# Supplementary Note: A model of miRNA gene birth and death.

**Two classes of miRNA genes – the background class and the putative functional class.**

The excesses of new miRNA genes on external branches of Fig. 1, especially the short ones, most likely represent evolutionarily transient miRNA genes that decay rapidly. While death of miRNAs may be common, there indeed exist many miRNA genes that are "immortal", being shared between nematodes, insects and vertebrates [1,2]. We therefore consider new miRNA genes as comprising two distinct classes - a putative functional class and a background class. These two classes may have very different birth and death rates. Obviously, it is the putative functional class that is biologically relevant.

Birth of a new miRNA (putative functional or not) is defined by its fixation in the population, not by the occurrence of the mutation. Nevertheless, like any molecular evolution model, fixation rate (or birth rate) is the same as mutation rate if the mutations are neutral. After new miRNA genes have been fixed in the population, they face the pressure of decay due to subsequent accumulation of mutations. Random mutations on a hairpin structure are known to have a strong tendency to destabilize it [3]. We shall formulate the rate of decay as a survival function (i.e., the probability that a new miRNA survives to time *t*), which has to be estimated separately for the background miRNA genes and the putative functional ones. Under mutation pressure, background miRNA genes are expected to be relatively short-lived. In comparison, putative functional new miRNA genes may be considerably longer-living due to selection against degenerative mutations. The survivorship of putative functional miRNAs may thus be age-dependent.

In the miRNA birth and death model, we assumed that the birth rates of miRNA genes are constant at rate m genes per Myrs. Of the miRNA genes originating at any timepoint *t* ($0 \leq t \leq t6$, see Fig. 1 of text), a fraction of $f$ is "background miRNA genes" and the remaining fraction, 1-*f*, is putative functional miRNA genes. For each class of miRNA genes, the death rate is modeled by a survival function, i.e., the probability that a new miRNA gene survives t Myrs is $S(t) = f \cdot B(t) + (1 - f) \cdot F(t)$, where $B(t)$ is the probability that a new background miRNA gene

survives *t* Myrs and $F(t)$ is the probability that a new putative functional miRNA gene survives *t* Myrs.

**Maximum likelihood estimation of *f* and other parameters under the birth and death model**

Since relatively accurate species divergence time is only available within the *Drosophila* genus, we only consider the miRNA genes generating on the branch b1-b5 (Fig.1 ), and use *D. virilis* as outgroup. Any miRNA gene originating at branch b1-b5 would thus yield all or a subset of the 32 ($2^5$) possible phylogenetic patterns among *D. melanogaster, D. simulans, D. yakuba, D. ananassae,* and *D. pseudoobscura*. Given any survival function $S(t)$, we can calculate the probabilities of each of the 32 phylogenetic patterns. In Supplementary Table 3, we listed the survival functions of all the 32 phylogenetic patterns produced by miRNA genes generating on each of the five branch segment $S_{i,j}(t)$, where *j* (*j*=1,2,3,4,5) denotes the branch and *i* (*i*=1-32) denotes the *i*th phylogenetic pattern . For example, the total probability of the phylogenetic pattern 10000 is $P_1 = \sum_{j=1}^{5} \int_{t_j}^{t_{j+1}} S_{1,j}(t)dt$ . In Supplementary Table 4, we listed the detailed statistic formulas for all the 32 patterns. In general, the probability for the *i*th phylogenetic pattern is thus $P_i = \sum_{j=1}^{5} \int_{t_j}^{t_{j+1}} S_{i,j}(t)dt$ .Since the background hairpin approach starts from the hairpins identified in the genome of *D. melanogaster*, only 16 out of the 32 possible phylogenetic patterns are available for the background hairpin dataset (Table 2). Note in this dataset hairpins are present in *D. melanogaster* for all the 16 patterns. So we calculate the scaled probabilities of the 16 patterns with $P_i' = P_i / \sum_{j=1}^{16} P_i$ ( *i* = 1-16). The maximum likelihood value is thus $mle' = \sum_{i=1}^{16} \ln(P_i') \cdot k_i$ , where $k_i$ is the observed number of the *i*th phylogenetic pattern.

The objectives of this model are i) to estimate the birth rate of new miRNA genes and the relative proportion of the background class and putative functional; ii) to infer the death rates (or survival functions) of the two classes separately. The maximum likelihood estimation procedure consists of two steps. In Step I, we estimate the survival function of background

miRNAs from the genomic hairpin data of Table 2. In Step II, using the survival function of the background miRNAs as input, we estimate the survival function of the putative functional class from the observations of read-associated hairpins of Supplementary Table 2. The relative proportion of the two classes (background vs. putative functional) is estimated at the same time. The birth rate can be easily extracted when the rest has been estimated. The detailed estimation procedure is as the following:

In Step I, we used a gamma distribution to approximate the probability density of death at time $t$ for background miRNA genes, i.e., $B(t) = Gamma(\lambda_1, \alpha_1, t)$, where $\lambda_1$ denotes rate and $\alpha_1$ denotes shape parameters of a gamma distribution. In estimating survival function of the background miRNA genes, we set $f$=1, so we get $S(t) = B(t)$. Since *mle'* is only a function of $\lambda_1$ and $\alpha_1$, by iterating numerically, we can estimate the best fit $\lambda_1$ and $\alpha_1$ when *mle'* is the maximum.

In Step II, we use a similar procedure to estimate the survivorship of the putative functional miRNA genes. The observed miRNA genes of Table 2 are a mixture of the emerging functional class and the background class. We assume the proportions are $(1-f)$: $f$. We now assume that the death rate of the emerging functional class also follows a gamma distribution with different values for the $\lambda_2$ and $\alpha_2$ parameters. So we get $S(t) = f \cdot B(t) + (1-f) \cdot F(t) = f \cdot Gamma(\lambda_1, \alpha_1, t) + (1-f) \cdot Gamma(\lambda_2, \alpha_2, t)$. Since $\alpha_1$ and $\lambda_1$ have been estimated from the background miRNA genes, The maximum likelihood value is only a function of $\lambda_2$, $\alpha_2$ and $f$. Based on the probabilities of observing $Ai$ miRNA genes with phylogenetic pattern $i$, as presented in Table 2, we can also numerically estimate the best fit $\lambda_2$, $\alpha_2$ and $f$ using similar approach.

**The birth rate of miRNA genes.**

There are 42 *D. simulans* specific miRNAs using the high-stringency criteria. Because some of the miRNAs emerged on branch b2-b5 could have the phylogenetic pattern (100000) as well, the fraction of the 42 miRNAs that emerged in the *D. simulans* lineage since Node 1 is estimated by

$\int_{t1}^{t2} S_{1,j}(t)dt / \sum_{j=1}^{5} \int_{t_j}^{t_{j+1}} S_{1,j}(t)dt$ as given in Supplementary Table 4, which would be 0.7043 under the estimation in Fig. 2. Multiplying 42 with this fraction, we estimate that 30.3 new miRNAs emerged in the last 5 million years and have remained observable in the extant *D. simulans*. We then estimate the number of miRNAs that emerged in the last 5 Myrs, including those that have degenerated. Integrating the composite survival function of Fig. 2 from 0 to time *T* divided by *T*, we obtained a ratio of 0.5024 for $T$ = 5 Myrs. In other words, among those miRNAs that emerged in the last 5 Myrs, 50.2% are still observable now. Thus, the total number of miRNAs emerged in that period is 60.3 (30.3/0.5024). In other words, about 12 new miRNA genes originated every Myrs (60.3/5Myrs).

**Supplementary Methods.**

**The construction of small RNA libraries.**

The protocol for cloning miRNA[4] was modified and used in this study. Total RNA from male heads of *D. melanogaster, D. simulans* and *D. pseudoobscura* was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions, except that the 70% ethanol wash step was omitted. One thousand μg of total RNA was run on a 15% denaturing polyacrylamide gel. Upon visualization with SYBR Gold Nucleic Acid gel stain (molecular Probes), the gel slice ranging between 16 to 28 bases was excised. RNA was eluted with 0.3 M NaCl at 4°C overnight. The elution was treated with phenol/chloroform and the supernatant was transferred to a new tube. RNA was precipitated by adding 1 μl glycogen (20mg/ml) (Roche) and 2.5X volume ethanol and keeping at -20°C for more than 2 hours. RNA was collected by centrifugation at 12,000 rpm for 10 min. A pre-adenynated 3' linker (5'-rAppCTGTAGGCACCATCAAT/3ddC/ -3', IDT Inc.) was ligated to RNA 3' end with T4 RNA ligase (NEB) at 37°C for 40 min in 1X ligation buffer without ATP (50 mM Hepes pH 8.3, 10 mM MgCl2, 3.1 mM DTT, 10 ug/ml BSA, 8.3% glycerol). The ligation product was run on a 15% denaturing polyacrylamide gel and recovered from the gel again using the same procedure described above. A 5' RNA linker (5'-UCGUAGGCACCUGAAA-3') was ligated to RNA 5' end with T4 RNA ligase in 1 X ligation buffer with ATP (NEB) at 37°C for 40 min. The second ligation product was treated with

phenol/chloroform and precipitated by ethanol. Reverse transcription reaction was performed by using RT primer (5'-ATTGATGGTGCCTACA-3') and Superscript II (invitrogen) for 30 min at 42 °C. To increase cDNA production, the sample was denatured at 95°C for 5 min and cooled on ice. One µl superscript II (invitrogen) was added and the mixture was incubated at 42°C for another 30 min. The cDNA was run on a 12% denaturing polyacrylamide gel and the gel slice ranging between 90 to 102 bases was excised. cDNA was eluted from the gel, precipitated by ethanol and amplified by PCR using a fused primer pair (5'-N20a TCGTAGGCACCTGAAA-3' and 5'-N20bATTGATGGTGCCTACAGT-3'; N20a and N20b, 454 proprietary 20-nt primers). The PCR product was run on a 2% agarose gel. The gel slice containing the amplified DNA ranging between 90 to 102 bp was excised. DNA was purified by using Qiagen gel extraction kit. The purified DNA was used for 454 sequence collection (454 Life Sciences Inc.)

**Genome mapping of the small RNA reads.**

The genome sequences of *D. melanogaster* (dm2), *D. simulans* (droSim1) and *D. pseudoobscura* (dp4) were downloaded from UCSC Genome Bioinformatics Site (genome.ucsc.edu). After trimming the adaptor sequences, the small RNA reads (18-28 nt in length) were mapped on the corresponding genome sequences using BLAT[5]. Only reads that can be perfectly mapped on the genome sequences were kept for further analyses. In *D. melanogaster*, reads perfectly matching the characterized non-coding RNAs (rRNAs, tRNAs, snoRNA, miRNAs and snRNAs) and Transposable elements (FLYBASE R5.1) were excluded in new miRNA identification. In *D. simulans* and *D. pseudoobscura*, we BLAST [6] the read sequences (with flanking 70 nt at each side) against the ncRNAs and transposable element library of *D. melanogaster*. Fragments with BLAST E value $\leq 10^{-3}$ or masked by REPEATMASKER[7] were discarded.

**Searching new miRNA genes.**

For each miRNA gene, we defined the short and long arms based on the position of the mature miRNA in the miRNA precursor. The sequence flanking the mature miRNA and containing miRNA* was refered to “long arm” and the sequence flanking the mature miRNA and does not contain miRNA* was refered to “short arm”. By analyzing 2,636 characterized miRNA genes from 13 metazoan species in miRbase (V9.2), we found the length of the short arms is 12.6 nt,

with s.d. of 7.4, and none of the short arms is longer than 70 nt. The mean length of the long arms is 52.36 nt, and the s.d. is 8.5. Only 81 out of the 2,636 (~ 3%) long arms is greater than 70 nt. Based on this length survey, we took 70 nt at each side of the read to search new miRNA loci and folded the whole sequence using RNAFOLD [8] program. The substructures that meet with the three levels of stringency criteria were extracted for futhre analyses. However, some miRNA genes with long precursors might be missed by this size cutoff. So we also extended the side length to 100 nt and 150 nt and performed the same analyses. (The floatchart and three complementary figures were provided in our website http://pondside.uchicago.edu/wulab/microRNA .)

**Phylogenetic patterns of the miRNA genes.**

To obtain the orthologous alignments of each miRNA locus in six *Drosophila* species (*D. melanogaster, D. simulans, D. yakuba, D. ananassae, D. pseudoobscura,* and *D. virilis*), we used the following strategies. First, all miRNA sequences were BLAST against the other 5 genomic sequences to get the reciprocal best alignments (E value cutoff $10^{-5}$). Second, for each locus, the homologous sequences of the aforementioned species were extracted from the 15-way insect MULTIZ genome alignments (genome.ucsc.edu). Finally, the results of BLAST search and MULTIZ alignment were pooled based on sequence identity and syntenic information. In each species, the paralogous loci with BLAST E value $\leq 10^{-5}$ were combined and the one with the most stable hairpin structure was used. The orthologous sequences were then examined for miRNA-like features. In addition, the orthologous sequences were also required to have a secondary structure similar to the corresponding read-producing genes (similarity score greater than 0.3 by RNAFORESTER [9]).

**Background hairpins and type ii data.**

To identify the background hairpins in the whole genome of *D. melanogaster,* we divided the whole genome sequences into sliding windows with a window size of 150 nt and step size of 50 nt. The windows matching the known non-coding RNAs, transposable elements and putatively identified new miRNAs were discarded. For each of the remaining window, the sequence was folded using RNAFOLD program in both the sense and antisense directions. The substructures

meeting with the low stringency criteria were then extracted. The orthologous sequences of *D. simulans, D. yakuba, D. ananssae , D. pseudoobscura,* and *D. virilis* were retrieved from the MULTIZ alignments and were then subjected to the three levels of stringency criteria.

There are about 110,000 background hairpins meeting with the high stringency criteria in *D. melanogaster*. Out of these hairpins, 1,861 have read sequences matching with reads collected from Ref.[10], and more than 95% of them have read only observed once (The total read number is over 2 millions). Those sets of 1,861 miRNA-like hairpins comprise type ii data in the main text. However, these hairpins with read expression did not show significantly different phylogenetic patterns from the rest background hairpins (Pearson's correlation coefficient $r=0.969$, $P<0.0001$). In our website http://pondside.uchicago.edu/wulab/microRNA, we listed the expression level of the miRNA-like genes with reads collected in both this study and Ref. [10]; we also presented the expression level of the 1,861 miRNA-like genes with reads collected only from Ref. [10] but not by this study.

**The settings of computer programs.**

**infile** is a file with one or multiple sequences with the FASTA format; **outfile** is the output file after running the computer programs. **genomedb** is the database file processed by the FORMATDB program. The following are the settings of the computer programs:

BLASTALL -i infile -d genomedb -o outfile -p blastn -e 1e-5 -W 7 -b 3 -v 3 -q -1 -F F
RNAFOLD -d0 < infile > outfile
RANDFOLD -d infile 1000 > outfile
REPEATMASKER –s -species drosophila infile
RNASHAPES -f infile -q -t 5 -F 0.001 -M 30 > outfile
RNAFORESTER –f infile --xml -2d -m -mc=-10 -cmin=0.5 -bm=1 -bd=-10 -br=0 -pm=10 -pd=-5 > outfile

**Duplication of newly identified miRNA genes.**

Using the 119 newly identified miRNA genes of *D. simulans* (with the high stringency criteria), we BLAST against the *D. simulans* genome and found 104 of them are singletons in the genome with an E value $\leq 10^{-6}$. Each of the remaining 15 genes (*miR-2003, miR-2005, miR-2009, miR-2014, miR-2018, miR-2021, miR-2025, miR-2029, miR-2033, miR-2034,miR-2037, miR-2052,*

*miR-2094, miR-2116* and *miR-2144*) have two paralogous copies in the genome that can form hairpin structures meeting with the high stringency criteria. Out of the 15 gene families, 12 families have the same mature miRNAs. Among the remaining 3 families, mature miRNA differ by 1 nt on paralogous copies *miR-2021* and *miR-2052*, and by 2 nt on paralogous copies of *miR-2116*. 14 out of the 15 duplication events are only restricted in *D. simulans*.

Using the 66 newly identified miRNA genes of *D. melanogaster* (with the high stringency criteria), we BLAST against the *D.melanogaster* genome and only found 2 genes (*miR-2003* and *miR-2583*) have paralogous copies in the *D. melanogaster* genome. *miR-2003* has 20 copies in *D. melanogaster* and all the 20 copies share the same mature miRNA; *miR-2583* has two copies in *D. melanogaster* and the two copies also share the same mature miRNA. The alignments of these duplicated genes were presented in in our website http://pondside.uchicago.edu/wulab/microRNA.

Thus among the 16 duplicated miRNA gene families identified in *D. melanogaster* and *D. simulans*, 15 are lineage specific and only one (*miR-2003*) have paralogous copies in both species. In this study, we are interested mainly in new miRNA loci that produce novel mature miRNAs which exist as distinct reads in our samples. Gene duplications resulting in identical miRNAs are not considered new miRNAs. By this definition, only 3/(119+66) = 1.7% of the newly identified miRNA genes have different paralogous mature miRNAs, pooling the 119 miRNA genes in *D. simulans* and the 66 in *D. melanogaster*. In other words, only 1.7% of new miRNAs were derived from gene duplication.

In the analysis above, we examine the newly discovered miRNA genes. The rate of gene duplication among the known established miRNAs is even lower. (It may be because the transcriptome is sensitive to the dosage of the established miRNAs.) Out of the 78 characterized miRNA genes in *D. melanogaster*, 28 are duplicated genes (including *miR-310/311/312/313* cluster) and 50 are singletons. Out of the 28 duplicated miRNA genes, 23 can be confirmed to originate before the radiation of *Drosophila* genus. *miR-310/311/312/313* cluster have four copies in *melanogaster* group but only three copies in *D. pseudoobscura* and *D. virilis*, so one

duplication event might occurred before the radiation of the *melanogaster* group. Additionally, across the 12 *Drosophila* genome, we only observed one tandem duplication of the whole *miR-310* cluster specifically in *D. erecta* (the paralogous clusters are located on position 10615282-10615763 and 10616740-10617220 of scaffold_4845, respectively. ) The lineage lengths across the 12 genomes is ~250 myrs, so the duplication rate is 2/250 = 0.008 duplication per Myrs. This rate is much lower than the estimated ~ 12 miRNA genes per Myrs by the de novo method.

**Searching for miRNA genes originated by inverted duplication.**

It has been elegantly demonstrated that inverted duplication of small segments can give birth to new miRNA genes [11-13]. We used two approaches to investigate whether inverted duplication is also a major mechanism for the origin of new miRNA genes identified in this study.

First, we BLAST the hairpin sequences of new miRNA genes of *D. melanogaster* against *D. simulans, D. pseudoobscura* and *D. virilis* genomes. If a miRNA gene originated through the inverted duplication mechanism, we should expect that the left and right arms of the hairpins (i.e., the mature miRNA and miRNA* plus the corresponding stem extension regions) should hit the same genomic location of the three genomes but with inverse directions. Out of the 66 new miRNA genes identified in *D. melanogaster*, none of them originated through this mechanism. We also BLAST the hairpin sequences of new miRNA genes of *D. simulans* against genomes *of D. melanogaster, D. pseudoobscura* and *D. virilis*, and still found none of the new miRNA genes originated through this mechanism.

Secondly, we carefully examined the multiple alignments of these newly identified miRNA loci and the flanking regions and still we failed to identify miRNA genes emerged through the inverted duplication mechanism.

**Simulation of de novo formation of hairpins by point mutations.**

To explore the likelihood of de novo creation of hairpin structures, we randomly selected 800 150-nt long segments from the genome of *D. melanogaster*. After excluding those segments that can have optimal substructures meeting with the low stringency criteria, 680 segments were retrieved for further simulation analyses.

For each segment, in each cycle, we substitute the nucleotide on one randomly selected site using a substitution matrix derived from polymorphisms of the 4-fold degenerate sites of 9,628 genes of *D. simulans* (see our website http://pondside.uchicago.edu/wulab/microRNA) and subject the derived sequence to the miRNA definition criteria (see Text). This process is continued until the structure of the derived sequence meeting with the high stringency criteria of miRNA definition. The number of cycles was counted as the number of substitutions required to form the hairpin structures. If a segment cannot form an optimal substructure meeting with the high stringency criteria after 100 rounds of substitutions, 100 was chosen as the number of substitutions. For each segment, this simulation process was repeated 100 times and the median number of substitutions was recorded. The number of substitutions required to convert a non-hairpin forming sequence into a hairpin meeting with the high stringency criteria is presented in Supplementary Fig. 2. The branch length of the *D. simulans* lineage is 0.05 substitutions per site, or 150 x 0.05 = 7.5 substitutions for a 150 nt segment. In our simulations, 2.34% (16/680) of random segments (that are 150 nt long and not hairpin forming) gave rise to a miRNA-like hairpin with 8 mutations or fewer. Extrapolating this calculation to the entire *Drosophila* genome of 139 Mb, we arrived at a number of nearly 4000 such events every Myrs ($139 \times 10^6$ / 150 x 2.34% /5, assuming the branch length of *D. simulans* lineage is 5 Myrs).

To exclude the possibility that the simulated hairpin birth rate is augmented by the dead hairpins in the genome that can be reactivated with relative few substitutions, we redo the computer simulation in the following way: We permuted each of the 78 known miRNA genes in *D.melanogaster* 100 times. We chose 6600 of the permutated sequences that cannot form hairpin with free energy smaller than -15 Kcal/mol for further analysis. The mean length of the 6600 sequences is 86.8 nt with s.d. of 12.9. For each of the 6600 sequences, we repeated the simulation process as shown above. After correcting the sequence length to 150 nt, among the 6600 sequences, 340 (5.16%) sequences can generate hairpins meeting with the high stringency

criteria with 8 or fewer substitutions (The figure is implemented in our website http://pondside.uchicago.edu/wulab/microRNA). The two sets of simulation results were well converged.

**The survival function of background hairpins modeled by neutral evolution simulation.**

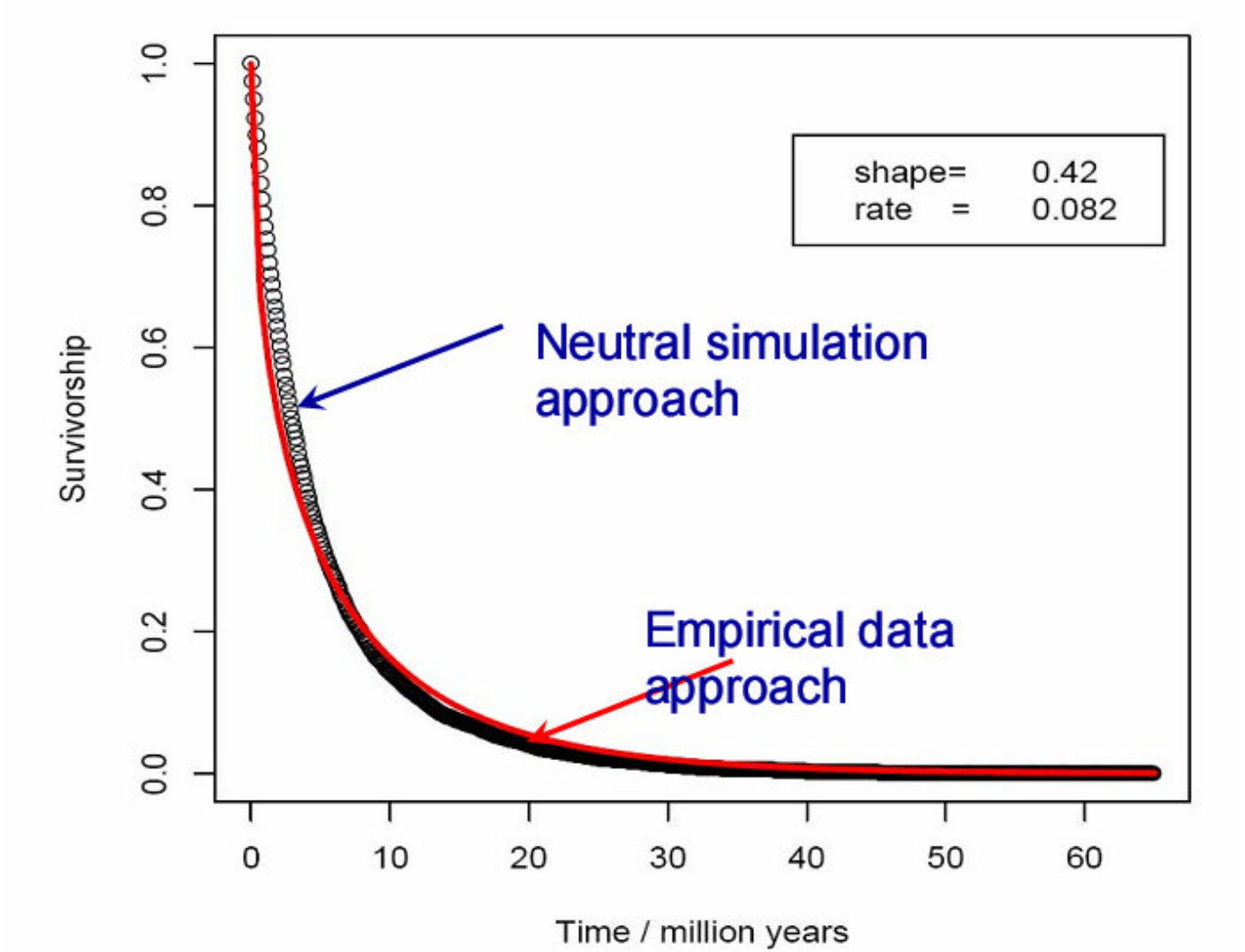
10,000 hairpin sequences meeting with the high stringency criteria were randomly selected from the background hairpin dataset. For each hairpin sequence, in each cycle, we substitute the nucleotide on one randomly selected site using a substitution matrix and examine whether the derived sequences can still meet with the high stringency criteria. The number of mutations was then recorded. Next we convert the number of mutations into the survival time. The mean substitution rates on the genome sequences were assumed as 0.5 mutations per 100 nt per Myr [14]. Let L be the length of the hairpin and S be the number of mutations it survives. We generated S+1 exponential random number with rate = 0.5/100*L and sum them up to simulate the survival time. In Supplementary Fig. 1 we plotted the probability that the sequences that survives longer than t Myrs (Y axis) against t Myrs (X axis).

**MiRNA expression level difference between species is not associated with difference in secondary structures.**

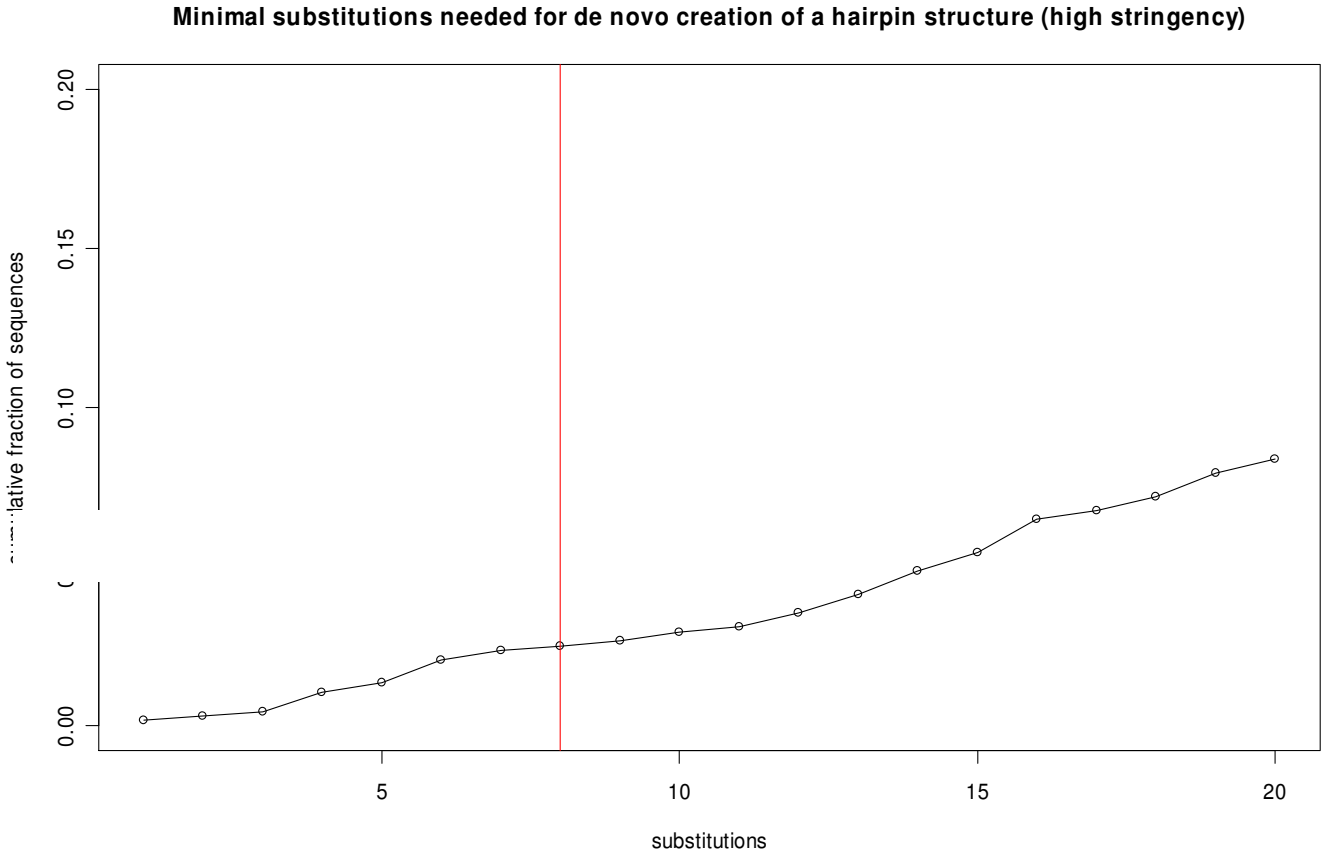
The difference in miRNA expression between species (in terms of expression fold changes) is significantly correlated with neither the difference in probabilities of folding into the most stable hairpin structure (determined by RNASHAPES software) nor the difference in deltaG. However, we observed a significant correlation between the expression difference and secondary structure divergence as determined by RNAFORESTER program(r =-0.409, P-Value = 0.011, counting miRNAs that have more than 20 reads observed in *D. melanogaster*). But this significant correlation is caused by *miR-284*. After removing this point, the correlation disappear.

**References**

1. Pasquinelli, A. E. et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-9 (2000).
2. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
3. Borenstein, E. & Ruppin, E. Direct evolution of genetic robustness in microRNA. *Proc Natl Acad Sci U S A* **103**, 6593-8 (2006).
4. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, 858-62 (2001).
5. Kent, W. J. BLAT---The BLAST-Like Alignment Tool
10.1101/gr.229202. Article published online before March 2002. *Genome Res.* **12**, 656-664 (2002).
6. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402 (1997).
7. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-3.0. 1996-2004 <http://www.repeatmasker.org>.
8. Hofacker, I. L. et al. Fast Folding and Comparison of RNA Secondary Structures (The Vienna RNA Package). *Monatsh. Chem.* **125**, 167--188 (1994).
9. Hochsmann, M., Voss, B. & Giegerich, R. Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans Comput Biol Bioinform* **1**, 53-62 (2004).
10. Ruby, J. G. et al. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.*, gr.6597907 (2007).
11. Allen, E. et al. Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat Genet* **36**, 1282-90 (2004).
12. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* **20**, 3407-25 (2006).
13. Fahlgren, N. et al. High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. *PLoS ONE* **2**, e219 (2007).
14. Halligan, D. L. & Keightley, P. D. Ubiquitous selective constraints in the Drosophila genome revealed by a genome-wide interspecies comparison. *Genome Res* **16**, 875-84 (2006).




Supplementary Fig. 1 The survival function estimated by the background hairpin approach (empirical data approach) is very close to that obtained by computer simulation (neutral simulation approach), with respect to the "shape" of the survival functions.




Supplementary Fig. 2 The number of substitutions required to convert a 150-nt sequence that cannot form a hairpin structure (under the low stringency criteria) into a hairpin structure meeting with the high stringency criteria. The X axis is the number of mutations a non-hairpin sequence needed to form a hairpin sequence that can pass the high stringency criteria. The Y axis is the cumulative fraction of sequences. The red line is the number of substitutions accumulated in the *D. simulans* lineage for a 150-nt segment under neutral evolution.

## Tables

**Supplementary Table 1 The reads observed for the known and newly identified *Drosophila* miRNA genes (ranked by the descendant expression level in *D. melanogaster*).**

| Gene | Original Read Number | | | Normalized Read Number(per 100,000 reads) | | |
|---|---|---|---|---|---|---|
| | Dmel | Dsim | Dpse | Dmel | Dsim | Dpse |
| **Known miRNA genes** | | | | | | |
| *miR-8* | 5067 | 29852 | 2969 | 31,501.40 | 39,892.00 | 13,373.90 |
| *miR-276A* | 3800 | 6628 | 2262 | 23,624.50 | 8,857.20 | 10,189.20 |
| *miR-277* | 1203 | 10585 | 1211 | 7,479.00 | 14,145.00 | 5,455.00 |
| *miR-124* | 724 | 2638 | 881 | 4,501.10 | 3,525.20 | 3,968.50 |
| *miR-276B* | 588 | 2164 | 760 | 3,655.60 | 2,891.80 | 3,423.40 |
| *miR-305* | 420 | 1949 | 451 | 2,611.10 | 2,604.50 | 2,031.50 |
| *miR-34* | 414 | 1389 | 872 | 2,573.80 | 1,856.20 | 3,927.90 |
| *miR-278* | 310 | 670 | 519 | 1,927.30 | 895.3 | 2,337.80 |
| *miR-317* | 199 | 1500 | 885 | 1,237.20 | 2,004.50 | 3,986.50 |
| *miR-285* | 198 | 698 | 359 | 1,231.00 | 932.8 | 1,617.10 |
| *miR-274* | 164 | 387 | 262 | 1,019.60 | 517.2 | 1,180.20 |
| *miR-184* | 158 | 268 | 912 | 982.3 | 358.1 | 4,108.10 |
| *miR-13B-2* | 151 | 458 | 100 | 938.8 | 612 | 450.5 |
| *miR-13B-1* | 150 | 457 | 100 | 932.5 | 610.7 | 450.5 |
| *miR-1* | 149 | 2200 | 769 | 926.3 | 2,939.90 | 3,464.00 |
| *miR-210* | 146 | 649 | 337 | 907.7 | 867.3 | 1,518.00 |
| *miR-125* | 136 | 300 | 192 | 845.5 | 400.9 | 864.9 |
| *LET-7* | 122 | 1839 | 961 | 758.5 | 2,457.50 | 4,328.80 |
| *miR-14* | 102 | 45 | 65 | 634.1 | 60.1 | 292.8 |
| *BANTAM* | 56 | 89 | 104 | 348.2 | 118.9 | 468.5 |
| *miR-2A-2* | 51 | 905 | 1084 | 317.1 | 1,209.40 | 4,882.90 |
| *miR-284* | 49 | 24 | 139 | 304.6 | 32.1 | 626.1 |
| *miR-11* | 37 | 118 | 66 | 230 | 157.7 | 297.3 |
| *miR-263A* | 37 | 60 | 52 | 230 | 80.2 | 234.2 |
| *miR-2A-1* | 35 | 889 | 1039 | 217.6 | 1,188.00 | 4,680.20 |
| *miR-7* | 29 | 33 | 38 | 180.3 | 44.1 | 171.2 |
| *miR-12* | 26 | 740 | 68 | 161.6 | 988.9 | 306.3 |
| *miR-279* | 26 | 125 | 112 | 161.6 | 167 | 504.5 |
| *miR-133* | 23 | 42 | 60 | 143 | 56.1 | 270.3 |
| *miR-219* | 21 | 59 | 50 | 130.6 | 78.8 | 225.2 |
| *miR-315* | 21 | 42 | 36 | 130.6 | 56.1 | 162.2 |
| *miR-31A* | 19 | 75 | 32 | 118.1 | 100.2 | 144.1 |
| *miR-9A* | 16 | 84 | 43 | 99.5 | 112.3 | 193.7 |
| *miR-33* | 15 | 70 | 83 | 93.3 | 93.5 | 373.9 |
| *miR-13A* | 14 | 61 | 92 | 87 | 81.5 | 414.4 |
| *miR-307* | 13 | 31 | 163 | 80.8 | 41.4 | 734.2 |
| *miR-2B-2* | 11 | 104 | 77 | 68.4 | 139 | 346.8 |
| *miR-79* | 10 | 28 | 20 | 62.2 | 37.4 | 90.1 |
| *miR-2C* | 8 | 84 | 102 | 49.7 | 112.3 | 459.5 |
| *miR-2B-1* | 7 | 71 | 0 | 43.5 | 94.9 | 0 |
| *miR-100* | 7 | 16 | 4 | 43.5 | 21.4 | 18 |
| *miR-316* | 6 | 21 | 0 | 37.3 | 28.1 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *miR-306* | 5 | 37 | 45 | 31.1 | 49.4 | 202.7 |
| *miR-282* | 5 | 59 | 35 | 31.1 | 78.8 | 157.7 |
| *miR-281-1* | 4 | 42 | 2 | 24.9 | 56.1 | 9 |
| *miR-281-2* | 4 | 51 | 2 | 24.9 | 68.2 | 9 |
| *miR-304* | 4 | 2 | 6 | 24.9 | 2.7 | 27 |
| *miR-10* | 3 | 53 | 24 | 18.7 | 70.8 | 108.1 |
| *miR-9B* | 3 | 40 | 5 | 18.7 | 53.5 | 22.5 |
| *miR-87* | 3 | 9 | 0 | 18.7 | 12 | 0 |
| *miR-31B* | 1 | 22 | 0 | 6.2 | 29.4 | 0 |
| *miR-92A* | 1 | 2 | 1 | 6.2 | 2.7 | 4.5 |
| *miR-263B* | 1 | 3 | 22 | 6.2 | 4 | 99.1 |
| *miR-6-1* | 1 | 0 | 0 | 6.2 | 0 | 0 |
| *miR-6-2* | 1 | 0 | 0 | 6.2 | 0 | 0 |
| *miR-6-3* | 1 | 0 | 0 | 6.2 | 0 | 0 |
| *miR-314* | 0 | 14 | 0 | 0 | 18.7 | 0 |
| *miR-308* | 0 | 11 | 51 | 0 | 14.7 | 229.7 |
| *miR-312* | 0 | 9 | 0 | 0 | 12 | 0 |
| *miR-283* | 0 | 4 | 1 | 0 | 5.3 | 4.5 |
| *miR-311* | 0 | 1 | 0 | 0 | 1.3 | 0 |
| *miR-313* | 0 | 1 | 0 | 0 | 1.3 | 0 |
| *miR-9C* | 0 | 6 | 5 | 0 | 8 | 22.5 |
| *miR-275* | 0 | 7 | 2 | 0 | 9.4 | 9 |
| *miR-310* | 0 | 2 | 0 | 0 | 2.7 | 0 |
| *IAB-4* | 0 | 2 | 0 | 0 | 2.7 | 0 |
| *miR-5* | 0 | 0 | 2 | 0 | 0 | 9 |
| **Newly identified miRNA genes** | | | | | | |
| *miR-2016* | 348 | 1233 | 416 | 2237.4 | 1742.8 | 2103.6 |
| *miR-2026* | 285 | 878 | 408 | 1832.3 | 1241.0 | 2063.1 |
| *miR-2024* | 265 | 1010 | 296 | 1703.7 | 1427.6 | 1496.8 |
| *miR-2005* | 133 | 507 | 177 | 855.1 | 716.6 | 895.0 |
| *miR-2031* | 54 | 434 | 160 | 347.2 | 613.4 | 809.1 |
| *miR-2023* | 24 | 69 | 59 | 154.3 | 97.5 | 298.3 |
| *miR-2029* | 17 | 23 | 35 | 109.3 | 32.5 | 177.0 |
| *miR-2007* | 12 | 94 | 37 | 77.2 | 132.9 | 187.1 |
| *miR-2037* | 12 | 33 | 9 | 77.2 | 46.6 | 45.5 |
| *miR-2006* | 10 | 23 | 32 | 64.3 | 32.5 | 161.8 |
| *miR-2030* | 9 | 38 | 38 | 57.9 | 53.7 | 192.2 |
| *miR-2019* | 7 | 22 | 15 | 45.0 | 31.1 | 75.8 |
| *miR-2014* | 5 | 16 | 38 | 32.1 | 22.6 | 192.2 |
| *miR-2034* | 5 | 43 | 7 | 32.1 | 60.8 | 35.4 |
| *miR-2008* | 4 | 28 | 25 | 25.7 | 39.6 | 126.4 |
| *miR-2038* | 4 | 0 | 4 | 25.7 | 0.0 | 20.2 |
| *miR-2057* | 3 | 9 | 6 | 19.3 | 12.7 | 30.3 |
| *miR-2123* | 3 | 8 | 0 | 19.3 | 11.3 | 0.0 |
| *miR-2021* | 2 | 24 | 6 | 12.9 | 33.9 | 30.3 |
| *miR-2308* | 2 | 1 | 0 | 12.9 | 1.4 | 0.0 |
| *miR-2025* | 1 | 93 | 28 | 6.4 | 131.4 | 141.6 |
| *miR-2039* | 1 | 6 | 20 | 6.4 | 8.5 | 101.1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *miR-2020* | 1 | 17 | 5 | 6.4 | 24.0 | 25.3 |
| *miR-2018* | 1 | 3 | 3 | 6.4 | 4.2 | 15.2 |
| *miR-2140* | 1 | 7 | 1 | 6.4 | 9.9 | 5.1 |
| *miR-2118* | 1 | 5 | 1 | 6.4 | 7.1 | 5.1 |
| *miR-2033* | 1 | 14 | 0 | 6.4 | 19.8 | 0.0 |
| *miR-2130* | 1 | 5 | 0 | 6.4 | 7.1 | 0.0 |
| *miR-2154* | 1 | 2 | 0 | 6.4 | 2.8 | 0.0 |
| *miR-2222* | 1 | 1 | 0 | 6.4 | 1.4 | 0.0 |
| *miR-2003* | 1 | 1 | 0 | 6.4 | 1.4 | 0.0 |
| *miR-2364* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2323* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2318* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2255* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2238* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2149* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2095* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2002* | 1 | 0 | 0 | 6.4 | 0.0 | 0.0 |
| *miR-2596* | 0 | 0 | 14 | 0.0 | 0.0 | 70.8 |
| *miR-2009* | 0 | 6 | 6 | 0.0 | 8.5 | 30.3 |
| *miR-2027* | 0 | 0 | 2 | 0.0 | 0.0 | 10.1 |
| *miR-2298* | 0 | 0 | 1 | 0.0 | 0.0 | 5.1 |
| *miR-2265* | 0 | 0 | 1 | 0.0 | 0.0 | 5.1 |
| *miR-2242* | 0 | 0 | 1 | 0.0 | 0.0 | 5.1 |
| *miR-2125* | 0 | 0 | 1 | 0.0 | 0.0 | 5.1 |
| *miR-2116* | 0 | 0 | 1 | 0.0 | 0.0 | 5.1 |
| *miR-2332* | 0 | 126 | 0 | 0.0 | 178.1 | 0.0 |
| *miR-2986* | 0 | 36 | 0 | 0.0 | 50.9 | 0.0 |
| *miR-2271* | 0 | 22 | 0 | 0.0 | 31.1 | 0.0 |
| *miR-2165* | 0 | 18 | 0 | 0.0 | 25.4 | 0.0 |
| *miR-2160* | 0 | 17 | 0 | 0.0 | 24.0 | 0.0 |
| *miR-2346* | 0 | 9 | 0 | 0.0 | 12.7 | 0.0 |
| *miR-2338* | 0 | 9 | 0 | 0.0 | 12.7 | 0.0 |
| *miR-2234* | 0 | 9 | 0 | 0.0 | 12.7 | 0.0 |
| *miR-2359* | 0 | 8 | 0 | 0.0 | 11.3 | 0.0 |
| *miR-2881* | 0 | 6 | 0 | 0.0 | 8.5 | 0.0 |
| *miR-2245* | 0 | 6 | 0 | 0.0 | 8.5 | 0.0 |
| *miR-3018* | 0 | 5 | 0 | 0.0 | 7.1 | 0.0 |
| *miR-2835* | 0 | 5 | 0 | 0.0 | 7.1 | 0.0 |
| *miR-2797* | 0 | 5 | 0 | 0.0 | 7.1 | 0.0 |
| *miR-2163* | 0 | 4 | 0 | 0.0 | 5.7 | 0.0 |
| *miR-2968* | 0 | 3 | 0 | 0.0 | 4.2 | 0.0 |
| *miR-2868* | 0 | 3 | 0 | 0.0 | 4.2 | 0.0 |
| *miR-2774* | 0 | 3 | 0 | 0.0 | 4.2 | 0.0 |
| *miR-2705* | 0 | 3 | 0 | 0.0 | 4.2 | 0.0 |
| *miR-2839* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2762* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2753* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2635* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2600* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2548* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *miR-2546* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2394* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2349* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2341* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2329* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2283* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2212* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2208* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2178* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2124* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-2011* | 0 | 2 | 0 | 0.0 | 2.8 | 0.0 |
| *miR-3084* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3079* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3045* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3031* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3030* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3021* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3020* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-3010* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2970* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2960* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2943* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2937* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2933* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2921* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2904* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2897* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2894* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2882* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2866* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2828* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2819* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2815* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2810* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2804* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2800* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2795* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2771* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2763* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2761* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2737* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2727* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2690* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2583* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2557* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2551* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2419* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2418* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2387* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2378* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *miR-2375* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2352* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2343* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2342* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2340* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2331* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2327* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2289* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2267* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2266* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2262* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2261* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2258* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2256* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2254* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2252* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2250* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2247* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2246* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2241* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2240* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2239* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2231* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2227* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2226* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2225* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2224* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2219* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2217* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2216* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2210* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2209* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2207* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2205* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2204* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2180* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2171* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2166* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2164* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2152* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2151* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2150* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2141* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2139* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2133* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2121* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2103* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2096* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2094* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2089* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *miR-2088* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2085* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2082* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2081* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2072* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2064* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2061* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2059* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2058* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2056* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2055* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2052* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2045* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2044* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2041* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2040* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2035* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2032* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |
| *miR-2001* | 0 | 1 | 0 | 0.0 | 1.4 | 0.0 |

**Supplementary Table 2 The phylogenetic patterns of the putative miRNAs, the background hairpins with and without reads expression data from Ref. [10] (The percentages are in the parentheses)**

| Phylogenetic Patterns | miRNA genes (s m y a p v) | background hairpins (m s y a p v) | background hairpins matching reads from Ref. [10] (m s y a p v) |
|---|---|---|---|
| 1-0-0-0-0-0 | 42 (25.61) | 89,979 (76.67) | 1,370 (73.62) |
| 1-0-0-0-0-1 | 0 (0) | 106 (0.09) | 8 (0.43) |
| 1-0-0-0-1-0 | 0 (0) | 191 (0.16) | 5 (0.27) |
| 1-0-0-0-1-1 | 0 (0) | 1 (0) | 1 (0.05) |
| 1-0-0-1-0-0 | 0 (0) | 345 (0.29) | 4 (0.21) |
| 1-0-0-1-0-1 | 0 (0) | 9 (0.01) | 0 (0) |
| 1-0-0-1-1-0 | 0 (0) | 12 (0.01) | 0 (0) |
| 1-0-1-0-0-0 | 5 (3.05) | 4,489 (3.82) | 87 (4.67) |
| 1-0-1-0-0-1 | 0 (0) | 21 (0.02) | 1 (0.05) |
| 1-0-1-0-1-0 | 0 (0) | 29 (0.02) | 1 (0.05) |
| 1-0-1-0-1-1 | 0 (0) | 2 (0) | 0 (0) |
| 1-0-1-1-0-0 | 1 (0.61) | 80 (0.07) | 3 (0.16) |
| 1-0-1-1-0-1 | 0 (0) | 3 (0) | 0 (0) |
| 1-0-1-1-1-0 | 0 (0) | 7 (0.01) | 0 (0) |
| 1-1-0-0-0-0 | 17 (10.37) | 16,014 (13.64) | 260 (13.97) |
| 1-1-0-0-0-1 | 0 (0) | 43 (0.04) | 0 (0) |
| 1-1-0-0-1-0 | 1 (0.61) | 315 (0.27) | 14 (0.75) |
| 1-1-0-0-1-1 | 1 (0.61) | 10 (0.01) | 0 (0) |
| 1-1-0-1-0-0 | 0 (0) | 193 (0.16) | 7 (0.38) |
| 1-1-0-1-0-1 | 0 (0) | 3 (0) | 0 (0) |
| 1-1-0-1-1-0 | 0 (0) | 56 (0.05) | 1 (0.05) |
| 1-1-0-1-1-1 | 1 (0.61) | 6 (0.01) | 0 (0) |
| 1-1-1-0-0-0 | 12 (7.32) | 4,542 (3.87) | 66 (3.55) |
| 1-1-1-0-0-1 | 0 (0) | 39 (0.03) | 2 (0.11) |
| 1-1-1-0-1-0 | 0 (0) | 323 (0.28) | 10 (0.54) |
| 1-1-1-0-1-1 | 3 (1.83) | 35 (0.03) | 1 (0.05) |
| 1-1-1-1-0-0 | 7 (4.27) | 210 (0.18) | 3 (0.16) |
| 1-1-1-1-0-1 | 2 (1.22) | 27 (0.02) | 0 (0) |
| 1-1-1-1-1-0 | 3 (1.83) | 193 (0.16) | 11 (0.59) |
| 1-1-1-1-1-1 | 69 *(42.07) | 79 (0.07) | 6 (0.32) |
| Total | 164 (100) | 117,362 (100) | 1,861 (100) |

* 50 are from the known that meet with the high stringency criteria and have expression evidence in this study.

Note that the phylogenetic positions of *D. melanogaster* and *D. simulans* are interchangeable.

**Supplementary Table 3 The survival function of a new miRNA gene after t Myrs.**

| Conservation pattern | ith | T6>t≥t5 | T5>t≥t4 | T4>t≥t3 | T3>t≥t2 | T2>t≥t1 | Sum |
|---|---|---|---|---|---|---|---|
| 1-0-0-0-0 | 1 | $P_{1,5}$ | $P_{1,4}$ | $P_{1,3}$ | $P_{1,2}$ | $P_{1,1}$ | $P_1$ |
| 1-0-0-0-1 | 2 | $P_{2,5}$ | | | | | $P_2$ |
| 1-0-0-1-0 | 3 | $P_{3,5}$ | $P_{3,4}$ | | | | $P_3$ |
| 1-0-0-1-1 | 4 | $P_{4,5}$ | | | | | $P_4$ |
| 1-0-1-0-0 | 5 | $P_{5,5}$ | $P_{5,4}$ | $P_{5,3}$ | | | $P_5$ |
| 1-0-1-0-1 | 6 | $P_{6,5}$ | | | | | $P_6$ |
| 1-0-1-1-0 | 7 | $P_{7,5}$ | $P_{7,4}$ | | | | $P_7$ |
| 1-0-1-1-1 | 8 | $P_{8,5}$ | | | | | $P_8$ |
| 1-1-0-0-0 | 9 | $P_{9,5}$ | $P_{9,4}$ | $P_{9,3}$ | $P_{9,2}$ | | $P_9$ |
| 1-1-0-0-1 | 10 | $P_{10,5}$ | | | | | $P_{10}$ |
| 1-1-0-1-0 | 11 | $P_{11,5}$ | $P_{11,4}$ | | | | $P_{11}$ |
| 1-1-0-1-1 | 12 | $P_{12,5}$ | | | | | $P_{12}$ |
| 1-1-1-0-0 | 13 | $P_{13,5}$ | $P_{13,4}$ | $P_{13,3}$ | | | $P_{13}$ |
| 1-1-1-0-1 | 14 | $P_{14,5}$ | | | | | $P_{14}$ |
| 1-1-1-1-0 | 15 | $P_{15,5}$ | $P_{15,4}$ | | | | $P_{15}$ |
| 1-1-1-1-1 | 16 | $P_{16,5}$ | | | | | $P_{16}$ |
| 0-0-0-0-0 | 17 | $P_{17,5}$ | $P_{17,4}$ | $P_{17,3}$ | $P_{17,2}$ | $P_{17,1}$ | $P_{17}$ |
| 0-0-0-0-1 | 18 | $P_{18,5}$ | | | | | $P_{18}$ |
| 0-0-0-1-0 | 19 | $P_{19,5}$ | $P_{19,4}$ | | | | $P_{19}$ |
| 0-0-0-1-1 | 20 | $P_{20,5}$ | | | | | $P_{20}$ |
| 0-0-1-0-0 | 21 | $P_{21,5}$ | $P_{21,4}$ | $P_{21,3}$ | | | $P_{21}$ |
| 0-0-1-0-1 | 22 | $P_{22,5}$ | | | | | $P_{22}$ |
| 0-0-1-1-0 | 23 | $P_{23,5}$ | $P_{23,4}$ | | | | $P_{23}$ |
| 0-0-1-1-1 | 24 | $P_{24,5}$ | | | | | $P_{24}$ |
| 0-1-0-0-0 | 25 | $P_{25,5}$ | $P_{25,4}$ | $P_{25,3}$ | $P_{25,2}$ | | $P_{25}$ |
| 0-1-0-0-1 | 26 | $P_{26,5}$ | | | | | $P_{26}$ |
| 0-1-0-1-0 | 27 | $P_{27,5}$ | $P_{27,4}$ | | | | $P_{27}$ |
| 0-1-0-1-1 | 28 | $P_{28,5}$ | | | | | $P_{28}$ |
| 0-1-1-0-0 | 29 | $P_{29,5}$ | $P_{29,4}$ | $P_{29,3}$ | | | $P_{29}$ |
| 0-1-1-0-1 | 30 | $P_{30,5}$ | | | | | $P_{30}$ |
| 0-1-1-1-0 | 31 | $P_{31,5}$ | $P_{31,4}$ | | | | $P_{31}$ |
| 0-1-1-1-1 | 32 | $P_{32,5}$ | | | | | $P_{32}$ |
| Sum | | 1 | 1 | 1 | 1 | 1 | 1 |

$P_{i,j}$ is the survival probability that a new miRNA gene generated on brach j and generated the *i*th phylogenetic pattern. . $P_{i,j} = \int_{t_j}^{t_{j+1}} S_{i,j}(t)dt$ and the total probability for the *i*th phylogenetic pattern is $P_i = \sum_{j=1}^{5} \int_{t_j}^{t_{j+1}} S_{i,j}(t)dt$ .

**Supplementary Table 4 The detailed statistical formula of the survival function**

| Phylogenetic pattern | *i*th | survival function | detailed statistical formula |
|---|---|---|---|
| 10000 | 1 | $P_{1,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)]S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10000 | 1 | $P_{1,4}$ | $[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)]S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10000 | 1 | $P_{1,3}$ | $[S_{t3}(t)-S_{t1}(t)]S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]/S_{t3}(t)/S_{t2}(t)$ |
| 10000 | 1 | $P_{1,2}$ | $S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]/S_{t2}(t)$ |
| 10000 | 1 | $P_{1,1}$ | $S_{t1}(t)$ |
| 10001 | 2 | $P_{2,5}$ | $S_{t1}(t)^2[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10010 | 3 | $P_{3,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10010 | 3 | $P_{3,4}$ | $S_{t1}(t)^2[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10011 | 4 | $P_{4,5}$ | $S_{t1}(t)^3[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10100 | 5 | $P_{5,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10100 | 5 | $P_{5,4}$ | $[S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10100 | 5 | $P_{5,3}$ | $S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t3}(t)/S_{t2}(t)$ |
| 10101 | 6 | $P_{6,5}$ | $S_{t1}(t)^3[S_{t4}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10110 | 7 | $P_{7,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^3[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10110 | 7 | $P_{7,4}$ | $S_{t1}(t)^3[S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 10111 | 8 | $P_{8,5}$ | $S_{t1}(t)^4[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11000 | 9 | $P_{9,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)]S_{t1}(t)^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11000 | 9 | $P_{9,4}$ | $[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)]S_{t1}(t)^2/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11000 | 9 | $P_{9,3}$ | $[S_{t3}(t)-S_{t1}(t)]S_{t1}(t)^2/S_{t3}(t)/S_{t2}(t)$ |
| 11000 | 9 | $P_{9,2}$ | $S_{t1}(t)^2/S_{t2}(t)$ |
| 11001 | 10 | $P_{10,5}$ | $S_{t1}(t)^3[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11010 | 11 | $P_{11,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^3[S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11010 | 11 | $P_{11,4}$ | $S_{t1}(t)^3[S_{t3}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11011 | 12 | $P_{12,5}$ | $S_{t1}(t)^4[S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11100 | 13 | $P_{13,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^3/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11100 | 13 | $P_{13,4}$ | $[S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^3/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11100 | 13 | $P_{13,3}$ | $S_{t1}(t)^3/S_{t3}(t)/S_{t2}(t)$ |
| 11101 | 14 | $P_{14,5}$ | $S_{t1}(t)^4[S_{t4}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11110 | 15 | $P_{15,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^4/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 11110 | 15 | $P_{15,4}$ | $S_{t1}(t)^4/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |

| | | | |
|---|---|---|---|
| 11111 | 16 | $P_{16,5}$ | $S_{t1}(t)^5/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00000 | 17 | $P_{17,5}$ | $[1-S_{t5}(t)]+[S_{t5}(t)-S_{t4}(t)][S_{t5}(t)-S_{t1}(t)]/S_{t5}(t)+[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t3}(t)][S_{t4}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)+[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00000 | 17 | $P_{17,4}$ | $[1-S_{t4}(t)]+[S_{t4}(t)-S_{t3}(t)][S_{t4}(t)-S_{t1}(t)]/S_{t4}(t)+[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)+[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00000 | 17 | $P_{17,3}$ | $[1-S_{t3}(t)]+[S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t3}(t)+[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t3}(t)/S_{t2}(t)$ |
| 00000 | 17 | $P_{17,2}$ | $[1-S_{t2}(t)]+[S_{t2}(t)-S_{t1}(t)]^2/S_{t2}(t)$ |
| 00000 | 17 | $P_{17,1}$ | $[1-S_{t1}(t)]$ |
| 00001 | 18 | $P_{18,5}$ | $[S_{t5}(t)-S_{t4}(t)]S_{t1}(t)/S_{t5}(t)+S_{t1}(t)[S_{t4}(t)-S_{t3}(t)][S_{t4}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)+S_{t1}(t)[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00010 | 19 | $P_{19,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t3}(t)]S_{t1}(t)/S_{t5}(t)/S_{t4}(t)+[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)[S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00010 | 19 | $P_{19,4}$ | $[S_{t4}(t)-S_{t3}(t)]S_{t1}(t)/S_{t4}(t)+S_{t1}(t)[S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00011 | 20 | $P_{20,5}$ | $S_{t1}(t)^2[S_{t4}(t)-S_{t3}(t)]/S_{t5}(t)/S_{t4}(t)+S_{t1}(t)^2[S_{t3}(t)-S_{t2}(t)][S_{t3}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)^2[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00100 | 21 | $P_{21,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)]S_{t1}(t)/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)]S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00100 | 21 | $P_{21,4}$ | $[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)]S_{t1}(t)/S_{t4}(t)/S_{t3}(t)+[S_{t4}(t)-S_{t1}(t)]S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]^2/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00100 | 21 | $P_{21,3}$ | $[S_{t3}(t)-S_{t2}(t)]S_{t1}(t)/S_{t3}(t)+S_{t1}(t)[S_{t2}(t)-S_{t1}(t)]^2/S_{t3}(t)/S_{t2}(t)$ |
| 00101 | 22 | $P_{22,5}$ | $S_{t1}(t)^2[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t2}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)^2[S_{t4}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00110 | 23 | $P_{23,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t3}(t)-S_{t2}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00110 | 23 | $P_{23,4}$ | $S_{t1}(t)^2[S_{t3}(t)-S_{t2}(t)]/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]^2/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 00111 | 24 | $P_{24,5}$ | $S_{t1}(t)^3[S_{t3}(t)-S_{t2}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)+S_{t1}(t)^3[S_{t2}(t)-S_{t1}(t)]^2/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01000 | 25 | $P_{25,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]S_{t1}(t)/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01000 | 25 | $P_{25,4}$ | $[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]S_{t1}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01000 | 25 | $P_{25,3}$ | $[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]S_{t1}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01000 | 25 | $P_{25,2}$ | $[S_{t2}(t)-S_{t1}(t)]S_{t1}(t)/S_{t2}(t)$ |
| 01001 | 26 | $P_{26,5}$ | $S_{t1}(t)^2[S_{t4}(t)-S_{t1}(t)][S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |

| | | | |
|---|---|---|---|
| 01010 | 27 | $P_{27,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01010 | 27 | $P_{27,4}$ | $S_{t1}(t)^2[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01011 | 28 | $P_{28,5}$ | $S_{t1}(t)^3[S_{t3}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01100 | 29 | $P_{29,5}$ | $[S_{t5}(t)-S_{t1}(t)][S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01100 | 29 | $P_{29,4}$ | $[S_{t4}(t)-S_{t1}(t)]S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01100 | 29 | $P_{29,3}$ | $S_{t1}(t)^2[S_{t2}(t)-S_{t1}(t)]/S_{t3}(t)/S_{t2}(t)$ |
| 01101 | 30 | $P_{30,5}$ | $S_{t1}(t)^3[S_{t4}(t)-S_{t1}(t)][S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01110 | 31 | $P_{31,5}$ | $[S_{t5}(t)-S_{t1}(t)]S_{t1}(t)^3[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01110 | 31 | $P_{31,4}$ | $S_{t1}(t)^3[S_{t2}(t)-S_{t1}(t)]/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |
| 01111 | 32 | $P_{32,5}$ | $S_{t1}(t)^4[S_{t2}(t)-S_{t1}(t)]/S_{t5}(t)/S_{t4}(t)/S_{t3}(t)/S_{t2}(t)$ |

Note $S_{tj}(t) = \int_0^{tj} S(T \geq t)dt$ .