechellia</i>	Thorax, abdomen	SRR1205791	20,302.43	0.00
<i>D. yakuba</i>	Larva	SRR1275488	421.76	0.00	-
	Pupa	SRR1275484	70.77	0.00	-
	Pupa	SRR1275486	1,013.17	0.00	-
	Thorax, abdomen	SRR1205792	21,857.09	0.66	-
	Female larva	DRR018576 [†]	685.17	0.00	-
	Male larva	DRR018577 [†]	1,235.53	0.00	-
	Female pupa	DRR018578 [†]	596.41	0.00	-
	Male pupa	DRR018579 [†]	690.55	0.00	-
	Female adult	DRR018580 [†]	12,714.84	0.00	-
	Male adult	DRR018581 [†]	38,618.08	0.00	-
	<i>D. erecta</i>	Thorax, abdomen	SRR1205793	60,784.94	-
<i>D. pseudoobscura</i>	Embryo	SRR032435	2377.42	-	-
	Head	SRR032436	39,443.89	-	-
<i>D. virilis</i>	Larva	SRR1106727	375.22	-	-
	Larva	SRR1106728	313.77	-	-
	Adult	SRR1106729	1,241.07	-	-

* Number of reads per million mapped reads.

[†] Data were generated in this study.

Supplementary Table 3. The numbers of target genes of miR277, miR982, and miR954 identified using the TCC package (Sun et al. 2013) in combination with the PITA prediction software (Kertesz et al. 2007).

miRNA		Number of target genes				Total
		Larva	Pupa	Adult	Stage specific*	
miR277	Female	113	40	6	149	153
	Male	82	122	50	193	219
	Sex specific [†]	99	100	46	194 [‡]	200
	Total	147	131	51	253	286
miR982	Female	77	7	5	85	87
	Male	39	9	39	85	86
	Sex specific [†]	56	16	40	103	103
	Total	86	16	42	132	138
miR954	Female	5	8	8	19	20
	Male	7	11	15	31	32
	Sex specific	6	17	19	40	40
	Total	9	18	21	44	46

To narrow down the candidate target genes, only 7-mer target sites without any mismatches or 8-mer target sites with ≤ 1 mismatch were considered. Wobble pairing was not allowed.

* The number of target genes that were regulated by the miRNA at only one of the developmental stages examined.

[†] The number of target genes that were regulated by the miRNA in only one of the sexes.

[‡] The number of target genes that were regulated by the miRNA at only one developmental stage in one sex.

Supplementary Table 4. The numbers of target genes of miR277, miR982, and miR954 identified using the TCC package (Sun et al. 2013) in combination with the TargetScan prediction software (Ruby et al. 2007).

miRNA		Number of target genes				
		Larva	Pupa	Adult	Stage specific*	Total
miR277	Female	34	32	4	61	65
	Male	35	43	29	66	84
	Sex specific [†]	35	33	29	60 [‡]	63
	Total	52	54	31	86	106
miR982	Female	6	0	0	6	6
	Male	4	1	6	9	10
	Sex specific [†]	6	1	6	12	12
	Total	8	1	6	13	14
miR954	Female	0	0	1	1	1
	Male	1	1	1	3	3
	Sex specific	1	1	0	2	2
	Total	1	1	1	3	3

To narrow down the candidate target genes, TargetScan software was used with the default settings. Only 3'UTRs of transcripts were considered.

* The number of target genes that were regulated by the miRNA at only one of the developmental stages examined.

[†] The number of target genes that were regulated by the miRNA in only one of the sexes.

[‡] The number of target genes that were regulated by the miRNA at only one developmental stage in one sex.

Supplementary Table 5. Summary of the small RNA sequencing data for analyzing the levels and breadths of miRNA expression in *D. melanogaster*.

Stage	Tissue	Accession
Embryo (0-1 hr)	Whole body (WB)	SRR069840
Embryo (2-6 hrs)	WB	GSM286605, GSM286606
Embryo (6-10 hrs)	WB	SRR069838, SRR069839
Embryo (14-24 hrs)	WB	SRR069262
Larva (1st instar)	WB	GSM360256, GSM360257
Larva (3rd instar)	WB	SRR013600, SRR013602
	Imaginal disc	SRR069832
Pupa (0-2 days)	WB	GSM360262
Pupa (2-4 days)	WB	SRR013601, SRR013603
Adult	Female head	SRR068998, SRR013604
	Female body	SRR018039, SRR069837
	Ovary	SRR069232
	Male head	SRR069835, GSM322543
	Male body	SRR069836, SRR016854
	Testis	GSM280085

Supplementary Table 6. Enriched GO terms in the target genes of miR277.

GO term*	Description	FDR (<i>q</i> -value)	Enrichment	Category
GO:0006573	Valine metabolic process	1.95E-03	75.58	Process
GO:0009081	Branched-chain amino acid metabolic process	2.03E-07	64.78	Process
GO:0006551	Leucine metabolic process	4.93E-03	56.68	Process
GO:0046395	Carboxylic acid catabolic process	4.90E-03	9.12	Process
GO:0016054	Organic acid catabolic process	5.31E-03	9.12	Process
GO:0006631	Fatty acid metabolic process	1.47E-04	9.00	Process
GO:0044282	Small molecule catabolic process	3.39E-03	8.17	Process
GO:1901605	Alpha-amino acid metabolic process	2.52E-03	7.39	Process
GO:0032787	Monocarboxylic acid metabolic process	6.90E-03	5.36	Process
GO:0005811	Lipid particle	3.17E-03	4.79	Component
GO:0019752	Carboxylic acid metabolic process	4.10E-07	4.26	Process
GO:0006082	Organic acid metabolic process	4.28E-07	4.12	Process
GO:0043436	Oxoacid metabolic process	5.71E-07	4.12	Process
GO:0044255	Cellular lipid metabolic process	9.32E-03	3.96	Process
GO:0006520	Cellular amino acid metabolic process	2.44E-03	3.88	Process
GO:0044281	Small molecule metabolic process	7.17E-06	3.04	Process
GO:0044710	Single-organism metabolic process	6.64E-03	1.84	Process
GO:0003824	Catalytic activity	1.38E-04	1.62	Function

* GO terms with false discovery rate (FDR or *q*-values) < 0.01 are shown.

Supplementary Table 7. Summary of the small RNA and mRNA sequencing data for predicting the target genes of each miRNA in *D. melanogaster*.

Stage	Tissue	mRNA-seq accession	Small RNA-seq accession
Embryo	Whole body (WB)	SRR030237	SRR069838, SRR069839
Larva (1st instar)	WB	SRR030242, SRR030243	GSM360256, GSM360257
Larva (3rd instar)	WB	SRR030244	SRR013600, SRR013602
Pupa	WB	SRR030246	SRR013601, SRR013603, GSM360260, GSM360262
Adult	Female head	SRR039433-SRR039445	SRR068998, SRR013604
	Ovary	SRR032310-SRR032313	SRR069232
	Male head	SRR039436-SRR039438, SRR039446-SRR039452	SRR069835, GSM322543
	Testis	SRR032304-SRR032307	GSM280085

Supplementary Table 8. Consistency between bioinformatics predictions and experimental data.

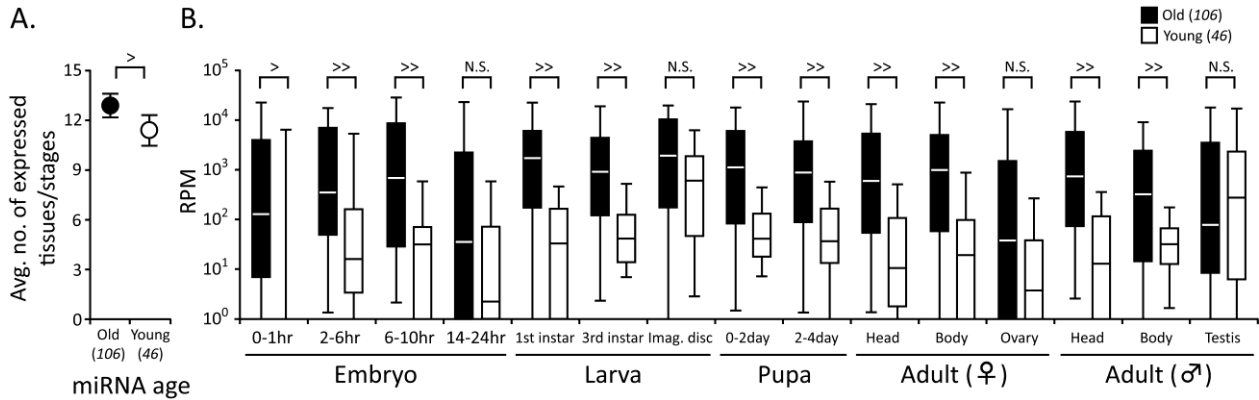
Prediction tool	Region considered	Rate of concordance		
		miR277	miR982	miR954
miRanda	Whole transcript	12.4% (415/4,140)	5.2% (138/2,649)	1.9% (99/5,130)
PITA	Whole transcript	10.1% (741/7,351)	5.3% (397/7,539)	2.0% (141/6,992)
TargetScan	3' UTR	12.0% (194/1,619)	8.5% (24/282)	2.1% (7/335)

The denominators in parentheses represent the total number of genes predicted by the bioinformatics tools (miRanda, PITA, or TargetScan), whereas the numerators indicate the number of genes that were detected as differentially expressed (DEGs) among the predicted genes. miRanda and TargetScan were used with the default settings, whereas only 7-mer target sites without any mismatches or 8-mer target sites with ≤ 1 mismatch were considered in the PITA software.

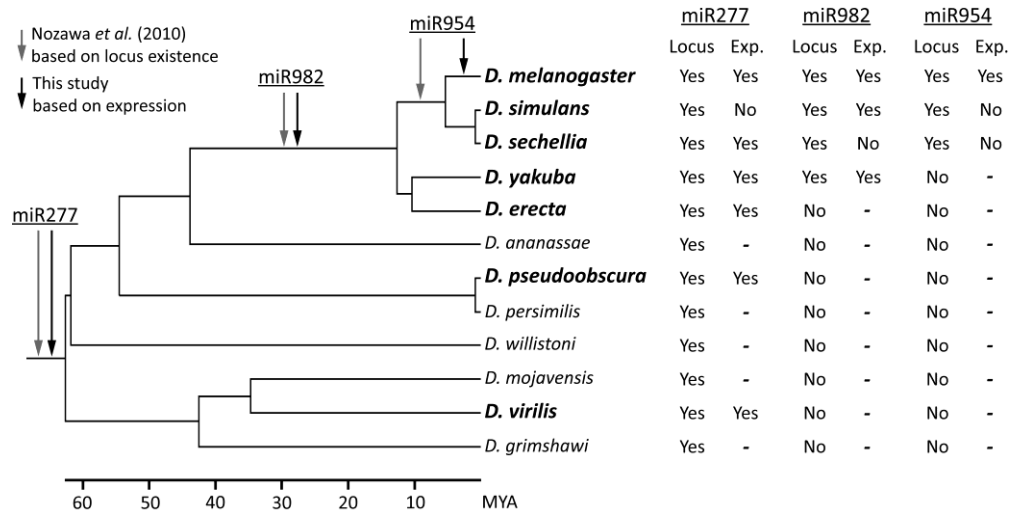
Supplementary Table 9. Difference in the expression level of miR277, miR982, and miR954 between the transgenic and wild-type *Drosophila melanogaster* flies.

miRNA	Expression level (RPM)		
	Wild type	Overexpression	Difference (OX-wild type)
<i>Larva</i>			
miR277	5511.7	131127.1	125615.4
miR982	41.7	283259.8	283218.1
miR954	198.1	350929.3	350731.2
<i>Pupa</i>			
miR277	5652.4	304036.6	298384.2
miR982	30.5	66103.6	66073.1
miR954	263.8	98337.5	98073.7
<i>Adult</i>			
miR277	86204.6	109674.7	23470.1
miR982	52.0	460704.0	460652.0
miR954	114.6	108461.0	108346.5

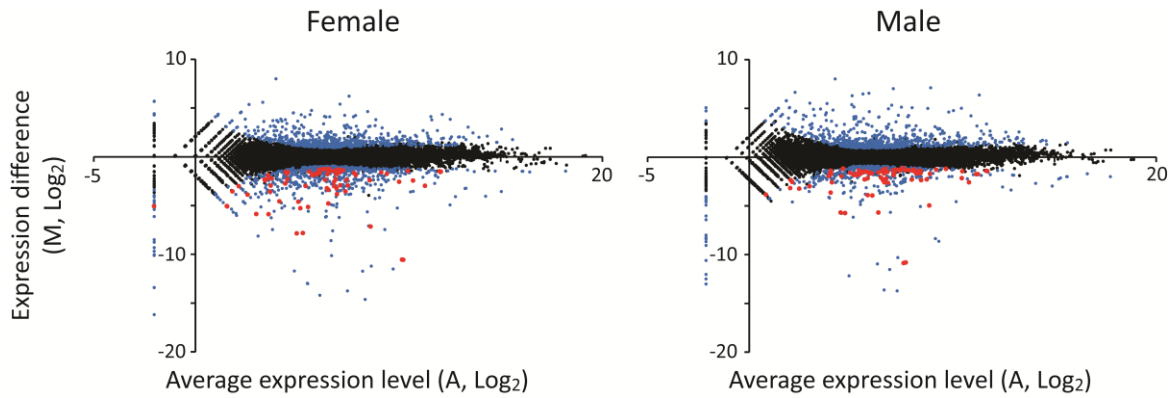
RPM (the number of reads per million mapped reads) in wild-type flies was the average of RPMs for each developmental stage in Supplementary Table 5. RPM in overexpression flies was estimated based on the RPM in wild-type flies and the fold changes in the expression level of a miRNA due to overexpression estimated by qPCR (see Supplementary Fig. 5).



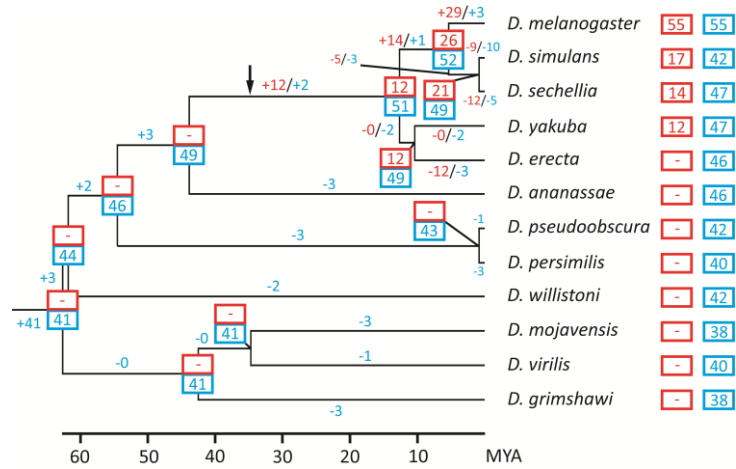
Supplementary Fig. 1. Relationships between the age of miRNAs and their tissue distribution (A) or expression level (B). “Old” and “young” are miRNA genes that existed before the radiation of *Drosophila* species analyzed and were generated in the lineage to *D. melanogaster* after the radiation, respectively. The numbers in parentheses indicate the number of miRNA genes in each group. (A) The error bars represent the 95% confidence intervals based on the 1,000 bootstrap resampling. The threshold of miRNA expression was set to be Reads Per Million mapped reads or RPM of 1. (B) The lines in the boxes represent the median; 50% of the values are included in the boxes, and 80% of the values are included within the bars. Statistical significance was based on Monte Carlo simulation with 1,000 replications. >>, $P < 0.001$; >, $P < 0.01$; N.S., not significant ($P \geq 0.05$).



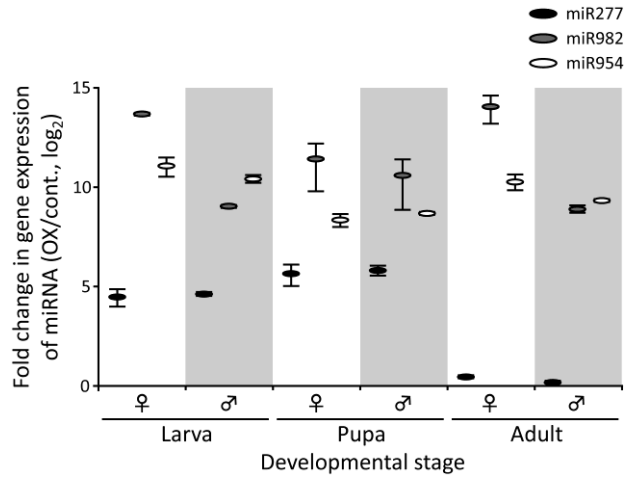
Supplementary Fig. 2. Origins of miR277, miR982, and miR954 in *Drosophila* species. The gray arrows indicate the origins of the miRNAs based on the study by Nozawa *et al.* (2010), in which the existence of the miRNA loci was used to determine their origins under the parsimony principle with the assumption of single origin of each miRNA. The black arrows represent the origins of the miRNAs estimated by the presense of miRNA expression as well as the existence of the locus. In this study, we used the latter criterion that is more stringent.



Supplementary Fig. 3. An example MA-plot showing the differences between the expression levels of genes in a larval stage control line and a larval stage line overexpressing miR277. The x-axis shows the average expression level of each transcript in both lines, and the y-axis shows the difference between the expression levels in each line. Negative values on the y-axis indicate that the expression level of a transcript was lower in the overexpression line than the control line. Each dot represents a transcript; the red dots indicate target transcripts regulated by the miRNA (miR277 in this example), and the blue dots indicate other differentially expressed transcripts (DETs). DETs were detected using the TCC package (Sun et al. 2013) with a cutoff threshold of $q = 0.01$. Target sites were predicted by miRanda (Enright et al. 2003).



Supplementary Fig. 4. Evolutionary changes in the number of miR982 target genes in *Drosophila*. The blue boxes show the numbers of orthologs of the genes identified as direct targets of miR982 in *D. melanogaster*. The red boxes show the numbers of orthologs with miR982 target sites at each node. The numbers along each branch indicate the gains (+) and losses (-) of these orthologs during evolution. The arrow indicates the origin of miR982.



Supplementary Fig. 5. Log₂ fold changes in the expression level of a miRNA due to overexpression of its own exogenous miRNA. Filled, gray, and open circles represent the fold changes in gene expression of miR277, miR982, and miR954, respectively. Error bars show standard errors. The expression levels of each miRNA in the flies overexpressing either miR277, miR982, or miR954 and the control flies were measured by using qPCR with the comparative threshold cycle method (Scheffe et al. 2006).

REFERENCES

- Enright AJ et al. 2003. MicroRNA targets in *Drosophila*. *Genome Biol* 5:R1.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet* 39:1278-1284.
- Nozawa M, Miura S, Nei M. 2010. Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biol Evol* 2:180-189.
- Ruby JG et al. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* 17:1850-1864.
- Scheffe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H. 2006. Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. *J Mol Med* 84:901-910.
- Sun J, Nishiyama T, Shimizu K, Kadota K. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* 14:219.