Supplemental Text and Methods for

Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus

Jaaved Mohammed, Alex S. Flynt, Alexandra M. Panzarino, Md Mosharrof Hussein Mondal, Matthew DeCruz, Adam Siepel and Eric C. Lai

Recovery of *D. melanogaster* miRNAs from simulated libraries

We simulated four libraries each representing 25 million randomly sampled reads from the pooled collection of *D. melanogaster* sRNAs from male bodies, female bodies, mixed embryos, and heads respectively. We created 100 samples of 100M reads each (25M X 4 libraries), and determined the recoverability of conserved and newly-evolved *D. melanogaster* miRNAs at varying minimum mature and star read expression thresholds. These artificial libraries, allowed us to investigate the amount of miRNAs at two different age groups (i.e. conserved and newly-evolved) that could be recovered in the other 11 *Drosophila* genomes. We defined conserved *D. melanogaster* miRNAs as those with unambiguous orthologs in the obscura-group species, *D. willistoni*, or the *Drosophila*-group species. Recently-evolved *D. melanogaster* miRNAs were defined as those with orthologs within the melanogaster-group species only.

False negative rate and false positive rate for miRNA discovery

We discussed the theoretical false negative rate in the subsampling description in the initial results (Figure 1A-B, Supplementary Figure S1). To examine how this played out in our experimental analysis, we address our recovery of miRBase miRNAs, which include the full set of miRBase loci that had never previously been cloned, and were predicted only on the basis of homology. Note that we exclude 47 annotations Drosophilid miRBase loci that we do not currently believe to be products of RNase III-biogenesis (i.e., not likely to be miRNAs, Supplemental Table S3).

If we disregard 13 exact duplicate paralog loci in miRBase (highlighted below in green), we missed only 38/1411 *Drosophila* species entries in miRBase (i.e. miRBase annotations we classified as "candidate" from our cloning data), for a false-negative rate of only 2.7%. However, nearly half of these are in *D. melanogaster*, and are very lowly-expressed, evolutionarily young, miRNAs. Although we do not demote these in our annotation pipeline (as with the 47 other loci in Table S3), we note they fall below current

annotation criteria and would not have been annotated de novo in this study. If we discount those loci, there is a false negative rate of only 23/1396 or 1.6%.

Dmel	15	dme-mir-2280 dme-mir-2281 dme-mir-2490 dme-mir-2495 dme-mir-4908 dme-mir-4911 dme-mir-4939 dme-mir-4946 dme-mir-4953 dme-mir-4956 dme-mir-4958 dme-mir-4964 dme-mir-4971 dme-mir-4975 dme-mir-4978
Dsim	3	dsi-mir-2578 dsi-mir-4966a dsi-mir-983b
Dsec	3	dse-mir-4966c dse-mir-4966d dse-mir-9684
Dere	1	der-mir-983b
Dpse	12 (green= exact duplicates)	dps-mir-2505 dps-mir-2524-1 dps-mir-2524-2 dps-mir-2525 dps-mir-2545a-2 dps-mir-2545a-3 dps-mir-2545a-4 dps-mir-2545a-5 dps-mir-2545a-6 dps-mir-2545b dps-mir-2574b dps-mir-307b
Dvir	2	dvi-mir-9706b-2 dvi-mir-9715
Dgri	15 (green= exact duplicates)	dgr-mir-285 dgr-mir-286-2 dgr-mir-3-1 dgr-mir-3-2 dgr-mir-309-1

		dgr-mir-309-2 dgr-mir-4-1 dgr-mir-4-2 dgr-mir-5-2 dgr-mir-6-1 dgr-mir-6-2 dgr-mir-6-3 dgr-mir-6-4 dgr-mir-6-5 dgr-mir-6-6

Regarding false positives, our annotation pipeline is proposed to exceed minimum confidence for substrates that have passed through a Dicer-dependent biogenesis pathway (either Drosha-Dicer or splicing-Dicer pathway). This does not mean that they are necessarily efficient substrates; indeed, we cannot know from these studies whether a majority or a small minority of input transcript is actually converted into small RNAs. Presumably, many evolutionarily recently-emerged miRNAs may not be as efficiently processed as conserved miRNAs. However, in our annotation efforts we looked for evidence of specific biogenesis based on specificity of small RNA read patterns and 3' overhangs from duplexes. Our annotations exceed minimum confidence, meaning that loci with features close to the borderline were set into the "candidate" category and were not used for further analysis. Moreover, we demoted 47 annotations from Drosophilid miRBase loci, which are usually taken as the "gold standard". Thus, we have taken substantial efforts to minimize false positives from our annotations.

Additional considerations for miRNA gene annotation

All miRNAs predicted from our miRNA identification pipeline were vetted manually and bioinformatically for miRNA and mirtron candidacy. In the manual phase, all miRDeep2 predictions (Friedlander et al. 2012), and intron and hairpin structures with $p > 0.5$ were examined for evidence of cleavage by Drosha and Dicer based on the sRNA read alignment, a hairpin secondary structure, and synteny with other miRNA predictions. Putative canonical miRNA were further classified bioinformatically using criteria based on (1) expression, (2) clonability of Drosha/Dicer products, such as the miR, miR*, loop, or 5’ or 3’ moR sequences, (3) structure pairing of the miR:miR* duplex, (4) 5' end consistency of miR and miR* reads, and (6) ratio of background to miRNA reads (**Supplemental Fig. S3**). Canonical miRNA or mirtron predictions that met all

criteria were labeled as “confident” while those that failed some or all criteria were labeled as “candidate” or “FALSE,” respectively (see **Supplementary Fig. S3** for exact criteria thresholds). “Candidate” loci that were orthologous to “confident” annotations were re-classified as “candidate-rescued.” Confidence classifications for all miRNA and mirtrons are provided in **Supplemental Table S4**. Finally, novel “confident”, “candidate-rescued”, and “candidate” miRNAs and mirtrons were segregated from miRBase annotation and their unannotated orthologs (i.e. known *Drosophila* miRNAs) in order to identify novel loci specific to this study. Mirtrons were classified using the same features as for canonical miRNAs except for Drosha cleavage, and an additional criterion for untemplated modifications of the 3’ arm reads was specified (**Supplementary Fig. S3**).

Identification of miRNA clusters and testes-restricted miRNAs

Although the boundaries of polycistronic transcripts in each species are unknown, we defined miRNA clusters based on genomic proximity within up to a 10kb window and expression similarity across the available libraries in a given species. Mirtrons were excluded from this classification. The strong majority of miRNA clusters identified in this study comprised genes with testes-restricted expression. Testes-restricted miRNAs were characterized as genes with >4-fold RPMM testis or male-body expression enrichment when compared against all other tissue and developmental-timepoint libraries. If >75% of miRNA genes within a cluster were classified as testes-restricted, then all genes within said cluster were labeled canonical, **T**estes-restricted, **R**ecently-evolved, **C**lustered miRNAs.

Birth and Death Model

To assess birth and death rate variation across classes of miRNAs and across *Drosophila* clades of interest, we designed and implemented a phylogenetic probabilistic graphical model. This model permits estimation of parameters of gene birth (λ) and death (μ) (**Fig. 6A**) based upon our assignments of miRNA presence and absence in each species per miRNA family alignment. We note that small RNAs were sampled more deeply in certain species, especially in *D. melanogaster* (**Fig. 1**). However, besides *D. melanogaster*, there is not generally a correlation between sampling depth and the number of confidently annotated miRNAs per species. This is due in part to the "rescue" approach (**Fig. 3**). Therefore, we chose to apply our estimates of miRNA flux using our full collection of annotations, rather than by attempting to make a new set of annotations

by subsampling a lower, fixed number of reads from across the species, since this would inevitably decrease annotation confidence.

Parameter estimation required two sets of precomputed data. The first datum needed was a binary encoding of miRNA presence (1) and absence (0) as leaf node labels of the phylogenetic model. In this regard, we labeled non-miRNAs and “candidate” miRNAs as cases of absences, and “confident” or “candidate-rescued” miRNAs as cases of presences. We assigned each miRNA for which no orthologs were identified to its own singleton miRNA alignment, to be counted as an independent birth event. The second datum needed was phylogenetic branch-length estimates for the 12 *Drosophila* species phylogeny (Clark et al. 2007). We estimated branch lengths (τ) in units of substitutions per site, using fourfold degenerate sites (i.e. sites within a codon in which all four possible nucleotide substitutions are synonymous) and the maximum-likelihood program RaXML (Stamatakis 2014). Fourfold degenerate sites were extracted from a *de novo* 12 *Drosophila* species whole genome alignment constructed using the LASTZ and MULTIZ programs and the chaining and netting protocol used for the UCSC Genome Browser. The resulting maximum-likelihood, newick format tree was: ((((((dm3:0.055153,(droSim1:0.027716,droSec1:0.023941):0.024430):0.050893,(droYak2:0.090814,droEre2:0.079010):0.032754):0.328435,droAna3:0.466508):0.162763,(dp4:0.018457,droPer1:0.018684):0.407262):0.120336,droWil1:0.593093):0.118858,((droVir3:0.244781,droMoj3:0.335567):0.082788,droGri2:0.319783):0.118858).

This method is implemented as a Java software package and available at http://compgen.cshl.edu/mirna/12flies/software/MirnaTreeML.zip.

Analysis of obscura-group sequence divergence and polymorphism data

To identify unambiguous sequence divergences between species, we focused on miRNAs with clear 1-to-1 orthologs, such as the miRNAs within the 3’ sub-cluster region of the *obscura*-subgroup *Dpse_3416→ Dpse-mir-2536* cluster (**Fig. 7B**). We utilized seed sequence identity to place miRNAs into seed identical families, in order to identify all homologs of a specific family, and used pre-miRNA locus positioning within the cluster to segregate true miRNA orthologs from paralogs.

To investigate sequence polymorphisms within a population of *D. pseudoobscura* individuals, we downloaded whole genome sequencing data from http://pseudobase.biology.duke.edu/. This data provided polymorphisms from 11 North American *D. pseudoobscura* strains and 2 *D. pseudoobscura bogotana* sub-species

(McGaugh et al. 2012). Pre-miRNA alignments of the reference sequence for *D. pseudoobscura* and *D. persimilis* species and the corresponding *D. pseudoobscura* individual sequences permitted the visualization of sequence divergences and polymorphisms, respectively.

Construction of UAS-DsRed-miRNA expression vectors

Pri-miRNA fragments, including ~200bp upstream and downstream of the pre-miRNA hairpins, were amplified from genomic DNA using the primers below and cloned into the 3' UTR position of pUAS-DsRed (Brennecke et al. 2005) using the Seamless Ligation Cloning Extract (SLiCE) method for directional, homology arm dependent cloning (Zhang et al. 2014). Note that the *cwo* (373_164) constructs contain both miRNAs in single pri-miRNA fragments.

Primers for miRNA expression constructs

longmel_373_164F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCAGAGTTGTCAAGCGTCAGTAAGATC
TC

longmel_373_164R

TCACAAAGATCCTCTAGAGGTACCCTCGAGTTCGCAGCTTCTATCTTGTATCTCCTG

longpse_373_164F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCTGTGTGTGTGCGAGAGAGTGTAC

longpse_373_164R

CAAAGATCCTCTAGAGGTACCCTCGAGGGAGAAAGATATATCATACTGGCAGTTTA
TTAG

longvir_373_164F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCCGCAGATGCCGGATAACTGATTTAC

longvir_373_164R

TCACAAAGATCCTCTAGAGGTACCCTCGAGCGCACATCCTCGATCATTTGCG

mel4984F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCCTGGTTATGCAGCTGCGACAAG

mel4984R

TCACAAAGATCCTCTAGAGGTACCCTCGAGGCCATGGCATTTGTAGAAGAAGCG

sim4984F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCGGAATCTGGTTATGCAGCTGCG

sim4984R

TCACAAAGATCCTCTAGAGGTACCCTCGAGTGCCATGGCATTTGTAGAAGAAGC

yak4984

GCCGCCACCACCTGTTCCTGTAGCGGCCGCACGGACGCATCAGTATCTGGAC

yak4984

TCACAAAGATCCTCTAGAGGTACCCTCGAGCCACACTCTAATTCGGAATCGG

per2484F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCCTGTTGTTGACTGCAAAGTAAGATGT
GAG

per2484R

TCACAAAGATCCTCTAGAGGTACCCTCGAGGACTTCTCAATTGGGAATTCGTCGTG

pse41F

GCCGCCACCACCTGTTCCTGTAGCGGCCGCTCTAATTTACGTTCGTGCTATGTAGT
CCC

pse41R

TCACAAAGATCCTCTAGAGGTACCCTCGAGGGAGTACTCACATTTGATCTAAGAGA
CACAAG

Northern blot assays

S2 cells were transfected, following manufacturer protocols, with 1.7 μg of the UAS-dsRed-miRNA constructs along with 0.8 μg of a second plasmid encoding GAL4 under control of a ubiquitin promotor (Ub-Gal4) using TransitIT ® Insect from Mirus Bio. After a 48 hour incubation RNA was harvested using Tri-Reagent ® from Fisher Scientific. Northern blotting was carried out as previously described (Okamura et al.

2007) using P32 labeled DNA oligonucleotide probes. A single probe was used to detect *D. melanogaster, simulans* and *yakuba* mir-4984 orthologs which are identical, while an equal mixture of two pse-41/per2484 probes was used due to the single nucleotide difference between their mature products.

Northern Probes

pse41 probe TTGCCAGGCTAGAGCTGTAATA

per2484 probe TTGCCAGGCTAGAGCTGTGATA

probe4984 CCAGCGAATACGTCAAAGAATT

probe373 CAGCTGCCTCGTCGTGCAACA

probe164 AACGTAGTCAGCAAACAAACA

Luciferase sensor assays

Luciferase sensors containing only seed match pairing to test miRNAs were generated by ligating annealed oligonucleotides into a modified psiCheck dual luciferase vector (Okamura et al. 2007). Each insert encoded 4 miRNA binding sites, which were identical in the mel-4984 sensor. Due to a single nucleotide difference in the seed of the pse-41/per2484 homologs, two of the sites were targets of the *D. pseudoobscura* version, and the other two target sites were for the *D. persimilis* version. Luciferase assays were performed as 4 replicates where each well was transfected with 50ng Ub-Gal4 (200ng), 50ng psiCheck sensor, 100ng UAS-dsRed-miRNA, using TransitIT ® Insect from Mirus Bio–following manufacturer protocols. Dual luciferase assays were performed with a Dual-Glo ® kit from Promega. A Synergy H1® biotek plate reader was used to measure luciferase activity. Statistical significance was determined by ANOVA and Tukey HSD using R.

Primers for generating psiCheck sensors

miR-4984 DMEL/SIM/YAK sense

GGCCTCAAAGAATTTACTTCAAAGAATTCGAGTCAAAGAATTGTCCTCAAAGAATT

miR-4984 DMEL/SIM/YAK antisense

TCGAAATTCTTTGAGGACAATTCTTTGACTCGAATTCTTTGAAGTAAATTCTTTGA

pse41/per2484 DPSE/PER sense

GGCCAGCTGTGATAGGACAGCTGTAATAAGATAGCTGTGATAAGTAAGCTGTAATA

pse41/per2484 DPSE/PER antisense

TCGATATTACAGCTTACTTATCACAGCTATCTTATTACAGCTGTCCTATCACAGCT

Supplemental References

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of MicroRNA-Target Recognition. *PLoS biology* **3**: e85.

Clark AG Eisen MB Smith DR Bergman CM Oliver B Markow TA Kaufman TC Kellis M Gelbart W Iyer VN et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203-218.

Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40**: 37-52.

McGaugh SE, Heil CS, Manzano-Winkler B, Loewe L, Goldstein S, Himmel TL, Noor MA. 2012. Recombination modulates how selection affects linked sites in Drosophila. *PLoS biology* **10**: e1001422.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130**: 89-100.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Zhang Y, Werling U, Edelmann W. 2014. Seamless Ligation Cloning Extract (SLiCE) cloning method. *Methods in molecular biology* **1116**: 235-244.

Zhu C, Byrd RH, Lu P, Nocedal J. 1997. L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software* **23**: 550-560.

Inventory of Supplemental Figures and Tables

Supplemental Figure S1: Conserved and newly-evolved *D. melanogaster* miRNAs recovered at varied read depth thresholds using *in silico* simulated libraries. These libraries are composed of randomly sampled reads across all *D. melanogaster* sRNA-seq male-body, female-body, head, and mixed embryo libraries used within this study. miRNA recovery rates are computed per read-depth sample at various miR or miR* read thresholds. Error bars depict the standard error of the recovery rate across 100 simulations.

Supplemental Figure S2: Read length distribution for all small RNA libraries sequenced in this study. We extended our previous broad and deep analysis of *D. melanogaster* by sampling 11 additional *Drosophila* species as listed to the right, by analyzing mixed embryos, adult heads, male bodies and female bodies; a testis library was also generated for *D. simulans*. A subset of libraries were sequenced in replicates, especially ones where the expected dominant miRNA-sized peak (21-22 nt) peak was not initially observed. A piRNA peak is seen in most of the body libraries. Due to the technical difficulty in culturing *D. grimshawi*, it was only feasible to generate two libraries for male and female bodies.

Supplemental Figure S3: Detailed flow-chart of miRNA and mirtron identification pipeline and scoring criteria.

Supplemental Figure S4: Drosophilid miRbase loci demoted for lack of compelling small RNA evidence. Examples of 47 demoted miRBase (v21) miRNAs and mirtrons. 37 of these represent piRNAs within *D. pseudoobscura* and *D. virilis* previously classified as miRNAs.

Supplemental Figure S5: Total miRNA and mirtron annotation count within 12 *Drosophila* species. Annotations are further subdivided within (1) three confidence categories- “confident”, “candidate-rescued, and candidate”, and (2) between known and novel annotations. Note that “candidate” annotations were not utilized for analyses of miRNA flux in this study.

Supplemental Figure S6: Distribution of miRNAs for three classes of miRNAs within each *Drosophila* species. These classes are defined by biogenesis pathway and canonical miRNAs are further divided by their testes-restricted expression. Only "confident" and "candidate-rescued" loci are included; loci that are considered "candidate" only and lack further rationale to be rescued based on a confidently processed miRNA ortholog are not included in these pie charts.

Supplemental Figure S7: Read alignments for *mir-10404*, a conserved non-canonical miRNA generated from the ITS1 spacer in ribosomal RNA, across the Drosophilid phylogeny.

Supplemental Figure S8: Other well-conserved miRNAs identified within this study. Alignment and representative hairpin structure for each conserved miRNA. Included is *dme_474* which is not well-conserved but is one of the Drosha-cleaved hairpins within the pasha 5' UTR. Note that unlike most other conserved miRNAs, these loci generally accumulate modest amounts of small RNAs and/or have atypical structural features. This might reflect that their processing is atypical and/or regulated, or that their conservation reflects a role other than, or in addition to miRNA-type function. For example, besides the pasha 5' UTR hairpins, two of these loci are located in CDS or 3'UTR, and thus cleavage could mediate host mRNA downregulation.

Supplemental Figure S9: Alignment and small RNA read details for novel conserved miRNAs annotated in this study.

Supplemental Figure S10: Evolutionary patterns of 5' end cleavage precision for all conserved *D. melanogaster* miRNAs.

Supplemental Figure S11: Additional examples of novel sense/antisense miRNA pairs. (A) Example of novel sense/antisense miRNA pair from *D. willistoni* (*dwi_62/dwi-98*). (B) Example of novel sense/antisense miRNA pair from *D. ananassae* (*dan_100/dan_244*).

Supplemental Figure S12: Annotation of novel Testes-restricted, recently-evolved, clustered (TRC) miRNA clusters identified within *D. ananassae*.

Supplemental Figure S13: Annotation of a novel Testes-restricted, recently-evolved, clustered (TRC) miRNA cluster identified in *D. willistoni*. Tissue code indicates the miRNAs are all highest expressed in male-body libraries.

Supplemental Figure S14: Annotation of novel Testes-restricted, recently-evolved, clustered (TRC) miRNA clusters identified in *obscura-subgroup* species, *D. pseudoobscura* and *D. persimilis*. (A) Genomic organization and small RNA read density of orthologous *dps_3416 → dps-mir-2536* TRC clusters in *D. pseudoobscura* and *D. persimilis*. (B) Genomic organization and small RNA read density of orthologous *dps-mir-2510 → dps_23* TRC clusters in *D. pseudoobscura* and *D. persimilis*. Tissue code indicates the miRNAs are all highest expressed in male-body/testis libraries. Note that there are also additional copies of some of these TRC loci located outside of these clusters.

Supplemental Figure S15: Annotation of novel Testes-restricted, recently-evolved, clustered (TRC) miRNA clusters identified in virilis clade species. (A) The *dvi_66 → dvi_40* cluster. This cluster has a clear homologs in mojavensis. The small RNA mappings to their respective genomic loci are shown. (B) The *dvi_24637 → dvi_197* cluster has clear homologs in D. mojavensis. The small RNA mappings to their respective genomic loci are shown. (C) Annotation of three additional novel Testes-restricted, recently-evolved, clustered (TRC) miRNA clusters identified in virilis clade species. The *dvi_43 → dvi_207* cluster has two copies in *D. virilis* (i.e. roughly similar members can be found on scaffold_12723 and scaffold_12963). The other cluster is *D. mojavensis dmo_62* and *dmo_330*.

Supplemental Figure S16: Expression difference between *D. pseudoobscura*-specific or *obscura-subgroup*-specific TRC and solo canonical miRNAs. Points reflect the maximum expression per locus assessed over all *D. pseudoobscura* libraries. P-value computed from two-tailed Wilcoxon Rank Sum Test.

Supplemental Figure S17: All possible phylogenetic reconstruction of ancestral miRNA presence and absence for 3 miRNA classes using a phylogenetic probabilistic graphical model with universal parameters of $\lambda = 0.292$ and $\mu = 0.694$. These parameters were computed by running the phylogenetic reconstruction algorithms on all mirtrons and

miRNAs pooled together. These trees illustrate how the method's maximum likelihood reconstruction performs for all possible configurations of extant miRNAs presence and absence per alignment. Blue text indicates count of alignments with this particular configuration in each class. Summary of miRNA birth and death (Figure 6) are based upon these estimates of ancestral miRNA presence and absence.

Supplemental Figure S18: Individual phylogeny of extant and inferred ancestral miRNA presence and absence for all canonical miRNAs that are not in the testis-restricted clustered subclass.

Supplemental Figure S19: Individual phylogeny of extant and inferred ancestral miRNA presence and absence for all mirtrons.

Supplemental Figure S20: Individual phylogeny of extant and inferred ancestral miRNA presence and absence for all testis-restricted clustered (TRC) canonical miRNAs.

Supplemental Figure S21: Examples of mirtrons with atypical emergence and decay patterns. Expression profiles and mirtron alignments are shown per example to highlight the non-clade specificity of mirtron presence.

Supplementary Figure S22. Evolution of *mir-4984* processing and function across related melanogaster subgroup species. (A) *mir-4984* hairpin is similar across 5 related melanogaster subgroup species and its mature (green) arm is identical in all these species. However, small RNA sequencing indicates substantial accumulation only in *Dmel*, very modest in *Dsim/Dyak*, and not in *Dyak/Dere*. (B) Experimental tests of UAS-DsRed-mir-4984 expression constructs transfected into S2 cells shows that only the Dmel construct was effectively processed into miRNAs. (C) DsRed expression confirms that all constructs were expressed. (D) Luciferase sensor assays of the three orthologs of mir-4984 onto a common multimer sensor contain 4 miR-4984 seed matches. Only Dmel-mir-4984 conferred specific repression of the sensor.

Supplemental Figure S23. Differential expression of miRNAs between *obscura-subgroup* species. Scatterplot depicting the correlation of miRNA expression of all *D. pseudoobscura* and *D. persimilis* ortholog pairs (RPMM = Reads Per Million Mapped

MiRNA Reads). All miRNA alignments with orthologs in both species are shown. Points that lie on or near the diagonal represent similarly expressed ortholog pairs. Orthologs with > 6-fold RPMM difference (denoted by the blue-dashed line and labeled points) are examples of significantly differentially expressed orthologs. Points are colored by miRNA age, and shapes represent miRNAs with or without miR:miR* duplex region substitutions (fraction of duplex sites with substitutions are labeled). Note that the *mir-309* cluster (yellow) loci are expected to be expressed in the very early embryo, and given that the embryo development and timing were not controlled in library preparation, their differential accumulation may not be genuine. Amongst loci changed by >6-fold, *dme-mir-2b-1* and *dme-mir-310* are deeply conserved, but all others are specific to the *obscura-subgroup* species.

Supplemental Figure S24: 3' end untemplated nucleotide additions for canonical miRNAs and mirtrons in 12 *Drosophila* species. (A) Proportion of AG ending 3' arm miRNAs and mirtrons that contain reads within mono-A, C, G or U untemplated additions. Error bars represent the standard error of the mean. More mirtrons contain untemplated uridylation than comparable 3' end AG-ending canonical miRNAs. (B) Species-specific empirical cumulative distribution function of mono-uridylation for mirtrons and canonical miRNAs with 3' end 'G' nucleotide or non-'G' nucleotides (i.e. IUPAC ambiguity code 'H'). P-value computed from two-tailed Wilcoxon Rank Sum Test between canonical 3'-end 'H' miRNAs and mirtrons. Significant differences in mono-uridylation distributions between these two classes are noted in blue text. P-values from comparisons between canonical 3' end 'H' miRNAs and 3' end 'G' miRNAs are all non-significant and not shown.

Supplemental Table S1: Small RNA libraries for 11 *Drosophila* species created for and analyzed in this study.

Supplemental Table S2: Small RNA libraries for *Drosophila* species acquired from public repositories and analyzed in this study. (**A**) Libraries for *D. melanogaster*. (**B**) Libraries for *D. simulans, D. yakuba, D. erecta, D. pseudoobscura*, and *D. virilis*.

Supplemental Table S3: miRBase Drosophilid loci that were demoted from miRNA status.

Supplemental Table S4: Master list of all known and novel miRNAs in *Drosophila.*

Supplemental Table S5: List of all sense and antisense miRNA pairs identified and analyzed in this study.

Supplemental Table S6: 3' Untemplated nucleotide addition counts to canonical miRNA-3p and mirtron-3p species.

Supplemental Table S7: miRNA sRNA read, sequence, and structure features for Random Forest classifier.