



Genetic Analysis of Loop Sequences in the Let-7 Gene Family Reveal a Relationship between Loop Evolution and Multiple IsomiRs

Tingming Liang¹, Chen Yang¹, Ping Li¹, Chang Liu¹, Li Guo^{2*}

1 Jiangsu Key Laboratory for Molecular and Medical Biotechnology, College of Life Science, Nanjing Normal University, Nanjing, 210023, China, **2** Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 211166, China

Abstract

While mature miRNAs have been widely studied, the terminal loop sequences are rarely examined despite regulating both primary and mature miRNA functions. Herein, we attempted to understand the evolutionary pattern of loop sequences by analyzing loops in the let-7 gene family. Compared to the stable miRNA length distributions seen in most metazoans, higher metazoan species exhibit a longer length distribution. Examination of these loop sequence length distributions, in addition to phylogenetic tree construction, implicated loop sequences as the main evolutionary drivers in miRNA genes. Moreover, loops from relevant clustered miRNA gene families showed varying length distributions and higher levels of nucleotide divergence, even between homologous pre-miRNA loops. Furthermore, we found that specific nucleotides were dominantly distributed in the 5' and 3' terminal loop ends, which may contribute to the relatively precise cleavage that leads to a stable isomiR expression profile. Overall, this study provides further insight into miRNA processing and maturation and further enriches our understanding of miRNA biogenesis.

Citation: Liang T, Yang C, Li P, Liu C, Guo L (2014) Genetic Analysis of Loop Sequences in the Let-7 Gene Family Reveal a Relationship between Loop Evolution and Multiple IsomiRs. *PLoS ONE* 9(11): e113042. doi:10.1371/journal.pone.0113042

Editor: John J. Rossi, Beckman Research Institute of the City of Hope, United States of America

Received: July 26, 2014; **Accepted:** October 11, 2014; **Published:** November 14, 2014

Copyright: © 2014 Liang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by National Basic Research Program of China (973 Program) (No. 2012CB947600 and 2013CB911600), the National Natural Science Foundation of China (No. 31171137, 31271261, 31401009 and 61301251), the Program for New Century Excellent Talents in University by the Chinese Ministry of Education (No. NCET-11-0990), the Program for the Top Young Talents by the Organization Department of the CPC Central Committee, the Collaborative Innovation Center for Cardiovascular Disease Translational Medicine (Nanjing Medical University), the Research Fund for the Doctoral Program of Higher Education of China (No. 20133234120009), the National Natural Science Foundation of Jiangsu (No. BK20130885), the Natural Science Foundation of the Jiangsu Higher Education Institutions (No. 12KJB360001 and 13KJB330003), NSFC for Talents Training in Basic Science (J1103507), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: gl8008@163.com

Introduction

MicroRNAs (miRNAs), a class of small non-coding RNA, are widely studied as crucial regulatory molecules able to modulate broad regulatory networks at the post-transcriptional levels [1,2]. miRNA is generated from primary miRNA (pri-miRNA) and precursor miRNA (pre-miRNA) in animals [3], with the pre-miRNA presenting a stable stem-loop structure. Both of the two arm products, miR-#-5p and miR-#-3p, have been reported to form mature and functional miRNAs [4–8], with the loop sequences connecting miR-#-5p and miR-#-3p in the stem-loop structure typically degraded during miRNA biogenesis. These transitory intermediates are rarely examined due to an unseen direct role in the miRNA regulation process. Instead, most studies have focused on the functional and regulatory roles of miRNAs during miRNA-mRNA recognition as it relates to expression or translational repression. However, indeed, both pri-miRNA and pre-miRNA may contribute to the regulatory process [9]. Specifically, the loop nucleotides may tune and alter miRNA activity, controlling the processing precision during the miRNA maturation process [9].

Recently, the potential impact of hairpin loops has been examined, with short hairpin RNA (shRNA) loops possibly influencing effectivity [10–12] and alternative loop conformations in miRNAs potentially affecting expression [13,14]. The loop sequence may also affect Dicer recognition and possibly specificity, thus affect miRNA cleavage during the maturation process [11,15,16]. As mentioned above, the loop sequence in pre-miRNAs has an important role in regulating the activities and specificities of related molecules that can facilitate Drosha and Dicer, with a mutation in the loop possibly affecting miRNA processing [17,18]. Furthermore, miRNAs have been a key focal point because a miRNA locus can yield multiple sequences have diverse 5' and/or 3' ends and expression levels [19–25]. These miRNA variants, or physiological miRNA isoforms (isomiRs), may be mainly generated by imprecise Drosha and Dicer cleavage during pre-miRNA processing [22]. Additionally, the loop sequences may also impact pre-miRNA processing by affecting Drosha and/or Dicer directly or indirectly. Even though pre-miRNA loop sequences have been shown to directly or indirectly impact Drosha and Dicer during miRNA processing, few studies have performed an evolutionary analysis of the loop sequences,

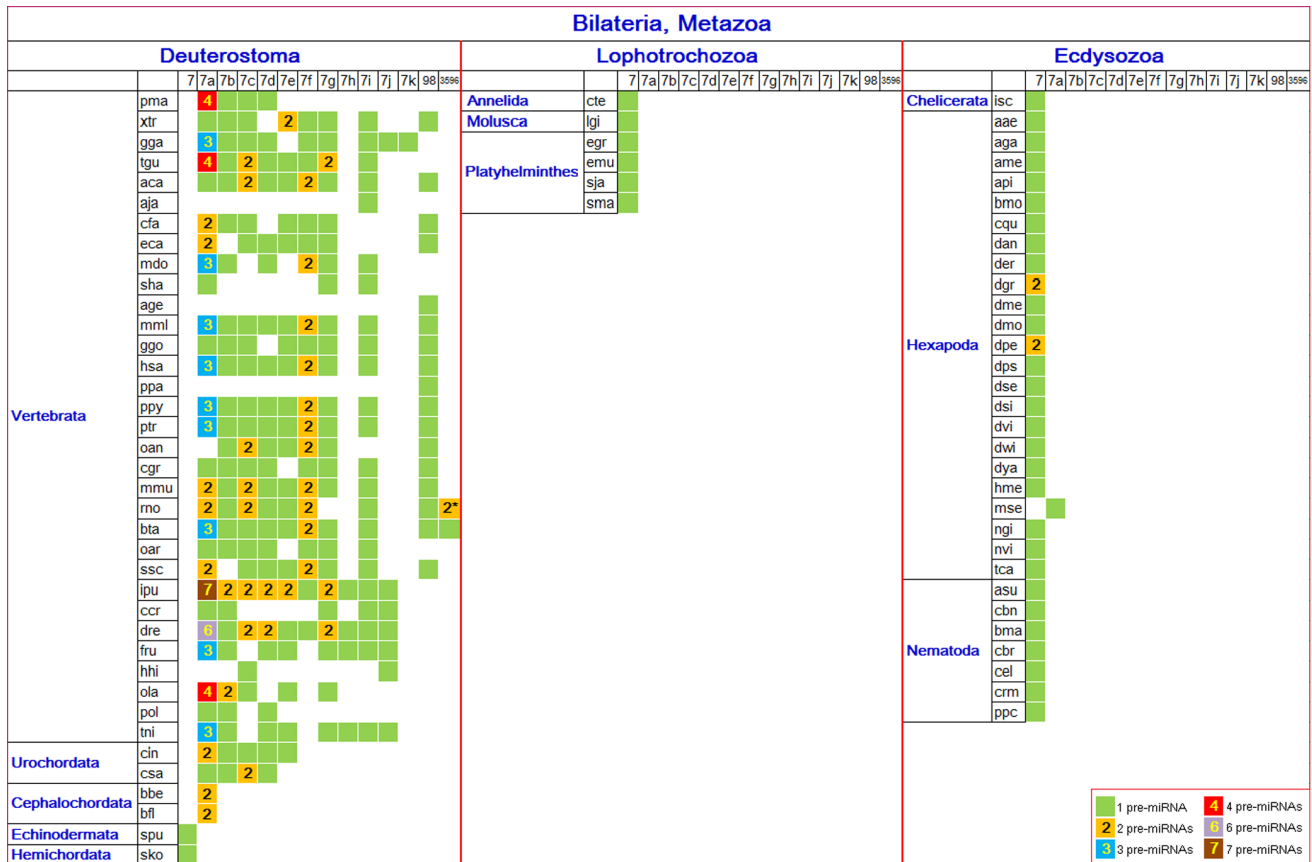


Figure 1. The distribution of the let-7 family among metazoan animals. The let-7 gene was named and homologous miRNAs were further found and named. These miRNA names were consistent with the current annotations in the miRBase database. "2*" indicates that there are two genes, rno-mir-3596b and rno-mir-3596d, and not two pre-miRNAs for the mature miRNA. doi:10.1371/journal.pone.0113042.g001

especially one focused on homology and clustered miRNAs across different animal species.

Herein, we attempt to perform an evolutionary analysis of the loop sequences in let-7 and locate related miRNAs across different animal species. The let-7 gene family has been widely detected in metazoans, with its associated miRNAs thought to have co-expanded with the HOX gene clusters [26]. A series of homologous miRNAs can be found in a miRNA gene family, including multicopy pre-miRNAs from specific animal species. Simultaneously, some members can be located in a cluster with other related miRNAs, with miRNAs within a close proximity on a chromosome being co-transcribed as a polycistronic transcript [27–30]. Therefore, a classical miRNA gene family and related miRNAs are typically selected to track and reveal the evolutionary patterns of the ever present yet ignored loop sequences. Thus, this study will examine the roles of loops during miRNA maturation and processing.

Materials and Methods

MiRNA members in the let-7 gene family, to include homologous miRNAs and miRNAs from different animal species, were obtained from the miRBase database (http://www.mirbase.org/cgi-bin/mirna_summary.pl?fam=MIPF0000002). According to the location distributions, some were located within a gene cluster and all pre-miRNA and miRNA let-7 sequences and other related miRNAs were simultaneously collected.

pre-miRNA loop sequences were collected, including miR-#-5p and miR-#-3p sequences. If the miR-#-5p or miR-#-3p sequences were not annotated in the miRBase database, the complementary antisense miRNA consensus sequences was collected following alignment with the Clustal X 2.0 software [31]. Despite the sequence diversity of miRNAs among different animal species, miRNAs are phylogenetically conserved, particularly in the functional seed sequences (nucleotides 2–8). miRNAs can also vary in their length distributions, which can be attributed to being derived from various 3' ends. Additionally, multiple isomiRs have been widely detected from miRNA loci, which indicates that flexible 5' and 3' ends mediate these loop sequences. The present study mainly focused on the core loop sequences. The 5' and 3' ends of loop sequences are cleavage sites of Dicer, with previous studies showing that imprecise cleavage by Drosha and Dicer can lead isomiRs generation via biased cleavage [24,32,33]. Based on this phenomenon, in addition to biased cleavage, the continuous 5, 3 and 1 nucleotides in the 5' and 3' ends of the loop sequences were specifically analyzed. To understand the effect of changes in the pre-miRNA loop sequences, the minimum free energy of relevant pre-miRNA sequences were predicted using the RNAfold webserver [34].

Phylogenetic trees were constructed using pre-miRNAs and loop sequences were constructed based on the Neighbor-Joining (NJ) method using the MEGA 6.06 software [35]. Evolutionary relationships among loop sequences within the hsa-let-7 gene family were reconstructed using SplitsTree 4.10 [36]. Sequence

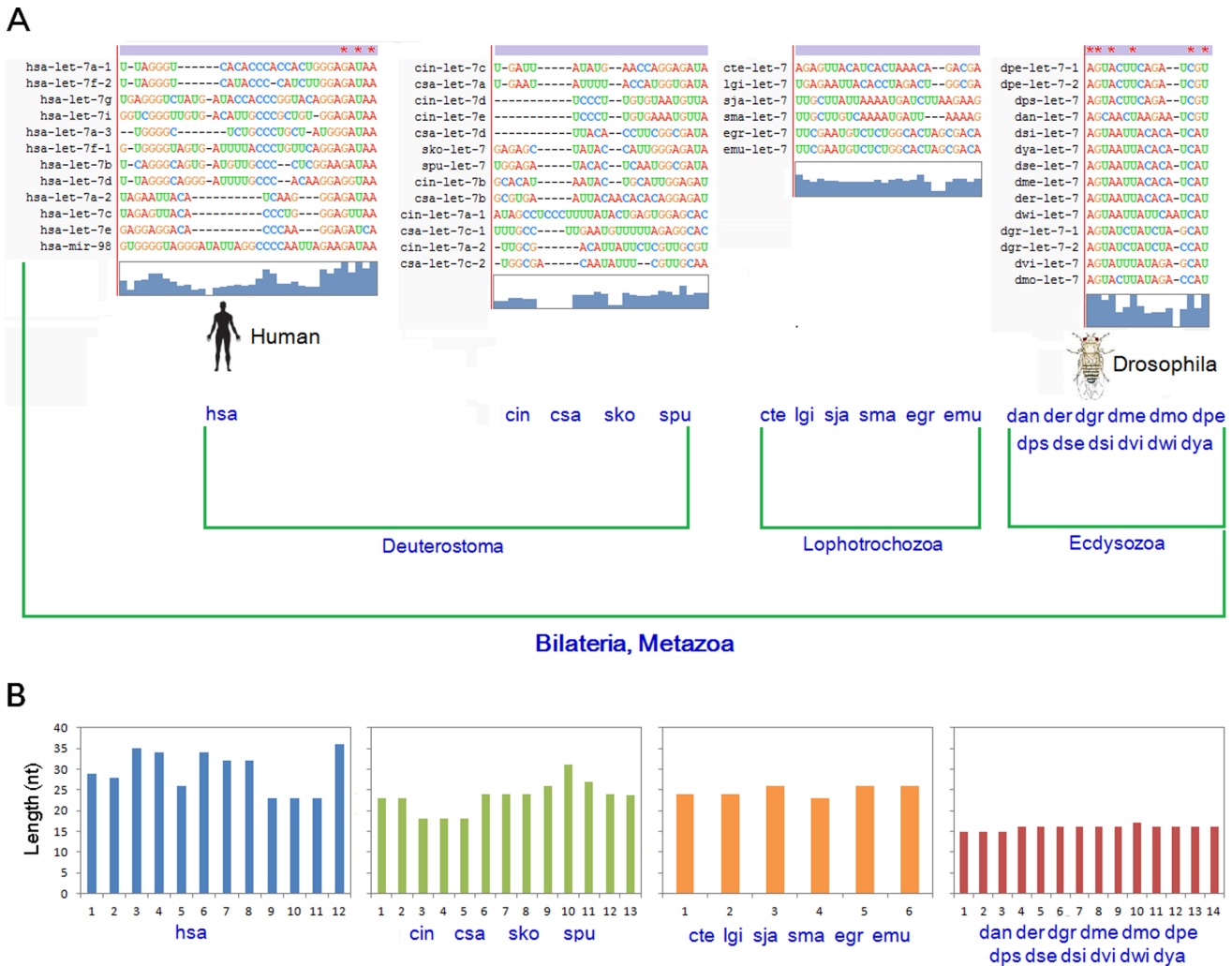


Figure 2. Sequence diversity and length distribution of the loop sequences among different animal species. (aae, *Aedes aegypti*; aca, *Anolis carolinensis*; aga, *Anopheles gambiae*; age, *Ateles geoffroyi*; aja, *Artibeus jamaicensis*; ame, *Apis mellifera*; api, *Acyrtosiphon pisum*; asu, *Ascaris suum*; bbe, *Branchiostoma belcheri*; bfl, *Branchiostoma floridae*; bma, *Brugia malayi*; bmo, *Bombyx mori*; bta, *Bos taurus*; cbn, *Caenorhabditis brenneri*; cbr, *Caenorhabditis briggsae*; ccr, *Cyprinus carpio*; cel, *Caenorhabditis elegans*; cfa, *Canis familiaris*; cgr, *Cricetulus griseus*; cin, *Ciona intestinalis*; cqu, *Culex quinquefasciatus*; crm, *Caenorhabditis remanei*; csa, *Ciona savignyi*; cte, *Capitella teleta*; dan, *Drosophila ananassae*; der, *Drosophila erecta*; dgr, *Drosophila grimshawi*; dme, *Drosophila melanogaster*; dmo, *Drosophila mojavensis*; dpe, *Drosophila persimilis*; dps, *Drosophila pseudoobscura*; dre, *Danio rerio*; dse, *Drosophila sechellia*; dsi, *Drosophila simulans*; dvi, *Drosophila virilis*; dwi, *Drosophila willistoni*; dya, *Drosophila yakuba*; eca, *Equus caballus*; egr, *Echinococcus granulosis*; emu, *Echinococcus multilocularis*; fru, *Fungus rubripes*; gga, *Gallus gallus*; ggo, *Gorilla gorilla*; hhi, *Hippoglossus hippoglossus*; hme, *Heliconius melpomene*; hsa, *Homo sapiens*; ipu, *Ictalurus punctatus*; isc, *Ixodes scapularis*; lgi, *Lottia gigantea*; mdo, *Monodelphis domestica*; mml, *Macaca mulatta*; mmu, *Mus musculus*; mse, *Manduca sexta*; ngi, *Nasonia giraulti*; nvi, *Nasonia vitripennis*; oan, *Ornithorhynchus anatinus*; oar, *Ovis aries*; ola, *Oryzias latipes*; pma, *Petromyzon marinus*; pol, *Paralichthys olivaceus*; ppa, *Pan paniscus*; ppc, *Pristionchus pacificus*; ppy, *Pongo pygmaeus*; ptr, *Pan troglodytes*; rno, *Rattus norvegicus*; sha, *Sarcophilus harrisii*; sja, *Schistosoma japonicum*; sko, *Saccoglossus kowalevskii*; sma, *Schistosoma mansoni*; spu, *Strongylocentrotus purpuratus*; ssc, *Sus scrofa*; tca, *Tribolium castaneum*; tgu, *Taeniopygia guttata*; tni, *Tetraodon nigroviridis*; xtr, *Xenopus tropicalis*).

doi:10.1371/journal.pone.0113042.g002

logos of relevant loop sequences across homologous miRNAs or clustered miRNAs were analyzed using the WebLogo program (<http://weblogo.berkeley.edu/logo.cgi>) [37]. An un-paired *t* test was used to estimate length distribution differences between the two groups and the chi-square (χ^2) test was performed to estimate differences in loop nucleotide composition. Length distribution differences among multiple groups were estimated using ANOVA analysis ($P < 0.05$), followed by a *q* test if a significant *P*-value was obtained. Adjusted *P*-values were obtained based on pairwise comparison. All the relevant statistical analysis was performed using the Stata software (Version 11.0).

Results

Overview of the distribution and sequence characteristics of let-7 loop sequences

Let-7 gene family members have been found in 75 metazoan animal species (Figure 1). Multiple homologous miRNA genes, some being multicopy pre-miRNAs, were more prevalent in vertebrates and urochordates, while single miRNA genes were more common in other metazoan species (Figure 1). Among these miRNAs, let-7a showed higher pre-miRNA multicopy numbers (1–7) relative to other homologous miRNAs. Although these multicopy pre-miRNAs could generate the same let-7a-5p

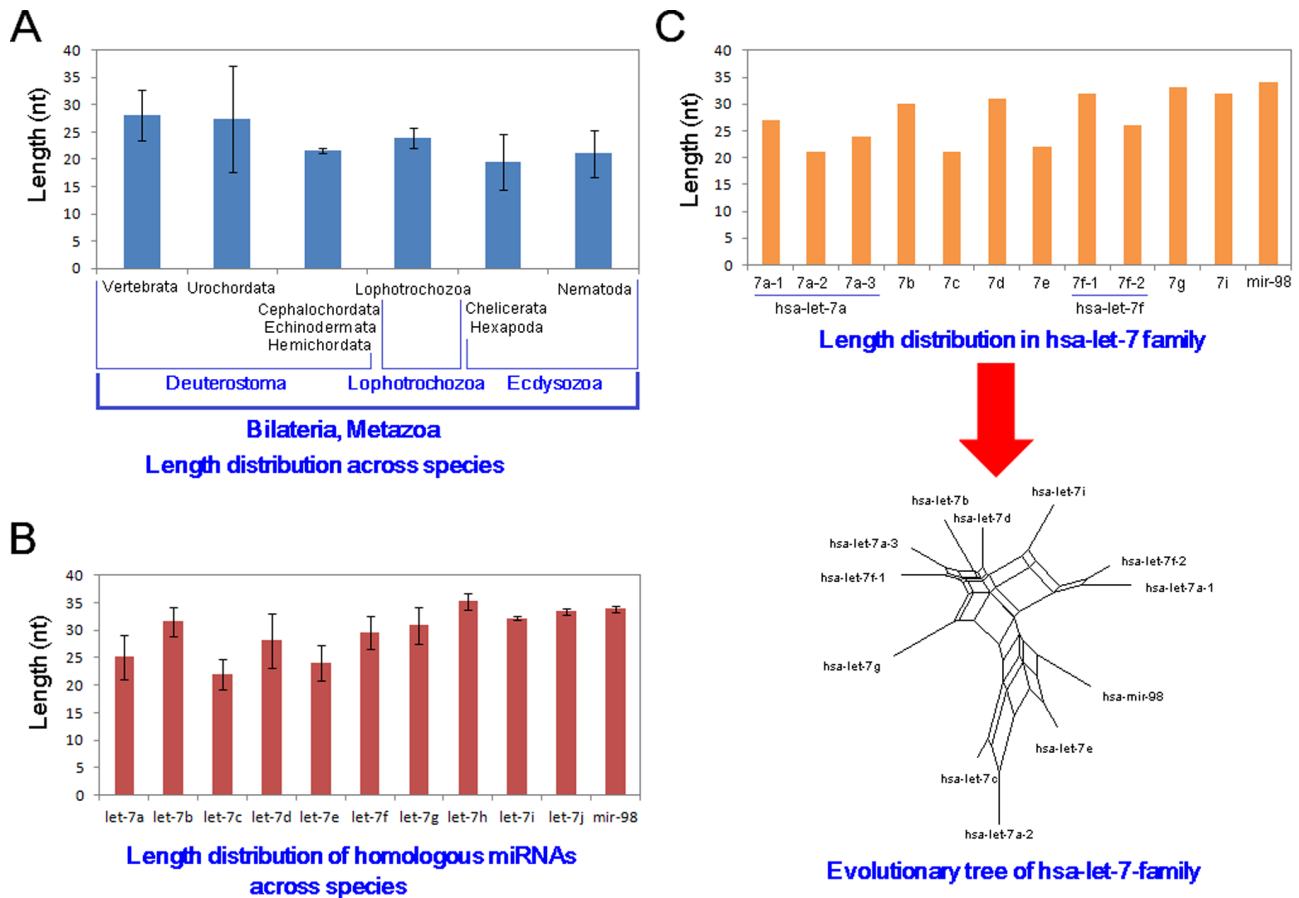


Figure 3. Length distributions of loops and their relevant evolutionary tree. (A) Length distribution of the let-7 loop among different animal species; (B) length distribution of loops in homologous miRNAs within the let-7 family across different animal species; and (C) length distribution of homologous hsa-let-7 loops and relevant evolutionary tree. doi:10.1371/journal.pone.0113042.g003

sequences, their loop sequences showed larger genetic distances (Figure S1 in File S1). Phylogenetic tree construction showed that the pre-miRNA multicopy loop sequences may be divided into two clusters based on loop and pre-miRNA sequences, with distribution differences noted (Figure S1 in File S1).

Although loop sequences were believed to not be well-conserved relative to the miR-#-5p and miR-#-3p sequences, loop sequences in *Drosophila* were found to be highly conserved (Figure 2A). Loop sequences among different species or from homologous miRNAs or multicopy pre-miRNAs showed higher levels of sequence divergence. Nucleotide insertions/deletions were common in the loop sequences, even between homologous miRNAs in the same species or between multicopy pre-miRNAs. Interestingly, varying length distributions between loop sequences from different animal species were noted (Figure 2B, $P_{adj} < 0.05$, except between Urochordata and Lophotrochozoa, Table S1 in File S1).

Based on the whole let-7 population, changes in length distribution across different species were found to be significant ($F = 19.14$, $P < 0.0001$, Figure 3A). The longest average loop length was found in vertebrates (28.13 ± 4.66 nts), followed by those in urochordates (27.36 ± 9.78 nts), which were longer loop sequences than those found in other metazoans (Figure 3A). Each member in the let-7 gene family showed an inconsistent length distribution (Figure 3), despite being homologous miRNAs with higher sequence similarity. Specifically, let-7a (25.12 ± 3.99 nts), let-7c (22.03 ± 2.81 nts) and let-7e (24.17 ± 3.21 nts) had shorter

length distributions than other homologous miRNAs (from 28.12 ± 4.87 nts to 35.25 ± 1.50 nts) ($F = 39.97$, $P < 0.0001$, Figure 3B). To further examine potential relationships between length distributions and evolutionary patterns, hsa-let-7 family loop sequences were analyzed. Loop sequences from multicopy pre-miRNAs showed larger divergence lengths, with these relevant loop sequences located in different evolutionary clusters (Figure 3C and Figure S1C in File S1). Larger genetic distances were noted in homologous miRNAs, with shorter loops from the let-7a-2, let-7c and let-7e genes located in a single cluster (Figure 3C and Figure S1C in File S1).

Nucleotide characteristics in loop sequence

According to the 10 nucleotides collected from the 5' and 3' ends of the loop sequences, guanine was the most prevalent (40.99%), followed by adenine (29.09%), while the cytosine presence was very low (5.32%) (Table S2 in File S1). Furthermore, if 3 continuous nucleotides were selected from either the 5' or 3' ends, no significant difference in nucleotide composition could be detected between the ends ($\chi^2 = 6.20$, $P = 0.102$). However, significant differences could be found between the two ends if only one terminal nucleotide was analyzed ($\chi^2 = 63.90$, $P < 0.0001$). Uracil and guanine were the dominant nucleotides at the 5' terminal ends (61.11% and 28.95%), while uracil was the most dominant nucleotide at the 3' terminal ends (81.58%) (Figure 4, Table S2 in File S1 and Figure S2A in File S1). Guanine

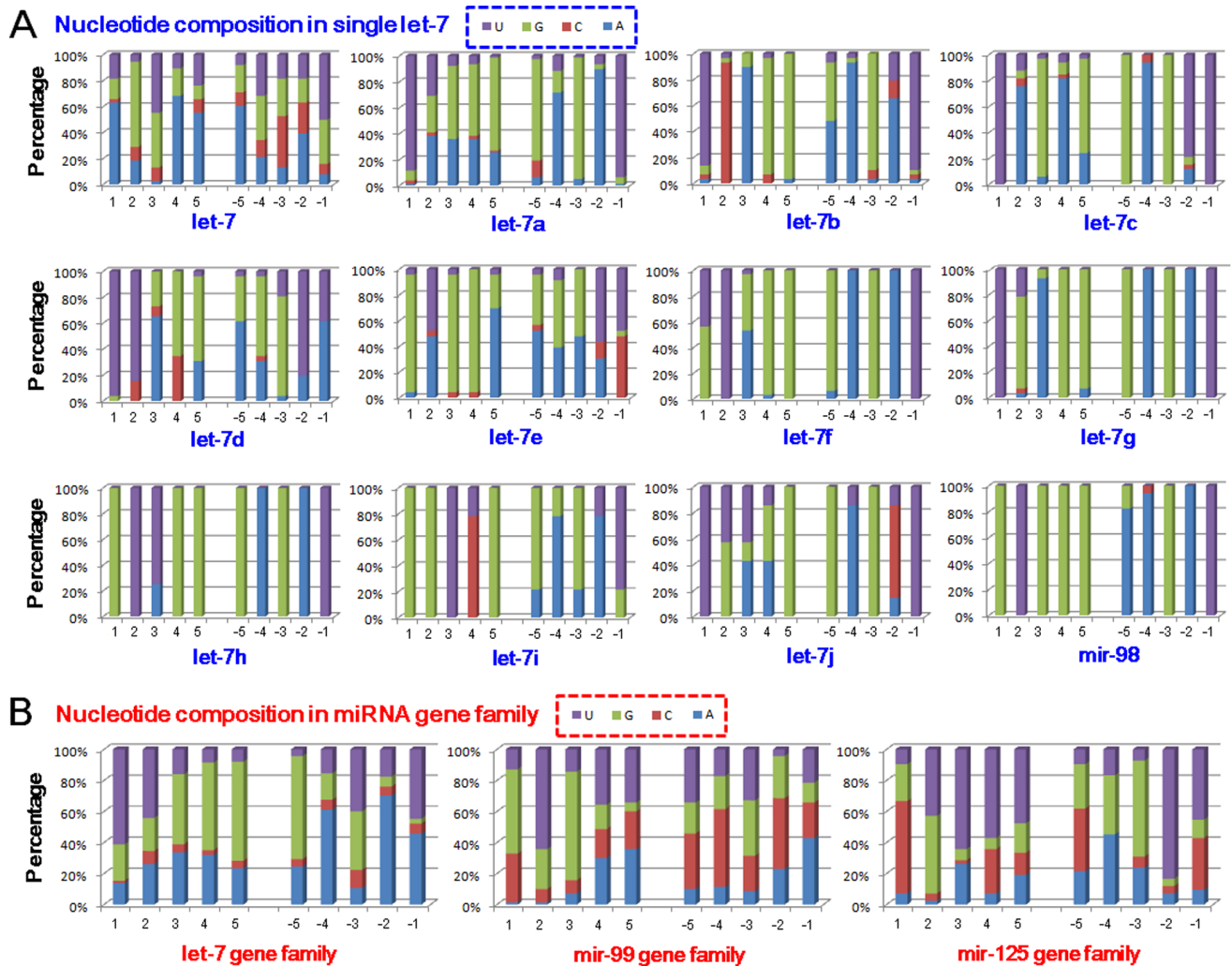


Figure 4. Nucleotide composition of the 5 continuous nucleotides at the 5' and 3' ends of the loop sequence. (A) Loop nucleotide composition in single let-7 genes (including let-7, let-7a, let-7b, etc.), with rare let-7 genes, such as let-7k and mir-3596, were not analyzed. (B) Loop nucleotide composition of relevant clustered miRNA gene families. doi:10.1371/journal.pone.0113042.g004

was present at an elevated rate, appearing at especially high frequencies in positions 4 and 5 at the 5' ends and positions -3 and -5 at the 3' ends (Table S2 in File S1), with a dominance also noted in the middle of the loop (Figure 5 and Figure S2A in File S1).

As a larger miRNA gene family, loop sequences showed larger nucleotide divergence than let-7-5p and let-7-3p (Figure 5A and Figure 5C). Let-7-5p acts as a canonical miRNA and is well conserved across animal species, particularly in the seed sequences. Among these homologous let-7 sequences, their loops showed varying levels of sequence similarity when examining single miRNA genes across animal species (Figure 5A and Figure S2A in File S1). While mature miRNAs, including miR-#-5p and miR-#-3p, are well conserved (especially for canonical mature miRNA) (Figure S2B in File S1), the loops commonly contain diverged nucleotides (Figure S2 in File S1) relative to the terminal regions that are relatively conserved.

Loop sequences in related miRNAs clustered with let-7

Based on the close physical distance between let-7 and other miRNAs, relevant miRNAs clustered with let-7 were collected

from different species. Similarly to the let-7 gene family, loops in related miRNA gene family were not as conserved (Figure 5B and Figure 5C). Interestingly, let-7 loops showed longer length distributions among their miRNA cluster (let-7 loops: 95% CI, 25.28–27.00 nts; clustered loops: 95% CI, 14.50–16.40 nts; $t = 16.14, P < 0.0001$), with a similar relation seen among the human let-7 genes and its clustered miRNAs (Table S3 in File S1). Based on the 5 continuous nucleotides extending from the 5' and 3' ends, significant difference could be detected between the three physically distant relevant miRNA gene families ($\chi^2 = 301.41, P < 0.0001$). Most miRNA gene families showed divergence in the dominant nucleotides between 5' and 3' ends (Table S2 in File S1). Curiously, we found that the let-7 cluster showed inconsistent dominant nucleotides at the 3' terminal across the let-7 gene family. Furthermore, diverse dominant nucleotides and nucleotide compositions could be detected between relevant alternative miRNA gene families (Table S2 in File S1).

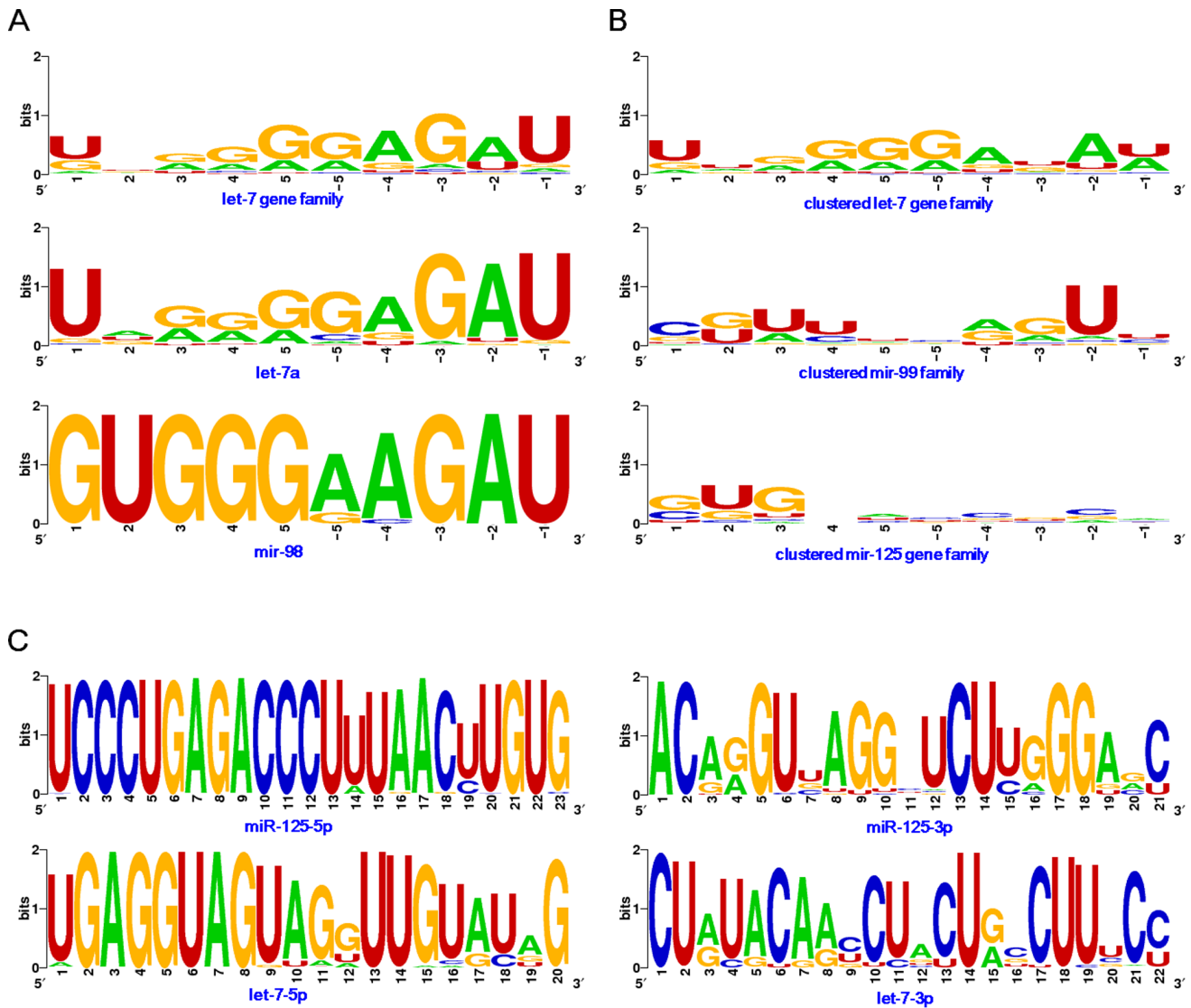


Figure 5. Sequence logo of loop sequences and miRNA sequences. (A) Sequence logo of all of the let-7 loop sequences, including the single let-7a and mir-98 loop sequences; (B) sequence logo of loop sequences from relevant clustered miRNA gene families; and (C) sequence logo of relevant mature miRNAs (including miR-#-5p and miR-#-3p). doi:10.1371/journal.pone.0113042.g005

Discussion

miRNAs are generated from pre-miRNA through Dicer recognition and cleavage of the terminal loop [38], with the loop connecting to miR-#-5p and miR-#-3p, which may contribute to the hairpin activity [39]. As a class of small non-coding RNAs, the negative regulatory RNAs have been widely studied, but little focus has been placed on the loop sequences. Indeed, the loops, including the loop structure and sequence, may tune and alter miRNA activity [9], which may further affect miRNA expression [13,14] by affecting Dicer recognition and cleavage during miRNA maturation [11,15]. The potential roles of loop sequences during miRNA biogenesis have been studied, especially for occurrence of multiple isomiRs. These isomiRs are mainly derived from imprecise and alternative cleavage by Drosha and Dicer during miRNA processing and maturation [19–24]. These loop sequences may provide more information for miRNA biogenesis, especially based on the analysis of loop sequences across different

animal species. In the present study, let-7 gene family loop sequences and other relevant miRNAs were analyzed and clustered to elucidate evolutionary and functional roles relating to miRNA biogenesis.

Multicopy miRNA genes may be located on different chromosomes and are mainly derived from historical duplication events [40–43]. Although the same mature miRNAs can be derived from these multicopy pre-miRNAs, the arms and loop sequences can be involved in larger divergence events, particularly the loops (Figure S1 in File S1) [7]. These loops may show diverse genetic distances relative to other homologous miRNA genes, and are always grouped in different clusters (Figure 3C and Figure S1 in File S1). Based on the phylogenetic relationships of pre-miRNAs and loops, similar distributions suggest that loop sequences dominate nucleotide divergence and evolutionary patterns in pre-miRNAs (Figure S1 in File S1). While both miR-#-5p and miR-#-3p sequences are conserved, loop sequences exhibit divergence through both varied nucleotides (including insertion/deletion)

and lengths (Figure 2, Figure 5 and Figure S2 in File S1). In *Drosophila*, the loop sequences and mature miRNAs are well conserved as well, with varied nucleotides and insertion/deletions found in the middle region of the loop (Figure 2A). The two terminal ends of the loops are relative conserved, which may contribute to stem-loop structure and cleavage by Dicer. These results suggest rapid evolution of the loops that further drives the evolution of miRNA genes.

Interestingly, we found changed length distributions of the loop sequences across different animal species and among different homologous miRNAs (Figure 2, Figure 3, Table S1 in File S1 and Table S3 in File S1). In higher metazoan species, the let-7 loops tend to be longer than those seen in lower species, with varying length distributions also seen different homologous miRNAs (Figure 3B and 3C). Clustered miRNAs tend to have similar length distribution, which implicates that the loop lengths may be affected by evolutionary relationships (Figure 3 and Figure S1 in File S1). Additionally, longer loop sequences may be an evolutionary trend in let-7 gene family, which may be of importance during miRNA biogenesis. Specifically, loop length may influence the stem-loop structure and stability (Table S4 in File S1), with longer loop sequences providing a possibility to dominate the evolution of miRNA genes across different animal species and homologous miRNAs within a specific species. These length variances further increase loop sequence diversity, which contributes to the evolutionary divergence between different miRNA genes, including homologous miRNAs (especially for multicopy pre-miRNAs). Indeed, although varied loop sequences exist, we still found the potential nucleotide characteristics in the 5' and 3' ends of the loops (Figure 4, Figure 5, Figure S2 in File S1 and Table S2 in File S1). Dominant nucleotides, such as uracil and guanine, are present at higher frequencies at the terminal ends, with these biased nucleotide compositions possibly influencing Dicer cleavage and contributing to the phenomenon of multiple isomiRs. While many isomiRs can be produced from a miRNA loci, only several isomiRs are dominantly expressed [21,24,33].

Some let-7 sequences are located in gene clusters with homologous miRNAs or other miRNAs, such as with the mir-125 and mir-99 gene families. Loops of clustered miRNA gene families may have different lengths, suggesting the length difference of loops from different miRNA genes (Table S3 in File S1). Significant differences in 5' and 3' terminal nucleotide compositions are noted among these clustered miRNAs, with both uracil and guanine dominating in the two terminus ends (Table S2 in File S1). Compared to the stable length distributions of miRNAs, rapid loop sequence evolution can drive miRNA gene

evolution and may further affect isomiR expression profiles during pre-miRNA processing [11,15]. Expression analysis indicates stable isomiR expression profiles [21,25], even across different tissues and animal species, suggesting stable Droscha and Dicer cleavage during miRNA processing and maturation. Furthermore, loop sequences also show higher levels of nucleotide divergence between homologous miRNA genes, especially between multicopy pre-miRNAs (Figure 2 and Figure S1 in File S1). These findings suggest that nucleotide divergence does not influence cleavage precision, but that nucleotide bias in the 5' and 3' ends of the loop potentially influences miRNA maturation.

Herein, canonical miRNA loop sequences were collected, with variations in the loops potentially based on the phenomenon of multiple isomiRs, with the canonical miRNA sequences not necessarily the most dominant sequence. However, according to length distributions and terminal end nucleotide compositions, it can be concluded that loop sequences tend to be longer in higher animal species, with rapid evolution of the loop sequences further driving miRNA gene evolution. Stable isomiR expression profiles indicate relative cleavage, which may be closely related to the dominant nucleotide distributions. This study further enriches the understanding of miRNA biogenesis as it relates to loop sequences across different animal species and among homologous miRNAs, particularly considering the phenomenon of multiple isomiRs.

Supporting Information

File S1 Figure S1. Examples of neighbor-joining tree of the loop and pre-miRNA sequences in the let-7 gene family. Figure S2. Sequence logo of loop sequences and miRNA sequences. (A) Sequence logo of loop sequences from single let-7 genes and (B) sequence logo of relevant mature miRNAs (including miR-#-5p and miR-#-3p). **Table S1. Loop sequence length distribution from Figure 2B. Table S2. Nucleotide composition and frequency. Table S3. 95% confidence interval (CI) and difference between length distributions of let-7 and related clustered miRNAs. Table S4. Effect on minimum free energy (MFE) when changing loop sequence lengths in hsa-let-7a-1.** (DOC)

Author Contributions

Conceived and designed the experiments: TML CL LG. Performed the experiments: TML CL LG. Analyzed the data: TML CY PL. Contributed reagents/materials/analysis tools: CY PL CL. Contributed to the writing of the manuscript: TML LG.

References

- Lai EC (2003) microRNAs: runts of the genome assert themselves. *Curr Biol* 13: R925–936.
- Flynt AS, Lai EC (2008) Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nature Reviews Genetics* 9: 831–842.
- Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. *Embo Journal* 21: 4663–4670.
- Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, et al. (2008) The regulatory activity of microRNA star species has substantial influence on microRNA and 3' UTR evolution. *Nature Structural & Molecular Biology* 15: 354–363.
- Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, et al. (2009) Hierarchical rules for Argonaute loading in *Drosophila*. *Mol Cell* 36: 445–456.
- Okamura K, Liu N, Lai EC (2009) Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol Cell* 36: 431–444.
- Guo L, Lu Z (2010) The Fate of miRNA* Strand through Evolutionary Analysis: Implication for Degradation As Merely Carrier Strand or Potential Regulatory Molecule? *Plos One* 5: e11387.
- Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang HB, Arrese EL, et al. (2010) Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC Genomics* 11: 52.
- Yue SB, Trujillo RD, Tang Y, O'Gorman WE, Chen CZ (2011) Loop nucleotides control primary and mature miRNA function in target recognition and repression. *RNA Biol* 8: 1115–1123.
- Brummelkamp TR, Bernards R, Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296: 550–553.
- Hinton TM, Wise TG, Cottee PA, Doran TJ (2008) Native microRNA loop sequences can improve short hairpin RNA processing for virus gene silencing in animal cells. *J RNAi Gene Silencing* 4: 295–301.
- Vlassov AV, Korba B, Farrar K, Mukerjee S, Seyhan AA, et al. (2007) shRNAs targeting hepatitis C: effects of sequence and structural features, and comparison with siRNA. *Oligonucleotides* 17: 223–236.
- Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* 32: 4776–4785.
- Boudreau RL, Martins I, Davidson BL (2009) Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. *Mol Ther* 17: 169–175.

15. McManus MT, Petersen CP, Haines BB, Chen J, Sharp PA (2002) Gene silencing using micro-RNA designed hairpins. *Rna* 8: 842–850.
16. Vermeulen A, Behlen L, Reynolds A, Wolfson A, Marshall WS, et al. (2005) The contributions of dsRNA structure to Dicer specificity and efficiency. *Rna* 11: 674–682.
17. Liu G, Min H, Yue S, Chen CZ (2008) Pre-miRNA loop nucleotides control the distinct activities of mir-181a-1 and mir-181c in early T cell development. *PLoS One* 3: e3592.
18. Michlewski G, Guil S, Semple CA, Caceres JF (2008) Posttranscriptional Regulation of miRNAs Harboring Conserved Terminal Loops. *Molecular Cell* 32: 383–393.
19. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.
20. Morin RD, Aksay G, Dolgoshina E, Ebhardt HA, Magrini V, et al. (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Research* 18: 571–584.
21. Guo L, Yang Q, Lu J, Li H, Ge Q, et al. (2011) A Comprehensive Survey of miRNA Repertoire and 3' Addition Events in the Placentas of Patients with Pre-eclampsia from High-throughput Sequencing. *PLoS ONE* 6: e21072.
22. Neilsen CT, Goodall GJ, Bracken CP (2012) IsomiRs - the overlooked repertoire in the dynamic microRNAome. *Trends Genet* 28: 544–549.
23. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, et al. (2010) Complexity of the microRNA repertoire revealed by next generation sequencing. *Rna-a Publication of the Rna Society* 16: 2170–2180.
24. Guo L, Chen F (2014) A Challenge for miRNA: Multiple IsomiRs in miRNAomics. *Gene* 544: 1–7.
25. Burroughs AM, Ando Y, de Hoon MJL, Tomaru Y, Nishibu T, et al. (2010) A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Research* 20: 1398–1410.
26. Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. *Journal of Molecular Biology* 339: 327–335.
27. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T (2003) New microRNAs from mouse and human. *Rna-a Publication of the Rna Society* 9: 175–179.
28. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP (2003) Vertebrate MicroRNA genes. *Science* 299: 1540–1540.
29. Kim VN, Nam JW (2006) Genomics of microRNA. *Trends in Genetics* 22: 165–173.
30. Xu JZ, Wong CW (2008) A computational screen for mouse signaling pathways targeted by microRNA clusters. *Rna-a Publication of the Rna Society* 14: 1276–1283.
31. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
32. Guo L, Li H, Lu J, Yang Q, Ge Q, et al. (2012) Tracking miRNA precursor metabolic products and processing sites through completely analyzing high-throughput sequencing data. *Mol Biol Rep* 39: 2031–2038.
33. Guo L, Chen F, Lu Z (2013) Multiple IsomiRs and Diversity of miRNA Sequences Unveil Evolutionary Roles and Functional Relationships Across Animals. *MicroRNA and Non-Coding RNA: Technology, Developments and Applications*. 127–144.
34. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, et al. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol* 6: 26.
35. Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729.
36. Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68–73.
37. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–1190.
38. Tsutsumi A, Kawamata T, Izumi N, Seitz H, Tomari Y (2011) Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nat Struct Mol Biol* 18: 1153–1158.
39. Schopman NC, Liu YP, Konstantinova P, ter Brake O, Berkhout B (2010) Optimization of shRNA inhibitors by variation of the terminal loop sequence. *Antiviral Res* 86: 204–211.
40. Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ (2008) MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America* 105: 2946–2950.
41. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, et al. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7: 25.
42. Zhang R, Peng Y, Wang W, Su B (2007) Rapid evolution of an X-linked microRNA cluster in primates. *Genome Research* 17: 612–617.
43. Guo L, Sun BL, Sang F, Wang W, Lu ZH (2009) Haplotype Distribution and Evolutionary Pattern of miR-17 and miR-124 Families Based on Population Analysis. *PLoS ONE* 4: e7944.