



# MicroRNA analysis in model species based on evolutionary rates

X.F. Mao and Y.C. Cao

Department of Molecular Biology, School of Bioscience and Bioengineering,  
South China University of Technology, Guangzhou, China

Corresponding author: X.F. Mao  
E-mail: mao.xf@mail.scut.edu.cn

Genet. Mol. Res. 15 (1): gmr.15017216  
Received August 15, 2015  
Accepted November 27, 2015  
Published March 28, 2016  
DOI <http://dx.doi.org/10.4238/gmr.15017216>

**ABSTRACT.** MicroRNAs (miRNAs) are major post-transcriptional regulators of gene expression. In an attempt to gain insights into miRNAs at the macroevolutionary level, we performed a systematic analysis of miRNAs in six model organisms based on their evolutionary rates. First, we calculated their miRNA evolutionary rates, and found that they did not correlate with the complexity of the organisms. A correlation between evolutionary rates and single nucleotide polymorphisms (SNPs) in the miRNA sequence suggested that slow-evolving miRNAs in humans tolerate more SNPs than miRNAs with similar evolutionary rates in other species. However, fast-evolving miRNAs had lower SNP densities in humans than in the fruit fly. We also found that evolutionary rates were correlated with the proportion of parasite or clustered miRNAs. This correlation exhibited a different pattern in zebrafish, which may be related to significant genome duplication in the early vertebrates. The minimized free energy of the miRNA stem-loop structure was not correlated with the evolutionary rates of any species in our analysis. After evaluating relative miRNA expression levels, we observed that newly emerged miRNAs in complex species would integrate into the gene network at a faster pace and be functionally important; therefore, miRNAs may have accelerated human evolution.

**Key words:** MicroRNA; Evolutionary rate; Molecular evolution; Model species; Comparative study

## INTRODUCTION

MicroRNAs (miRNAs) are a class of endogenous, small, non-coding RNAs that can activate, degrade, or suppress the expression of target mRNAs. They are about 22 nt in length, and account for 1-5% of the genome in a number of species, ranging from animals to plants. They regulate more than 30% of all genes, and participate in almost every biological process, including metabolism, cell differentiation, growth, the cell cycle, etc. Thus far, 24,521 pre-miRNA stem-loop sequences and 30,424 mature miRNA sequences from 206 species have been deposited in miRBase (release 20) (Kozomara and Griffiths-Jones, 2014). Many miRNAs in different species are conserved. For example, let-7 is present in worms, fruit flies, and humans.

miRNAs bind to the argonaute protein, and form an RNA-induced silencing complex that mainly targets the 3'-untranslated region of an mRNA to mediate mRNA degradation, or to suppress expression. However, the impact of miRNAs on the regulation of gene expression is controversial. It is widely believed that the two main functions of miRNAs are "tuning", which is related to regulating gene expression levels, and "buffering" (also called "canalization"), which is reducing the expression variants of the target gene. For example, in flies and mammals, miRNAs had been shown to reduce target gene-expression divergence (Cui et al., 2007), consequently increasing species adaptability and flexibility. However, other studies have found that genes targeted by miRNAs have greater expression divergence (Lu and Clark, 2012). Since gene expression patterns can be treated as quantitative phenotypes, the miRNA regulation system may increase diversity within a species. Indeed, support for this theory comes from the studies of Lu and Clark (2012), who found that miRNAs increase variation in human gene expression. Therefore, by increasing gene expression variation, miRNAs may contribute to the high divergence and complexity across species. However, regardless of the theories, miRNAs are tightly bound to the evolutionary process.

The origin of miRNAs can be traced to early metazoans. miRNAs, distinct from transcription factors, evolve quickly and undergo a fast turnover (Nozawa et al., 2010). After integrating into the gene network, miRNAs perform important functions and are seldom lost during the evolutionary process (Wheeler et al., 2009). Intriguingly, during the evolution of bilateria and vertebrates, miRNA expansion coincided with animal body-plan formations and phenotype metamorphosis (Christodoulou et al., 2010; Berezikov, 2011). The largest expansion of miRNAs has thus far been observed in rodent and primate lineages (Guerra-Assunção and Enright, 2012). In the mammal lineage, miRNAs seemed to have evolved rapidly, with some newborn miRNAs getting lost while many others with increased expression were conserved in the nervous system, thus displaying robustness in the gene networks (Meunier et al., 2013). Such rapid evolutionary rates have also been observed during the early evolution of marsupials and the late evolutionary history of humans (Iwama et al., 2013). All of these findings indicate that miRNAs are linked to phenotype formation and thus species evolution.

In general, the emergence of a new miRNA is more likely than a new protein-coding gene, because the formation of the stem-loop structure in the short miRNA sequence is simple (Chen and Rajewsky, 2007). A novel miRNA could evolve from the following genomic sources: 1) gene duplications, including tandem duplications and non-local duplications; 2) introns (Isik et al., 2010); 3) transposable elements (Yuan et al., 2011); 4) pseudogenes, snoRNAs, and tRNAs (Pederson, 2010); 5) antisense miRNA transcripts; and 6) *de novo* emergence (Liu et al., 2008). miRNAs with high evolutionary rates are mostly young, reside in introns, are expressed at low levels, and seem species-specific (Wang et al., 2011; Guerra-Assunção and Enright, 2012; Zhu et al., 2012). In humans, miRNAs with low expression levels have high evolutionary rates (Liang and Li, 2009).

miRNAs involved in the upstream activities of gene regulation pathways always evolve rapidly and tend to target proteins with low evolutionary rates, while miRNAs acting downstream of gene regulation evolve more slowly and target fast-evolving proteins (Qiu et al., 2010). In fruit flies, the evolutionary rates of miRNA backbones are negatively correlated with miRNA expression levels, but are positively correlated with miRNA biogenesis error rates (Shen et al., 2011). Young human miRNAs have been found to have low expression levels, high tissue specificity, high minimized free energy (MFE) in the stem-loop sequences, as well as high substitution rates, in both pre-miRNAs and flanking sequences (Zhu et al., 2012).

Since these studies have been conducted on closely related organisms, it is difficult to envisage miRNA evolution at the macro-level. In a previous study, we reported that miRNA systems differed between six model organisms, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens* (Mao et al., 2014). In this study, we performed additional data mining of the miRNA system in the six model organisms based on their evolutionary rates, in order to ascertain whether the results obtained by previous studies are generally applicable, and to improve our understanding of global miRNAs in a variety of organisms.

## MATERIAL AND METHODS

### miRNA data sources and genomes

In this study, we selected six model organisms: the worm *Caenorhabditis elegans* (Cel), the fruit fly *Drosophila melanogaster* (Dme), the zebrafish *Danio rerio* (Dre), the brown rat *Rattus norvegicus* (Rno), the mouse *Mus musculus* (Mmu), and the human *Homo sapiens* (Hsa). These model organisms are landmarks in evolutionary history, and have been well studied. For each species, we chose a closely related organism to retrieve homologous sequences of miRNAs, and their flanking regions, for calculating evolutionary rates. The genomes of all of the species were downloaded from the UCSC Genome Browser (Table 1). The miRNA precursor, mature sequence, and seed information were downloaded from miRBase (release 20).

**Table 1.** Genome versions of each model organism and their comparative species.

| Species                        | Genome version | Comparative species             | Genome version |
|--------------------------------|----------------|---------------------------------|----------------|
| <i>Caenorhabditis elegans</i>  | ce6            | <i>Caenorhabditis briggsae</i>  | cb3            |
| <i>Drosophila melanogaster</i> | dm3            | <i>Drosophila pseudoobscura</i> | dp3            |
| <i>Danio rerio</i>             | danRer7        | <i>Gasterosteus aculeatus</i>   | gasAcu1        |
| <i>Rattus norvegicus</i>       | m5             | <i>Mus musculus</i>             | mm10           |
| <i>Mus musculus</i>            | mm10           | <i>Oryctolagus cuniculus</i>    | oryCun2        |
| <i>Homo sapiens</i>            | GRch37         | <i>Macaca mulatta</i>           | rheMac3        |

### Determining miRNA evolutionary rates

Using BLAST alignments, miRNA precursors were mapped to their respective repeat-masked genomes. From the BLAST results, we identified the locations of miRNAs and sequences of non-miRNA regions with less than 5 nt mismatches. Flanking sequences of the same length at both ends of the miRNAs were retrieved. If a flanking sequence was overlapped by a clustered miRNA, we retrieved the non-overlapped sequence next to the clustered miRNA. Making use of the genome alignments from UCSC, we located homologous sequences of miRNA precursors

and flanking sequences in the compared species. Then, we used T-Coffee (Notredame et al., 2000) to align the miRNA precursors and flanking regions with their homologous sequences. From these alignments, we calculated the sequence divergence between the miRNA precursors and their homologous sequences using PAML (Yang, 1997), setting matrix K80 marked as  $D_p$ . We then calculated the sequence divergence between the flanking sequences and their homologous sequences, marked as  $D_f$ , in the same way. The miRNA evolutionary rate was calculated as  $R = D_p / D_f$ ;  $R = 1$  indicated neutral selection,  $R < 1$  indicated negative selection, and  $R > 1$  indicated positive selection. We sorted all of the miRNAs into five groups; miRNAs with rates  $< 0.3$  were clustered into Group I, 0.3-0.7 into Group II, 0.7-1.3 into Group III, 1.3-2.0 into Group IV, and  $> 2.0$  into Group V.

### **Assessing miRNA evolutionary rates using single nucleotide polymorphism (SNP) data**

We retrieved SNP sites for the model organisms from the SNP database, dbSNP (version 138) (Smigielski et al., 2000). Pre-miRNA sequences and flanking sequences in the same group were concatenated together. The ratio of the SNP frequency in the miRNA sequences to that in the flanking sequences was calculated to estimate the miRNA evolutionary rate. Then, Spearman's rank correlations were performed in the R program (R Core Team, 2013) to investigate associations between the SNP ratios and evolutionary rates.

### **Intergenic or intragenic miRNAs, clustered miRNAs, and miRNAs' secondary structure MFE**

By comparing the gene sites with the miRNA sites, we ascertained whether each miRNA was intergenic or intragenic. miRNAs were clustered together if the distance between two miRNAs was  $< 10$  kb, based on previous reports that the distance between two consecutive miRNAs is always  $< 5$  or  $> 50$  kb (Meunier et al., 2013). The MFE of the miRNA secondary structures was calculated using RNA-fold (Bomfünowerer et al., 2008). Using R programming, Spearman rank correlations were performed to investigate relationship between the miRNA evolutionary rate and the proportion of clustered miRNAs, and between MFEs and evolutionary rates.

### **miRNA target genes**

We collected two kinds of miRNA target gene sets: 1) experimentally validated target genes, and 2) *in silico* predicted target genes. Validated target gene data were downloaded from TarBase (Vergoulis et al., 2012), miRTarBase (Hsu et al., 2014), and miRecords (Xiao et al., 2009). The predicted gene data were downloaded from microRNA.org (Betel et al., 2008), TargetScan (Lewis et al., 2003), miRNAMap 2.0 (Hsu et al., 2008), and PITA (Kertesz et al., 2007). To reduce the probability of false positives, we only used the miRNA-target pairs that were predicted by no less than two tools.

### **miRNA functions**

To perform a functional analysis of the miRNAs we used the program DAVID (Huang et al., 2009), and the results obtained on molecular function, biological processes, KEGG pathways, and tissues were retrieved to calculate the proportions of functional miRNA-target pairs out of the

total miRNA-target pairs in each group. This analysis was performed to investigate the functional differentiation of miRNAs in species with different evolutionary rates.

### miRNA expression levels

We retrieved 60 small RNA sequencing libraries from the NCBI Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra>). The accession numbers and tissue information of all of the sequencing data are provided in [Table S1](#). All of the miRNA mature sequences were then aligned against the small RNA sequence library, allowing only one base mismatch. The relative pre-miRNA expression level in one library was measured as the maximum number of normalized reads mapped to its own mature sequence. These reads were normalized by dividing the number of mapped reads by the total number of mapped reads, multiplied by 1 million, and obtaining its  $\log_2$  value. The final miRNA expression level was estimated from the maximum expression levels in all of the libraries for each species. Finally, correlations between miRNA evolutionary rates and expression levels were investigated using R.

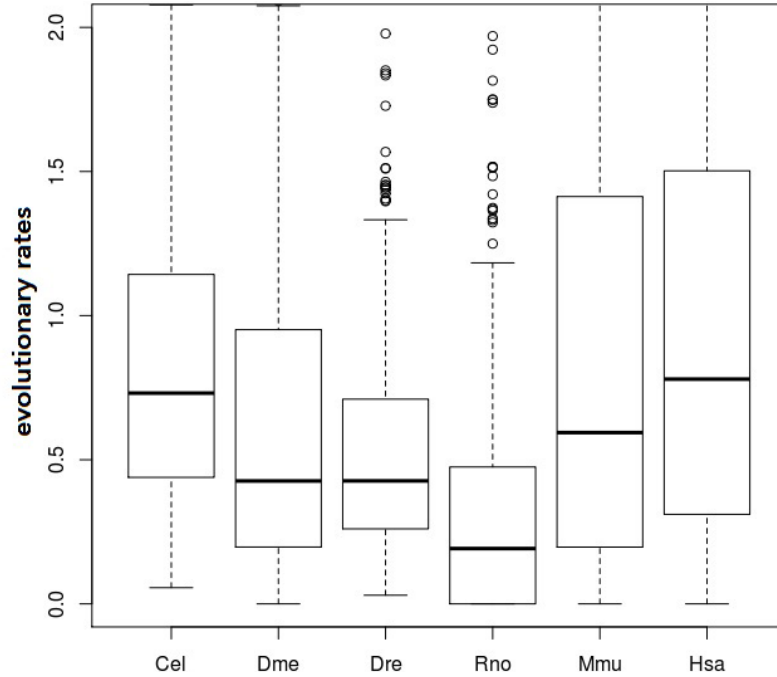
## RESULTS

### miRNA evolutionary rates

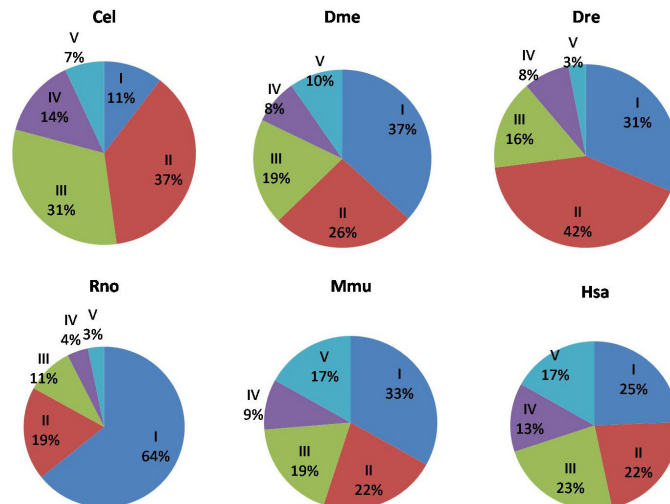
We used the same method that had been used in previous studies (Liang and Li, 2009; Qiu et al., 2010; Zhu et al., 2012; Jovelin, 2013), and classified the miRNAs into five groups for each species (see [Table S2](#) for details). The distributions and proportions of miRNAs in each group among the six organisms are presented in Figures 1 and 2. miRNAs in humans had the fastest evolutionary rates and those in rats had the slowest, which suggests that fast-evolving miRNAs in the rat have not yet been discovered. miRNAs in the worm had higher evolutionary rates than in the fruit fly. This suggests that the selective pressure in the worm may have been lower than in the fruit fly, so miRNAs would exhibit more polymorphisms. Previous studies have shown that fast-evolving miRNAs are younger, less conserved, closely related to species differentiations, and exhibit significant tissue specificity (Zhu et al., 2012). Consistent with this, we found that the proportions of fast-evolving miRNAs (Groups IV and V) increased in mice and humans, but decreased in worms and fruit flies, indicating that fast-evolving miRNAs could have played a more important role in complex species than in less complex species.

### miRNA evolutionary rates and SNPs

SNP sites obtained from the public databases were not suitable for calculating single-miRNA evolutionary rates because of the randomness and incompleteness of the SNP data. However, the ratios of SNPs in concatenated pre-miRNAs to SNPs in the same-length flanking regions could be used to estimate the evolutionary rates in a group (Liang and Li, 2009; Zhu et al., 2012). In general, SNP densities were much lower in the worm, zebrafish, and rat than in the other species. Therefore, we only used the SNP data from the fruit fly, mouse, and human to test groups classified by their evolutionary rates. The results revealed a consistency between the evolutionary rates and the SNP ratios, and suggested that fast-evolving miRNAs have more SNPs in their precursor sequences.

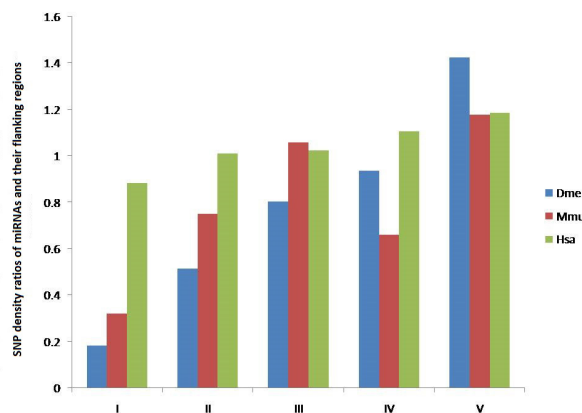


**Figure 1.** Boxplot of miRNA evolutionary rates of the six species studied. miRNAs in humans exhibited the highest evolutionary rate and those in rats the lowest, indicating that evolutionary rates did not simply correlate with organismal complexity. Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Rno, *Rattus norvegicus*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*.



**Figure 2.** Proportion of miRNA groups in the six model organisms studied. Slow-evolving miRNAs (Groups I and II) include major miRNAs in less complex organisms, such as the worm, fruit fly, and zebrafish; fast-evolving miRNAs expanded in more complex organisms, such as the mouse and human. Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Rno, *Rattus norvegicus*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*.

It is a known fact that polymorphisms and evolutionary rates are correlated (Liang and Li, 2009). However, this correlation exhibits different patterns in different species. We observed that in miRNA Group I, humans had a high SNP ratio (0.8823), although in the fruit fly and mouse the SNP ratios in Group I were only 0.1834 and 0.3210, respectively (Figure 3). This suggests that conserved miRNAs could be more diverse in humans than in other species. Among the five groups, the deviation was the lowest in humans. SNPs in pre-miRNAs could lead to variations in miRNA expression, SNPs in mature sequences could change the expression levels of target genes, and SNPs in seed regions could switch from one target gene to another (Wu et al., 2009; Lu and Clark, 2012). Humans have more complicated gene networks, and consequently greater flexibility to endure more SNPs, even in conserved miRNAs, which could have provided more diversity in the human population. Fast-evolving miRNAs (Group V) exhibited lower SNP density ratios in humans than in fruit flies. This suggests that fast-evolving miRNAs may perform important functions in higher species. High SNP densities are not found in functional miRNAs.



**Figure 3.** Single nucleotide polymorphism (SNP) density ratios of miRNAs and their flanking regions from the five miRNA groups in *Drosophila melanogaster* (Dme), *Mus musculus* (Mmu), and *Homo sapiens* (Hsa). Group I miRNAs in humans had a much higher ratio than those in the fruit fly or mouse, whereas Group V miRNAs in humans had a lower ratio than that in the fruit fly.

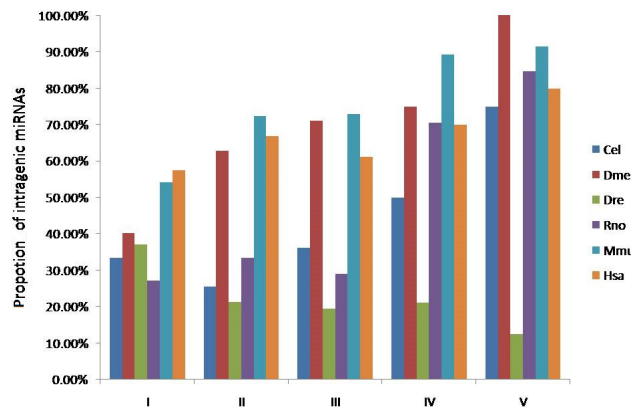
### miRNA evolutionary rates and genomic sources

miRNA genes have been shown to be intergenic or intragenic. When within a protein coding gene, they mainly reside in an intron. Only a small fraction of miRNAs have been located in the coding region of host genes (Berezikov et al., 2011). We found that fast-evolving miRNAs had the highest proportions of intragenic miRNAs, as found by Wang et al. (2011), except in the zebrafish (Figure 4). Young or fast-evolving miRNAs preferred to be intragenic, which in turn suggests that introns might be a good location for the emergence of new miRNAs. We also found that miRNAs were more likely to be parasitic in more complex species, such as the mouse and human.

Some miRNAs cluster together and share the same expression elements. The distance between any two consecutive miRNAs has always been <5 or >50 kb (Meunier et al., 2013). Therefore, we clustered miRNAs within 10 kb of each other (Figure 5). The results show that slow-evolving miRNAs are more likely to cluster, except in the zebrafish. Clustered miRNAs appear by gene duplication, and exhibit increased expression levels (Meunier et al., 2013; Sun et al.,



2013). This suggests that old or slow-evolving miRNAs were created mainly by gene duplication and assumed functional roles by increasing their expression levels, thus displaying the early evolutionary patterns of miRNAs.



**Figure 4.** Proportion of intragenic miRNAs in each group in the six model species studied. In all of the species except the zebrafish, miRNAs in fast-evolving groups were more likely to be intragenic; miRNAs in fast-evolving groups were less likely to be intragenic in the zebrafish. Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Rno, *Rattus norvegicus*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*.

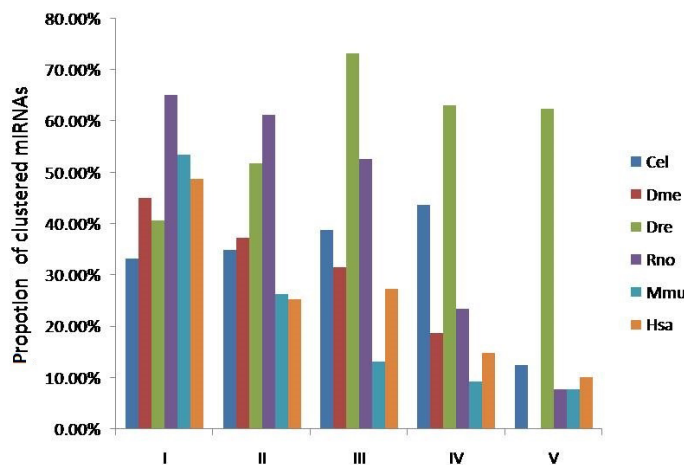
We further observed that regardless of the proportion of intragenic miRNAs or clustered miRNAs, the trend in the zebrafish was the opposite of that observed in the other five species. Zebrafish had a small proportion of intragenic miRNAs but a large proportion of clustered miRNAs, particularly in the fast-evolving miRNA groups. We also found that zebrafish had a number of miRNAs that share the same precursor sequences (74 in 319, 23.2%), which was much higher than in the other species (all below 4%). In addition, we found 38 non-miRNA regions with less than 5 nt mismatches to the precursor miRNA sequences, which were also higher than in the two simpler species, the fruit fly and the worm (6 and 16, respectively). All of these data indicate a different evolutionary pattern of miRNAs in the zebrafish, and that new miRNAs were mainly created by duplication. This pattern could have occurred because of the high rate of duplication in the early vertebrate genome (Gu et al., 2009; Sun et al., 2013), which could also have increased duplication errors and provided genetic sources for later miRNA evolution.

### miRNA evolutionary rates and MFE

Research on the fruit fly has revealed that the robustness of miRNAs is due to the acquisition of new functions (Price et al., 2011). miRNA MFEs are stable in miRNAs with different evolutionary rates and ages (Shen et al., 2011; Zhu et al., 2012). To determine whether these principles are universally applicable to all miRNAs, we performed Pearson's correlation tests to ascertain whether associations existed between miRNA evolutionary rates and MFEs in all six model species. We found a significant correlation between evolutionary rates and MFEs ( $P < 0.05$ ) in the fruit fly, mouse, and rat. miRNA evolution in the zebrafish is undergoing a significant duplication, as mentioned earlier. Therefore, many miRNAs in the zebrafish share similar MFEs. The absence of a correlation in the human genome suggests that human miRNAs could have



evolved to stabilize structures that enable them to perform important functions. We found that Group I miRNAs had the lowest standard variation in MFEs in all six species (data not shown). This suggests that conserved miRNAs are more likely to have conserved and stable secondary structures, rather than to gain lower MFEs, for recognition by the miRNA genesis pathway.



**Figure 5.** Proportion of clustered miRNAs in each group in the six model species studied. In all of the species except the zebrafish, miRNAs in fast-evolving groups were less likely to be clustered; miRNAs in fast-evolving groups were more likely to be intragenic in the zebrafish. Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Rno, *Rattus norvegicus*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*.

To investigate differences in miRNA MFEs across species, we performed pairwise Wilcoxon rank sum tests between the six species. As shown in Table 2, significant differences existed between most values. Interestingly, mouse miRNAs had high MFEs (not shown in Table 2), and were not significantly different to the fruit fly. Humans were significantly different to the other species, except for the worm and zebrafish. Next, we compared mammalian miRNA MFEs with non-mammalian miRNA MFEs, and vertebrate miRNA MFEs with non-vertebrate miRNA MFEs, and performed two-sample Wilcoxon rank sum tests. The corresponding P values were 0.05549 and 9.723E-6, respectively. It is evident from these results that significant differences exist between vertebrates and non-vertebrates, which suggest that the efficiency of precursor miRNA recognition by the miRNA genesis pathway may have evolved in early vertebrates.

**Table 2.** Wilcoxon rank sum test P values for minimized free energy between two samples among the six species studied.

|     | Cel | Dme      | Dre       | Rno       | Mmu       | Hsa       |
|-----|-----|----------|-----------|-----------|-----------|-----------|
| Cel |     | 0.009518 | 0.06933   | 1.951E-05 | 0.01594   | 0.1708    |
| Dme |     |          | 3.288E-07 | 1.159E-15 | 0.3717    | 2.19E-07  |
| Dre |     |          |           | 0.002003  | 3.725E-09 | 0.5242    |
| Rno |     |          |           |           | 2.20E-16  | 1.896E-06 |
| Mmu |     |          |           |           |           | 1.69E-13  |

Cel = *Caenorhabditis elegans*; Dme = *Drosophila melanogaster*; Dre = *Danio rerio*; Rno = *Rattus norvegicus*; Mmu = *Mus musculus*; Hsa = *Homo sapiens*.

## miRNA evolutionary rates and target gene functions

In a previous study, we found that miRNAs have distinct target preferences across species (Mao et al., 2014). To investigate whether miRNA evolutionary rates play a role in this, we collected experimentally validated and *in silico* predicted miRNA target genes for all six species. We annotated the functions of miRNA-target pairs using the program DAVID, and calculated the proportion of every function-related miRNA-target pair in the five groups among the six species. Pearson's correlation tests were used to elucidate the relationships between these proportions and their respective evolutionary rates. From these analyses, we were able to infer whether miRNAs in the same group regulated certain functions.

Since our previous study indicated that higher-species miRNAs were more likely to target genes performing "binding" functions, we specifically investigated this relationship in this study. However, no correlation was found between miRNA evolutionary rates and "binding" related to miRNA-target pair proportions. The proportions of some functionally related miRNA-target pairs were significantly positively correlated ( $P < 0.05$ ) with miRNA evolutionary rates among human *in silico* target data, such as "regulation of myeloid cell differentiation", "positive regulation of gliogenesis", and "positive regulation of translation". However, the proportions of some functional miRNAs, such as "embryo gut development", "cellular response to heat", and "organic anion transport", were significantly negatively correlated ( $P < 0.05$ ) with miRNA evolutionary rates. In the validated data, we found significant, positive correlations ( $P < 0.05$ ) between the proportions of functional miRNAs, such as "eye morphogenesis", "cell death", and "regulation of Ras protein signal transduction", and miRNA evolutionary rates. "Regulation of vesicle-mediated transport", "cell adhesion", and "energy derivation by oxidation of organic compounds" were significantly negatively correlated ( $P < 0.05$ ) with evolutionary rates.

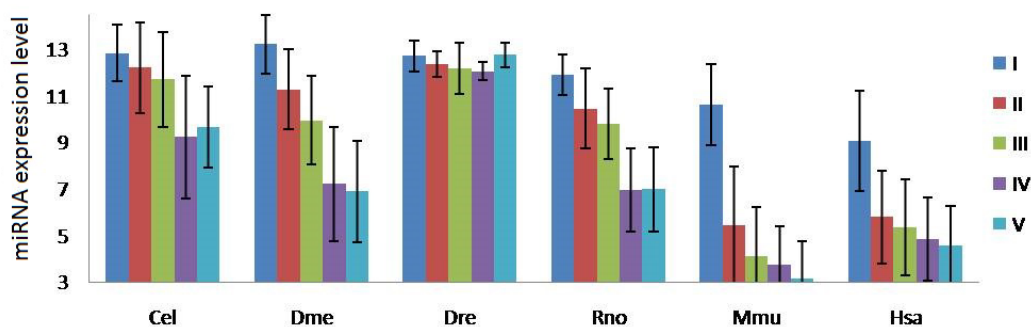
We also found significant correlations ( $P < 0.05$ ) in tissue-related miRNA-target data in each miRNA group in humans. Among the validated and *in silico* target sets, fast-evolving miRNAs were more likely to target genes expressed in the "retina", "cerebellum", "B-cell", "neuron", "pancreas", etc., many of which are nervous system components. This might have contributed to the complexity of *Homo sapiens*, particularly in nerve tissue. Our results confirmed that fast-evolving or young miRNAs may participate in specific mechanisms and increase species complexity.

## miRNA evolutionary rates and expression levels

Using small sequencing data from the NCBI SRA database, we calculated relative miRNA expression levels (see Methods). To minimize bias from spatial and temporal variations in miRNA expression patterns, we chose the maximum normalized level in 10 different tissues to evaluate miRNA expression in each species. The average expression levels of each group in the six species are shown in Figure 6. We observed significant negative correlations between miRNA evolutionary rates and expression levels in all of the species, except the zebrafish. Previously, it was shown that fast-evolving miRNAs have low expression levels in the worm, fruit fly, and human (Liang and Li, 2009; Shen et al., 2011; Zhu et al., 2012; Jovelin, 2013); our results indicate that this is not applicable to all species. Details of the expression levels are provided in [Table S3](#).

To investigate further, we performed Pearson's correlation tests between miRNA evolutionary rates and expression levels (Table 3). The results supported our hypothesis that a negative correlation occurs in all species, except the zebrafish. Because the zebrafish is in the lineage of the early vertebrates, it probably experienced significant genomic duplications.

Therefore, most miRNAs probably exhibited lower evolutionary rates, as we observed in Groups I and II. Pre-miRNAs sharing same sequences in the zebrafish genome probably increased the number of miRNA copies and caused a high level of expression.



**Figure 6.** miRNA expression levels in each group in the six species studied. Expression levels were normalized for each organism; therefore, the expression levels of different groups can only be compared within the same species, rather than between different organisms. Except in zebrafish, slow-evolving miRNAs exhibited high expression levels, while fast-evolving miRNAs exhibited low expression levels. Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Rno, *Rattus norvegicus*; Mmu, *Mus musculus*; Hsa, *Homo sapiens*.

**Table 3.** Pearson's correlation test P values and rho values of miRNA evolutionary rates and expression levels in the six species studied.

|         | Cel      | Dme      | Dre     | Rno      | Mmu      | Hsa      |
|---------|----------|----------|---------|----------|----------|----------|
| P value | 1.19E-02 | 1.41E-11 | 0.3847  | 6.94E-14 | 2.20E-16 | 3.11E-13 |
| rho     | -0.2338  | -0.4584  | -0.0544 | -0.3627  | -0.3294  | -0.2054  |

Cel = *Caenorhabditis elegans*; Dme = *Drosophila melanogaster*; Dre = *Danio rerio*; Rno = *Rattus norvegicus*; Mmu = *Mus musculus*; Hsa = *Homo sapiens*.

In addition, we obtained very low P values in humans, and a rho value (indicating the correlation) close to zero. This indicates the existence of a significant correlation in humans, but the variations in expression due to evolutionary rates were the lowest in humans among the six species. This suggests that differences in expression levels between young and old miRNAs may have been decreased, and that young miRNAs could have already performed important functions in humans. Iwama et al. (2013) reported an increase in miRNA expression that was specific to the initial hominoid lineage. Therefore, fast-evolving miRNAs in humans may have been subjected to high, positive selection pressure, which may not have been the case in the other five species.

## DISCUSSION

miRNA evolution can be divided into two phases (Chen and Rajewsky, 2007). Due to the length of miRNA seeds, nascent miRNAs can target multiple genes, which may be deleterious to an organism. Therefore, new miRNAs are, in general, expressed at low levels and are subjected to sequence divergence. In the second phase, malignant miRNAs would be purified by natural selection, while other miRNAs would perform important functions and join the gene networks. Then, selected miRNAs would increase their expression levels and gain time and space specificities. Such miRNA sequences tend to be conserved.

The SNP data we collected were consistent with the evolutionary rates. Additionally, slow-evolving miRNAs in higher species had more SNPs in their precursor sequences. This could be due to the complex gene regulation networks in higher species, which provide more buffering capacity and allow more SNPs in miRNA sequences. In agreement with the results of a previous study (Lu and Clark, 2012), we found that SNPs in miRNAs contribute to variations in target gene expression in complex species. Therefore, miRNAs probably increase both genotypic and phenotypic diversity in these species. However, in less complex species, SNPs seldom occur in conserved miRNAs due to the weak buffering ability of simple gene networks. Hence, these miRNAs probably perform “canalization” functions (Ebert and Sharp, 2012) in these species.

Analysis of the miRNA locations showed that slow-evolving miRNAs are usually clustered. This indicates that ancient miRNAs mainly evolved by gene duplication. Fast-evolving miRNAs tend to be intragenic, suggesting that *de novo* miRNAs tend to emerge from the introns of host genes. Zebrafish exhibited a unique pattern in these comparisons. In this species, slow-evolving miRNAs were the majority, and a large number of same sequence pre-miRNAs and miRNA-like regions were found. Correlations between evolutionary rates and intragenic and clustered miRNAs exhibited the opposite trends to those found in the other species, indicating that gene duplication could have been the main mechanism of miRNA evolution in this species, and led to the large expansion of miRNA copies and an increase in miRNA expression. Accumulated errors in the miRNA duplications probably became the source for later evolution.

Finally, the molecular functions of target genes could have decided miRNA functions. Young or fast-evolving miRNAs were mainly expressed at low levels. Their preferences for certain biological processes indicate the direction of miRNA evolution, and their expression levels indicated whether the miRNAs had integrated into gene networks. In humans, slow-evolving miRNAs had higher expression levels and lower SNP density ratios than in other species, except zebrafish, demonstrating that these miRNAs could be rapidly integrated into gene networks, thereby accelerating human evolution. Because this study is based on previously reported data in public databases, it is biased towards well-studied species and genes. However, it provides new insights into miRNA evolutionary rates and miRNA functional analysis.

## Conflicts of interest

All authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We would like to extend our thanks to Professor Cao who supervised us.

## REFERENCE

- Berezikov E (2011). Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* 12: 846-860. <http://dx.doi.org/10.1038/nrg3079>
- Berezikov E, Robine N, Samsonova A, Westholm JO, et al. (2011). Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res.* 21: 203-215. <http://dx.doi.org/10.1101/gr.116657.110>
- Betel D, Wilson M, Gabow A, Marks DS, et al. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 36: D149-D153. <http://dx.doi.org/10.1093/nar/gkm995>
- Bompfünnewerer AF, Backofen R, Bernhart SH, Hertel J, et al. (2008). Variations on RNA folding and alignment: lessons from Benasque. *J. Math. Biol.* 56: 129-144. <http://dx.doi.org/10.1007/s00285-007-0107-5>

- Chen K and Rajewsky N (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8: 93-103. <http://dx.doi.org/10.1038/nrg1990>
- Christodoulou F, Raible F, Tomer R, Simakov O, et al. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463: 1084-1088. <http://dx.doi.org/10.1038/nature08744>
- Cui Q, Yu Z, Purisima EO and Wang E (2007). MicroRNA regulation and interspecific variation of gene expression. *Trends Genet.* 23: 372-375. <http://dx.doi.org/10.1016/j.tig.2007.04.003>
- Ebert MS and Sharp PA (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149: 515-524. <http://dx.doi.org/10.1016/j.cell.2012.04.005>
- Gu X, Su Z and Huang Y (2009). Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. *J. Exp. Zool. B Mol. Dev. Evol.* 312B: 164-170. <http://dx.doi.org/10.1002/jez.b.21273>
- Guerra-Assunção JA and Enright AJ (2012). Large-scale analysis of microRNA evolution. *BMC Genomics* 13: 218. <http://dx.doi.org/10.1186/1471-2164-13-218>
- Hsu SD, Chu CH, Tsou AP, Chen SJ, et al. (2008). miRNAmap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res.* 36: D165-D169. <http://dx.doi.org/10.1093/nar/gkm1012>
- Hsu SD, Tseng YT, Shrestha S, Lin YL, et al. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 42: D78-D85. <http://dx.doi.org/10.1093/nar/gkt1266>
- Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Isik M, Korswagen HC and Berezikov E (2010). Expression patterns of intronic microRNAs in *Caenorhabditis elegans*. *Silence* 1: 5. <http://dx.doi.org/10.1186/1758-907X-1-5>
- Iwama H, Kato K, Imachi H, Murao K, et al. (2013). Human microRNAs originated from two periods at accelerated rates in mammalian evolution. *Mol. Biol. Evol.* 30: 613-626. <http://dx.doi.org/10.1093/molbev/mss262>
- Jovelin R (2013). Pleiotropic constraints, expression level, and the evolution of miRNA sequences. *J. Mol. Evol.* 77: 206-220. <http://dx.doi.org/10.1007/s00239-013-9588-6>
- Kertesz M, Iovino N, Unnerstall U, Gaul U, et al. (2007). The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39: 1278-1284. <http://dx.doi.org/10.1038/ng2135>
- Kozomara A and Griffiths-Jones S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42: D68-D73. <http://dx.doi.org/10.1093/nar/gkt1181>
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, et al. (2003). Prediction of mammalian microRNA targets. *Cell* 115: 787-798. [http://dx.doi.org/10.1016/S0092-8674\(03\)01018-3](http://dx.doi.org/10.1016/S0092-8674(03)01018-3)
- Liang H and Li WH (2009). Lowly expressed human microRNA genes evolve rapidly. *Mol. Biol. Evol.* 26: 1195-1198. <http://dx.doi.org/10.1093/molbev/msp053>
- Liu N, Okamura K, Tyler DM, Phillips MD, et al. (2008). The evolution and functional diversification of animal microRNA genes. *Cell Res.* 18: 985-996. <http://dx.doi.org/10.1038/cr.2008.278>
- Lu J and Clark AG (2012). Impact of microRNA regulation on variation in human gene expression. *Genome Res.* 22: 1243-1254. <http://dx.doi.org/10.1101/gr.132514.111>
- Mao X, Li L and Cao Y (2014). Evolutionary comparisons of miRNA regulation system in six model organisms. *Genetica* 142: 109-118. <http://dx.doi.org/10.1007/s10709-014-9758-5>
- Meunier J, Lemoine F, Soumillon M, Liechti A, et al. (2013). Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 23: 34-45. <http://dx.doi.org/10.1101/gr.140269.112>
- Notredame C, Higgins DG and Heringa J (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302: 205-217. <http://dx.doi.org/10.1006/jmbi.2000.4042>
- Nozawa M, Miura S and Nei M (2010). Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biol. Evol.* 2: 180-189. <http://dx.doi.org/10.1093/gbe/evq009>
- Pederson T (2010). Regulatory RNAs derived from transfer RNA? *RNA* 16: 1865-1869. <http://dx.doi.org/10.1261/rna.2266510>
- Price N, Cartwright RA, Sabath N, Graur D, et al. (2011). Neutral evolution of robustness in *Drosophila* microRNA precursors. *Mol. Biol. Evol.* 28: 2115-2123. <http://dx.doi.org/10.1093/molbev/msr029>
- Qiu C, Wang J, Yao P, Wang E, et al. (2010). microRNA evolution in a human transcription factor and microRNA regulatory network. *BMC Syst. Biol.* 4: 90. <http://dx.doi.org/10.1186/1752-0509-4-90>
- <eref>R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.</eref>
- Shen Y, Lv Y, Huang L, Liu W, et al. (2011). Testing hypotheses on the rate of molecular evolution in relation to gene expression using microRNAs. *Proc. Natl. Acad. Sci. USA* 108: 15942-15947. <http://dx.doi.org/10.1073/pnas.1110098108>
- Smigielski EM, Sirotkin K, Ward M and Sherry ST (2000). dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.* 28: 352-355. <http://dx.doi.org/10.1093/nar/28.1.352>

- Sun J, Gao B, Zhou M, Wang ZZ, et al. (2013). Comparative genomic analysis reveals evolutionary characteristics and patterns of microRNA clusters in vertebrates. *Gene* 512: 383-391. <http://dx.doi.org/10.1016/j.gene.2012.09.102>
- Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, et al. (2012). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.* 40: D222-D229. <http://dx.doi.org/10.1093/nar/gkr1161>
- Wang QH, Zhou M, Sun J, Ning SW, et al. (2011). Systematic analysis of human microRNA divergence based on evolutionary emergence. *FEBS Lett.* 585: 240-248. <http://dx.doi.org/10.1016/j.febslet.2010.11.053>
- Wheeler BM, Heimberg AM, Moy VN, Sperling EA, et al. (2009). The deep evolution of metazoan microRNAs. *Evol. Dev.* 11: 50-68. <http://dx.doi.org/10.1111/j.1525-142X.2008.00302.x>
- Wu CI, Shen Y and Tang T (2009). Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res.* 19: 734-743. <http://dx.doi.org/10.1101/gr.084640.108>
- Xiao F, Zuo Z, Cai G, Kang S, et al. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* 37: D105-D110. <http://dx.doi.org/10.1093/nar/gkn851>
- Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555-556.
- Yuan Z, Sun X, Liu H and Xie J (2011). MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. *PLoS One* 6: e17666. <http://dx.doi.org/10.1371/journal.pone.0017666>
- Zhu Y, Skogerbø G, Ning Q, Wang Z, et al. (2012). Evolutionary relationships between miRNA genes and their activity. *BMC Genomics* 13: 718. <http://dx.doi.org/10.1186/1471-2164-13-718>

## Supplementary Material

**Table S1.** Lists of the SRA accession numbers of small RNA sequencing libraries used in this study.

**Table S2.** Lists of the evolutionary rates of miRNAs in the six species studied.

**Table S3.** Lists of the expression levels of miRNAs in the six species studied.

[http://www.geneticsmr.com/year2016/vol15-1/pdf/gmr7216\\_supplementary.pdf](http://www.geneticsmr.com/year2016/vol15-1/pdf/gmr7216_supplementary.pdf)