Molecular evolution of microRNAs in bilaterian animals revealed by large-scale genomic analysis

Kahori Ikeda

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Graduate School of Media and Governance

KEIO UNIVERSITY
2014

Abstract

MicroRNAs (miRNAs) are small noncoding RNAs that regulate the expression of messenger RNAs (mRNAs). Although miRNAs may have strongly influenced the evolution of animals, the details of their evolution are yet to be resolved. Therefore, I have sought to clarify miRNA evolution by analyzing the conservation of miRNAs in both model and non-model species.

First, I predicted the fundamental regulatory relationships between miRNAs and their target genes that have been conserved throughout the evolution of the bilaterian animals. For this purpose, I designed a bioinformatics procedure to extract conserved miRNA/target-gene pairs, which predicted 31 evolutionarily conserved miRNA/target-gene pairs. The downregulation of six of these pairs was observed in HeLa cells, using a reporter-gene assay. I inferred that these pairs were present in the primitive gene regulatory network of the common bilaterian ancestor.

I also examined the "living fossil" *Triops cancriformis*, the tadpole shrimp. Because this non-model species has an interesting evolutionary history and morphology, I hypothesized that as-yet-undiscovered miRNA regulatory mechanisms and evolutionary trends would be revealed by comparing the miRNAs of this organism with those of model species. Deep-sequencing identified 180 miRNAs and six components of the RNAi machinery. The expression patterns of four of the conserved *T. cancriformis* miRNAs differed from those of *Drosophila melanogaster*. Most of the conserved *T. cancriformis* miRNAs share sequence similarities with those of the arthropods. However, the let-7 sequence and domains of DICER are more similar to those of the vertebrates, suggesting that the miRNA system of *T. cancriformis* evolved in a unique way.

In conclusion, even when the miRNA target genes are conserved among species, they may function differently depending on their expression patterns. I discuss the evolution of miRNAs in bilaterian animals with reference to the miRNA biology in model and non-model species.

Keywords: microRNA, evolution, development, genome informatics, comparative analysis

論文題目

「ゲノム情報を用いた左右相称動物における microRNA の分子進化学的解析」

論文要旨

MicroRNA (miRNA) は 22 塩基程度の小分子 RNA であり、標的となる messenger RNA (mRNA) の遺伝子発現を調節する役割を有している。近年、miRNA は動物の系統進化に多大な影響を及ぼしてきたと考えられているが、miRNA の進化については未解明な部分が多い。そこで私は、モデル生物及び非モデル生物を対象に miRNA の保存性を解析することによって、その進化に迫った。

まず、生物にとって重要な miRNA とその標的遺伝子は左右相称動物の進化を通して保存されてきているのではないかと仮説を立てた。そこで、5 種の左右相称動物のモデル生物種間で保存された 5 つの miRNA に着目し、それらの標的遺伝子を予測した。その結果、31 種の保存された miRNA/標的遺伝子ペアを情報学的手法によって抽出することができ、このうち 6 ペアに関しては miRNA による遺伝子発現制御を実験的に検証することができた。本解析から、これらの進化的に保存された miRNA と標的遺伝子は左右相称動物の祖先生物が既に有していたことが推測された。

次に生きた化石として知られているヨーロッパカブトエビ *Triops cancriformis* に着目した。この非モデル生物は進化と発生に特徴を有しているため、カブトエビの miRNA システムをモデル生物と比較することで、今まで知り得なかった miRNA の制御や進化が明らかになるのではないかと考えた。まず、カブトエビにおいて 180 種の miRNA 及び 6 種の RNAi 関連因子を同定した。次にカブトエビの発生段階における 6 種の保存された miRNA の発現パターンを、ショウジョウバエのそれと比較したところ、異なる挙動を示す miRNA を見つけることができた。また、保存性解析によって、カブトエビの保存された miRNA の大半が節足動物の miRNA と相同性を有している一方、カブトエビの let-7 配列や DICER のドメインは、節足動物より脊椎動物に近いタイプである可能性が示唆された。

これらの結果を総合して考えると、たとえ生物種間で保存されている miRNA と標的遺伝子であっても、miRNA の発現時期が異なることで、異なる役割を有する可能性があることが推察された。本学位論文ではこれらの研究から得られた知見を基に、左右相称動物の miRNA 進化について議論する。

キーワード：microRNA, 進化, 発生, ゲノム情報解析, 比較ゲノム解析

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A. mellifera	*Apis mellifera*
A. queenslandica	*Amphimedon queenslandica*
A. suum	*Ascaris suum*
AGO	Argonaute
AUB	Aubergine
B. belcheri	*Branchiostoma belcheri*
B. mori	*Bombyx mori*
BMP	bone morphogenetic protein
C. elegans	*Caenorhabditis elegans*
D. melanogaster	*Drosophila melanogaster*
D. rerio	*Danio rerio*
DGCR8	DiGeorge syndrome critical region protein 8
DNA	deoxyribonucleic acid
EC value	multiplied value of Enrichment and Coverage
G. gallus	*Gallus gallus*
H. sapiens	*Homo sapiens*
HEN1	Hua enhancer 1
HYL1	hyponastic leaves 1
ILTV	infectious laryngotracheitis virus
ITS	internal transcribed spacer
M. japonicus	*Marsupenaeus japonicus*
M. musculus	*Mus musculus*
miRISC	miRNA-induced silencing complex
miRNA	microRNA
mRNA	messenger RNA
N. vectensis	*Nematostella vectensis*
ncDNA	noncoding DNA
ncRNA	noncoding RNA
nt	nucleotide
PCR	polymerase chain reaction

piRNA	PIWI-interacting RNA
pre-miRNA	precursor microRNA
pri-miRNA	primary microRNA
pSILAC	pulsed stable isotope labelling with amino acids in cell culture
PUMA	proapoptotic p53 upregulated modulator of apoptosis
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
stRNA	small temporal RNA
T. cancriformis	*Triops cancriformis*
TRBP	TAR RNA binding protein
tRNA	transfer RNA
UTR	untranslated region
X. tropicalis	*Xenopus tropicalis*

Chapter 1

Introduction

1.1 Noncoding RNAs

How do animals differ? Animals comprise numerous and diverse types of cells, totaling 3.72×10^{13} in *Homo sapiens* (Bianconi et al. 2013). In all animals, biological information is stored in the deoxyribonucleic acid (DNA) sequences of their genomes. The double helical structure of DNA was discovered by Watson and Crick in 1953 (Watson and Crick 1953). The genetic information is transcribed from DNA as ribonucleic acid (RNA) species called "messenger RNAs" (mRNAs), which are translated in the synthesis of proteins. This flow of genetic information was called the "central dogma" by Francis Crick in 1958 (Crick 1958).

All the animals contain DNA, RNA, and proteins. This raises questions about the correlations between genome size, numbers of coding genes, and biological complexity. The human genome comprises approximately 3 billion base pairs and the number of genes encoding human proteins is estimated to be approximately 20,000 (Clamp et al. 2007; Ezkurdia et al. 2014). The size of the *Drosophila melanogaster* genome is approximately 120 million base pairs, and the number of protein-coding genes is approximately 13,600 (Adams et al. 2000). The alternative splicing of mRNA precursors increases the number of protein isoforms expressed. For this reason, the number of proteins expressed by most species is not precisely known. Although the number of protein-coding genes and the genome size seem to correlate with biological complexity, *D. melanogaster* has fewer protein-coding genes than *Caenorhabditis elegans* (19,735 genes) (Hillier et al. 2005). Moreover, the genome size and the

number of protein-coding genes in the plant *Triticum aestivum* (17 billion base pairs and 124,201, respectively) (International Wheat Genome Sequencing 2014) are greater than those of *H. sapiens*. The definition of biological complexity is controversial, and factors other than genome size and the number of protein-coding genes may contribute to it. Transcriptome analyses have revealed that approximately 70% of the genome is transcribed and more than 50% of the transcribed sequences represent noncoding RNAs (ncRNAs) in mouse (Carninci et al. 2005). Because evidence indicates that some ncRNAs are functional (Mattick and Makunin 2006), they are increasingly recognized as important molecules.

The number of noncoding DNAs (ncDNAs) is consistent with the biological complexity of an organism (Taft et al. 2007), suggesting a close relationship between the number of ncRNAs and biological complexity. Canonical ncRNAs include transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), which are required for protein synthesis in species ranging from bacteria to humans. Small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) are also classified as ncRNAs, and form complexes with proteins, that mediate RNA splicing and RNA modification, respectively (Dieci et al. 2009; Karijolich and Yu 2010). It has recently been demonstrated that small ncRNAs form a novel level in the genetic regulation of cellular functions (Rother and Meister 2011). In animals, PIWI-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and microRNAs (miRNAs) mediate gene silencing. piRNAs, which consist of approximately 25–30 nucleotides (nt), associate with members of the PIWI family of proteins (PIWI, AUB, and AGO3), which are

involved in the development of the gonads (Siomi et al. 2011). An siRNA molecule comprises approximately 21 nt, is produced from double-stranded RNA, and regulates the expression of target genes (Carthew and Sontheimer 2009). In this dissertation, I focus on miRNAs because miRNAs seem to be related to the complexity of the organisms.

1.2 Discovery of miRNAs

miRNAs are ncRNAs, of approximately 22 nt, that regulate the expression of their target mRNAs, mainly at the posttranscriptional level (Bartel 2004). miRNAs mediate processes such as cell proliferation and differentiation, development, apoptosis, and diseases (Adlakha and Saini 2014). So-called "seed sequences" of approximately 7 nt in the 5′ regions of miRNAs are important for the recognition of their miRNA targets (Lewis et al. 2003; Bartel 2009; Nahvi et al. 2009). Currently, 24,521 mature miRNAs from 206 species (animals, plants, and viruses) are registered in miRBase release 20.0 (Kozomara and Griffiths-Jones 2013). Because humans and flies have 2,588 and 466 mature miRNAs, respectively, the number of miRNAs may correlate with biological complexity.

The first miRNA, lin-4, was discovered in *C. elegans* in 1993, during study of a mutation that causes failure to develop (Lee et al. 1993; Wightman et al. 1993). The second miRNA to be discovered, let-7 of *C. elegans*, was reported in 2000 (Reinhart et al. 2000), and its sequence is conserved among vertebrates, ascidian, hemichordate, mollusc, annelid and arthropod. Its expression pattern is also conserved in some of these organisms (Pasquinelli et al. 2000). Because these miRNAs are expressed during specific developmental stages, they were originally called "small temporal RNAs" (stRNAs) (Pasquinelli et al. 2000). The term "RNA interference" (RNAi) was coined by Fire et al. in 1998 (Fire et al. 1998) to denote the regulation of gene expression by the degradation of specific mRNAs by double-stranded RNAs. It

has been shown that the molecular mechanism underlying RNAi involves the cleavage of mRNAs by 21–23 nt fragments produced from double-stranded RNAs (Zamore et al. 2000). In 2001, DICER was shown to generate small RNAs (Bernstein et al. 2001), and DICER1 may participate in the biogenesis of miRNAs (Grishok et al. 2001; Hutvagner et al. 2001). Simultaneously, miRNAs were established as a novel class of regulators by their identification in the genomes of *C. elegans*, *D. melanogaster*, and *H. sapiens* (Lagos-Quintana et al. 2001; Lau et al. 2001). In general, miRNAs control the levels of mRNA expression by binding to the 3′-untranslated region (UTR) of mRNAs, and some miRNA targets were identified by exploiting this mechanism (Krek et al. 2005). However, the evolutionary background of miRNAs and their target genes is unknown.

It is noteworthy that miRNAs are also encoded by viral genomes (Pfeffer et al. 2004). Viral miRNAs target their own mRNAs and those of their hosts during latent and lytic infection, immune evasion, apoptosis, viral replication, and the cell cycle (Takane and Kanai 2011). For example, infectious laryngotracheitis virus (ILTV) miR-I5 targets the mRNA of transcriptional activator ICP4, which is essential for viral growth. Its expression is repressed during latent infection, indicating that miR-I5 modulates the balance between lytic and latent viral infection (Waidner et al. 2011). miR-BART5 represses the expression of the mRNA encoding proapoptotic p53 upregulated modulator of apoptosis (PUMA), protecting the virus from apoptosis (Choy et al. 2008).

The development of deep-sequencing technologies facilitated the

identification of several miRNAs, allowing a more precise record of the variety and number of miRNAs expressed by diverse species to be established. Although some miRNAs are common among bilaterian animals, the evolution of miRNA sequences, miRNA targets, and miRNA components has not yet been clearly delineated.

1.3 Biogenesis of miRNAs

How are mature miRNAs synthesized? Is the miRNA processing pathway conserved among animals? A general scheme depicting the miRNA biosynthetic pathway is presented in Figure 1.1. Several steps are required to produce mature miRNAs. Initially, the primary miRNAs (pri-miRNAs) are transcribed from the genome in the nucleus (Lee et al. 2004), and are converted into hairpin RNAs of approximately 70 nt by Drosha, a member of the RNase III family, and DiGeorge-syndrome critical region protein 8 (DGCR8; also called Pasha) (Lee et al. 2003; Denli et al. 2004). Exportin-5 exports the processed RNAs, called "precursor miRNAs" (pre-miRNAs), to the cytoplasm, where the miRNA duplexes are processed by the RNase III DICER (Bernstein et al. 2001; Lund et al. 2004). The mature miRNAs ("guide strands") are then incorporated into the miRNA-induced silencing complex (miRISC) to play a role in regulating gene expression. The other strands ("passenger strands") are immediately degraded, although both the guide and passenger strands may be functional (Stark et al. 2007). In animals, the types of AGO proteins present in the miRISC differ from those in other organisms. In the arthropod *D. melanogaster*, miRNAs are processed by DICER1 and bind to AGO1, whereas in mammals, miRNAs are also processed by DICER1 but bind to AGO1–4 (Okamura et al. 2004; Peters and Meister 2007).

Plant cells synthesize mature miRNAs through a similar process. However, DCL1 and the RNA-binding protein hyponastic leaves 1 (HYL1) converts pri-miRNAs to pre-miRNAs, and the Drosha does not occur in plants (Kurihara et al. 2006). Other

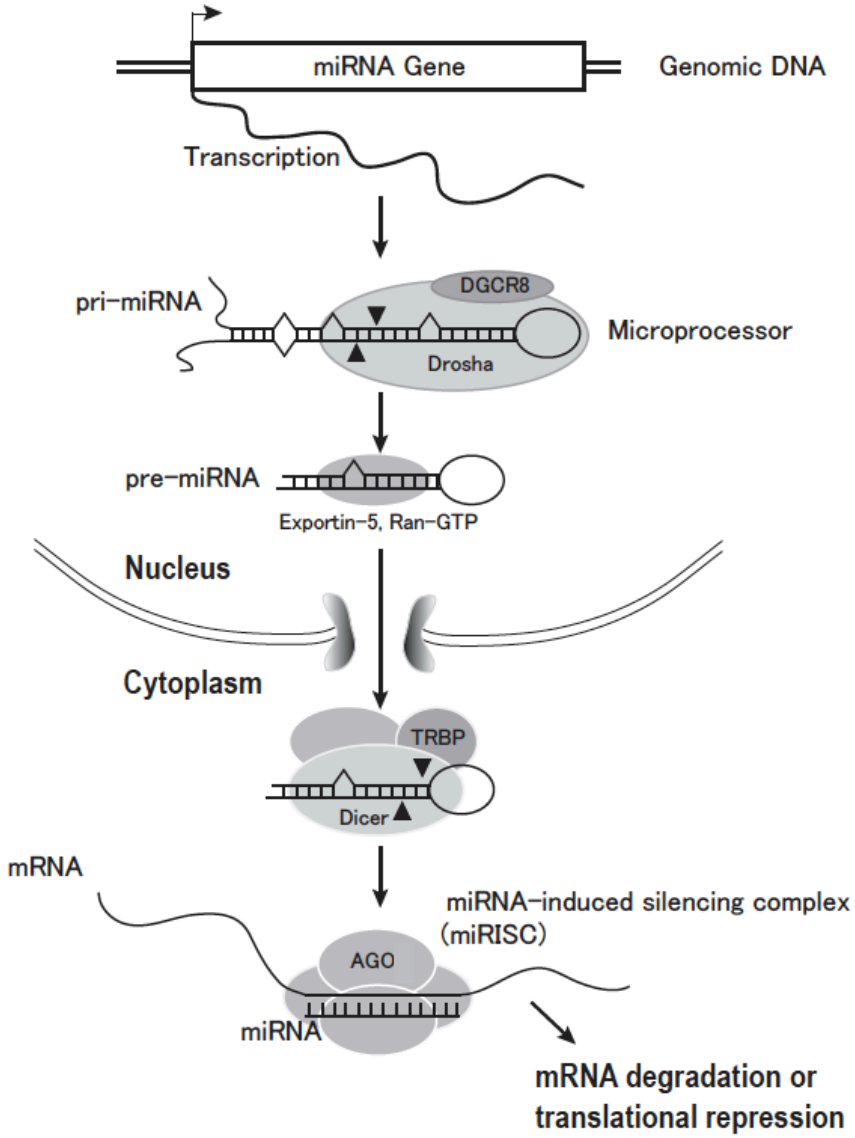


Figure 1.1 Canonical miRNA processing pathway

Initially, miRNA genes are transcribed to produce pri-miRNAs in the nucleus. The RNase III Drosha and its binding partner DGCR8, also known as microprocessor (also called Pasha), converts pri-miRNAs into pre-miRNAs. After export to the cytoplasm by a complex including Exportin-5, pre-miRNAs are processed into miRNA duplexes by the RNase III DICER with its cofactor TAR RNA binding protein (TRBP). One strand of the miRNA duplex is incorporated into the miRISC containing AGO that mediates the mRNA degradation or the translational repression. The triangles in the pri-miRNA and pre-miRNA molecules indicate the cleavage sites recognized by RNase III isoforms.

steps are required for miRNA biosynthesis in plants. For example, miRNA duplexes are methylated by the methyltransferase Hua enhancer 1 (HEN1) (Yu et al. 2005). In the plant miRISC, as mentioned above, DCL1 is required for miRNA production and binds mainly to AGO1, one of 10 *Arabidopsis thaliana* AGO paralogues (Voinnet 2009). Viral genomes do not encode miRNA regulatory proteins, and viruses therefore use the proteins of their hosts.

Although the basic miRNA biosynthesis is described here, some exceptions have been reported. Mirtrons are an example of miRNAs derived *via* another processing pathway. They are derived from the intron of an mRNA and are expressed by *D. melanogaster* and *C. elegans* (Ruby et al. 2007). The biosynthesis of miRNAs from mirtrons is independent of Drosha activity. Hundreds of mirtrons are present in mammals (Ladewig et al. 2012). Another alternative pathway for the synthesis of miRNAs is independent of DICER activity. Drosha cleavage occurs normally, but the pre-miRNA is not loaded onto DICER but directly onto AGO (Cheloufi et al. 2010). The components of these alternate pathways differ slightly, depending on the species, although they have not yet been identified in detail in non-model species. Based on these findings, I wondered whether our knowledge of miRNAs gained from model species is applicable to all bilaterian animals.

1.4 Aim: to understand the evolution of miRNAs by analyzing model and non-model species

As a first step towards answering the question raised in the previous section concerning the conservation of miRNAs among species, I analyzed the evolution of miRNAs in the bilaterian animals. The sequences of certain miRNAs are highly conserved among the bilaterian animals (Prochnik et al. 2007). However, Grimson et al. reported that *Nematostella vectensis*, which diverged before the appearance of the bilaterian animals, has no miRNAs with sequences similar to those of the bilaterian animals, except miR-100 (Grimson et al. 2008). The number of *N. vectensis* miRNAs is small (49 precursor miRNAs) compared with the numbers in more complex animals, and organismal complexity correlates with the number of miRNAs (Grimson et al. 2008). Furthermore, the interaction between the miRNA/target-gene pair is related to the phenotypic changes that have occurred in related species. For instance, Texel sheep have well-developed muscles compared with normal sheep. A nucleotide substitution in the GDF8 mRNA, which is a negative regulator of strong muscle growth, causes it to be targeted by miR-1 and miR-206, leading to the more muscular phenotype (Clop et al. 2006). These findings suggest that miRNAs are strongly related to the evolution of animals. Therefore, I surmised that the analysis of miRNA evolution would extend our understanding of the evolution of the bilaterian animals.

To better understand miRNA evolution, I focused my analysis on both model and non-model species. First, I wondered how many miRNA/target genes have

coevolved and how conserved miRNA/target genes function. Because conserved miRNA/target-gene pairs exist in a wide range of species, they are considered to have played important roles in evolution. To understand the functions of miRNAs, it is very important to identify the target mRNAs of individual miRNAs. Therefore, I predicted evolutionarily conserved miRNA/target-gene pairs in the bilaterian animals and identified them using bioinformatics and molecular biological techniques. From this analysis, it was possible to infer the gene regulation afforded by miRNAs in ancient animals (Chapter 2).

miRNAs have been identified and analyzed in various model species, including humans and flies. However, I wanted to know whether our knowledge of the miRNAs in model species can be extended to all bilaterian animals. As described above, miRNAs are strongly related to organismal morphology. By analyzing the miRNAs of non-model species, which have characteristic morphologies and evolutionary histories, I thought it would be possible to identify as-yet-undiscovered regulatory and evolutionary characteristics of miRNAs. Therefore, I focused on the tadpole shrimp *Triops cancriformis*, because both its evolutionary history and its morphology are interesting. To investigate the molecular evolution of miRNAs, I conducted a comparative analysis of the miRNAs of this organism and the components of its RNAi machinery (Chapter 3).

I emphasize again that understanding miRNA evolution is important for understanding animal evolution because miRNAs are intimately related to the evolution of the bilaterian animals (Niwa and Slack 2007; Prochnik et al. 2007;

Grimson et al. 2008). Therefore, my purpose was to understand the molecular evolution of miRNAs by analyzing their conservation, their targets, and the components of RNAi machinery among the bilaterian animals.

Chapter 2

Computational prediction and experimental validation of evolutionarily conserved microRNA target genes in bilaterian animals

2.1 Introduction

Currently, numerous microRNAs (miRNAs) with diverse sequences are being characterized in a wide range of species (Pasquinelli et al. 2000), suggesting that this small RNA molecule has a major effect on phylogeny. The importance of miRNAs is also suggested from recent research demonstrating that miRNA-guided gene regulation is involved in diverse biological functions, such as cell differentiation, development, carcinogenesis, and tumour suppression (He et al. 2005; Wienholds and Plasterk 2005; Shivdasani 2006; Johnson et al. 2007). For example, phylogenetically conserved miRNAs (e.g., let-7, miR-1, miR-124, and miR-125) are involved in cell differentiation and development (Reinhart et al. 2000; Chen et al. 2006; Caygill and Johnston 2008; Yu et al. 2008). In this case, let-7 regulates the expression of RAS proteins known as critical oncogene products (Johnson et al. 2005). Moreover, miR-34, another evolutionarily conserved miRNA, is a direct downregulator of p53 and is involved in a genetic pathway that promotes cell-cycle progression (He et al. 2007).

In recent years, more than thousands of miRNAs have been identified in humans (Griffiths-Jones et al. 2008), and this number is increasing. In a recent report by Friedman et al., the expression of a large number of target genes is predicted to be regulated by miRNAs (Friedman et al. 2009); however, relatively few of these have been verified experimentally. To overcome this problem, a series of computational methods has been developed to predict a large number of miRNA targets; e.g., TargetScan (Friedman et al. 2009), RNAhybrid (Rehmsmeier et al. 2004), MicroTar

(Thadani and Tammi 2006), PITA (Kertesz et al. 2007), miRanda (John et al. 2004), and PicTar (Krek et al. 2005). Nevertheless, these computational approaches often provide numerous target candidates with a large number of false positives because of the weak complementarity between miRNAs and 3′-untranslated regions (UTRs) (Watanabe et al. 2007). Recently, a phylogenetic profiling approach has been applied to overcome this limitation. For example, studies of the evolution of orthologous target sites have provided insights into the prediction of efficient miRNA targets (Gaidatzis et al. 2007). As for miRNAs, many miRNA families are found among various bilaterian animals, suggesting that several miRNAs and their target genes may have co-evolved; however, these features have yet to be systematically characterized.

I hypothesize here that the core regulatory relationship between miRNAs and their target genes were conserved throughout the evolution of bilaterian animals. In addition, by predicting these relationships, I sought to elucidate the core function of miRNAs in the primitive gene-regulatory network of the common bilaterian ancestor. Accordingly, I focused on five miRNAs (let-7, miR-1, miR-124, miR-125/lin-4, and miR-34) that are conserved among bilaterian species (*Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Drosophila melanogaster,* and *Caenorhabditis elegans*) and designed a procedure to extract conserved miRNA/target-gene pairs. I extracted evolutionarily conserved miRNA/target-gene pairs based on hybridization patterns and orthologous information. Further, I experimentally verified several candidate pairs to support my methodology. These results suggest a functional role of three major miRNAs (let-7, miR-1, and miR-124) that regulated genes related to development, muscle formation,

and cell adhesion. These results suggest a new role for the core function of miRNAs in the primitive gene-regulatory network of the common bilaterian ancestor.

2.2 Materials and methods

2.2.1 miRNA and 3'-UTR sequence data

I downloaded 2,404 mature miRNA sequences (885 for *H. sapiens*, 689 for *M. musculus*, 520 for *G. gallus*, 153 for *D. melanogaster*, and 157 for *C. elegans*) from the miRBase, version 13.0 (http://microrna.sanger.ac.uk/sequences/) (Griffiths-Jones et al. 2008). I downloaded sequences corresponding to 3′-UTR using the Ensembl transcript ID annotation in FASTA format (40,498 transcripts for *H. sapiens*, 3,332 for *M. musculus*, 13,089 for *G. gallus*, 16,822 for *D. melanogaster*, and 13,560 for *C. elegans*) from the Ensmart database Ensembl release 53 (http://www.ensembl.org/index.html) (Kasprzyk et al. 2004). Orthologous gene information was also downloaded from the Ensmart database Ensembl release 53. I obtained 145 experimentally verified miRNA/target-gene pairs from TarBase Version 5.0.1 (http://diana.cslab.ece.ntua.gr/tarbase/) (Papadopoulos et al. 2009).

2.2.2 Identification of miRNAs conserved among bilaterian animals

The 2,404 miRNA sequences were aligned using ClustalX (Larkin et al. 2007) with the following alignment parameters: gap opening: 22.50; gap extension: 0.83; and bootstrap value: 100. I checked the conservation of 2,404 miRNA sequences to extract evolutionarily conserved miRNAs. I defined the conservation threshold as an "overall sequence identity >75% with complete matching of the seed sequence (1–7, 2–8, or 3–9 nucleotides from the miRNA's 5′ end)"; furthermore, I introduced information on the

phylogenetic relationship among miRNAs to extract reliably conserved miRNAs and used highly conserved miRNA families (category I) (Huang and Gu 2007).

2.2.3 Extraction of evolutionarily conserved miRNA/target-gene pairs among bilaterian animals

To extract evolutionarily conserved miRNA/target-gene pairs among bilaterian animals, I devised a three-step filtering approach (Figure 2.1). In step 1, I predicted genes targeted by each of the five miRNAs using RNAhybrid, which is fast and flexible software for miRNA target prediction, with the free-energy option and the seed-sequence option (Rehmsmeier et al. 2004). The RNA duplex free-energy filter was defined as the appropriate value that led to the efficient extraction of experimentally verified miRNA/target-gene pairs. I also considered a complete match across the seed sequence (1–7, 2–8, or 3–9 nucleotides from the miRNA's 5′ end), which was used as a filter by adding the seed option of RNAhybrid.

In step 2, I used four binding parameters of the hybridization pattern of the miRNA/messenger RNA (mRNA) duplexes. According to a recent study, a binding rule is likely to exist for the recognition of target mRNAs by miRNAs (Kiriakidou et al. 2004). Moreover, G-U wobble pairs within miRNA/mRNA duplexes play a key role in the interaction with target mRNAs (Lai 2005). Subsequently, potential candidates were extracted using four binding parameters (number of mismatches of mRNA within the whole miRNA sequence, number of mismatches of miRNA within the whole miRNA sequence, number of G-U wobble pairs within the whole miRNA sequence, and

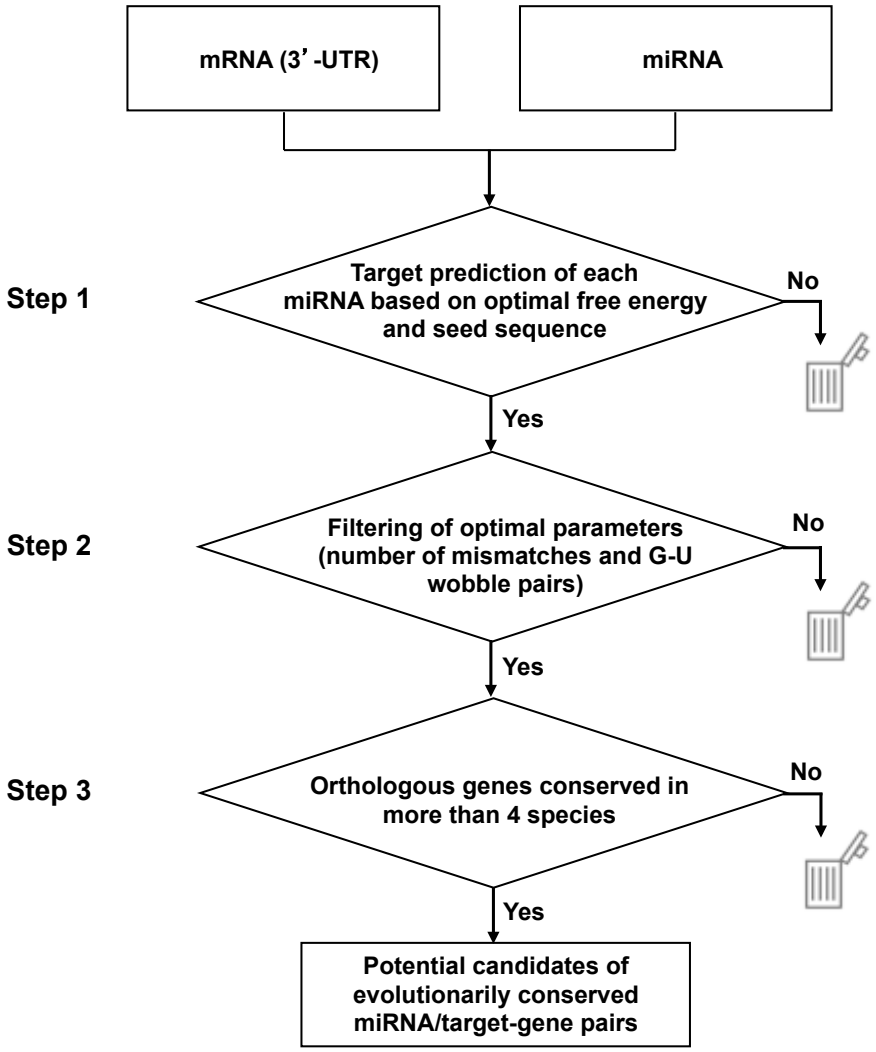


Figure 2.1 Computational extraction of conserved miRNA/target-gene pairs among bilaterian animals

Evolutionarily conserved miRNAs were extracted from the five model species (*H. sapiens*, *M. musculus*, *G. gallus*, *D. melanogaster*, and *C. elegans*). For each miRNA, potential target genes were predicted using the following criteria: optimal free-energy threshold and complete matching of nucleotide sequences between the seed sequence of miRNA/mRNA duplexes (step 1), binding pattern of the 3′-UTR of miRNA/mRNA duplexes (step 2), and orthologous gene information (step 3).

number of G-U wobble pairs within the seed sequence) of the miRNA/mRNA duplexes (Figure 2.2). I used the hybridization pattern of experimentally verified and predicted miRNA/mRNA pairs to calculate coverage, by changing these binding parameters one by one (28–0 for the number of mismatches of mRNA within the whole miRNA sequence; 14–0 for the number of mismatches of miRNA within the whole miRNA sequence; 10–0 for the number of G-U wobble pairs within the whole miRNA sequence; and 6–0 for the number of G-U wobble pairs within the seed sequence). The range of each of the four binding parameters was determined based on the coverage of experimentally verified and predicted miRNA/mRNA pairs. Parameter combinations were then plotted on a 2D graph by calculating the "ratio of experimentally verified miRNA/mRNA" and "Enrichment" using the points in the four parameter ranges. The criterion "Enrichment" was defined and calculated as the value of "ratio of experimentally verified miRNA/target-gene pairs" divided by the "ratio of predicted miRNA/target-gene pairs". I obtained the most effective combination of four binding parameters for extracting miRNA/target-gene pairs based on the EC value (multiplied value of Enrichment and Coverage). Parameter combination with highest EC value was selected.

In step 3, orthologous gene information was used to extract orthologous genes targeted by the same type of miRNA. For the retrieval of evolutionarily conserved miRNA/target-gene pairs from various bilaterian animals, I set the orthologous gene information criteria as orthologous genes conserved in at least four species, each containing the miRNA target site of interest.

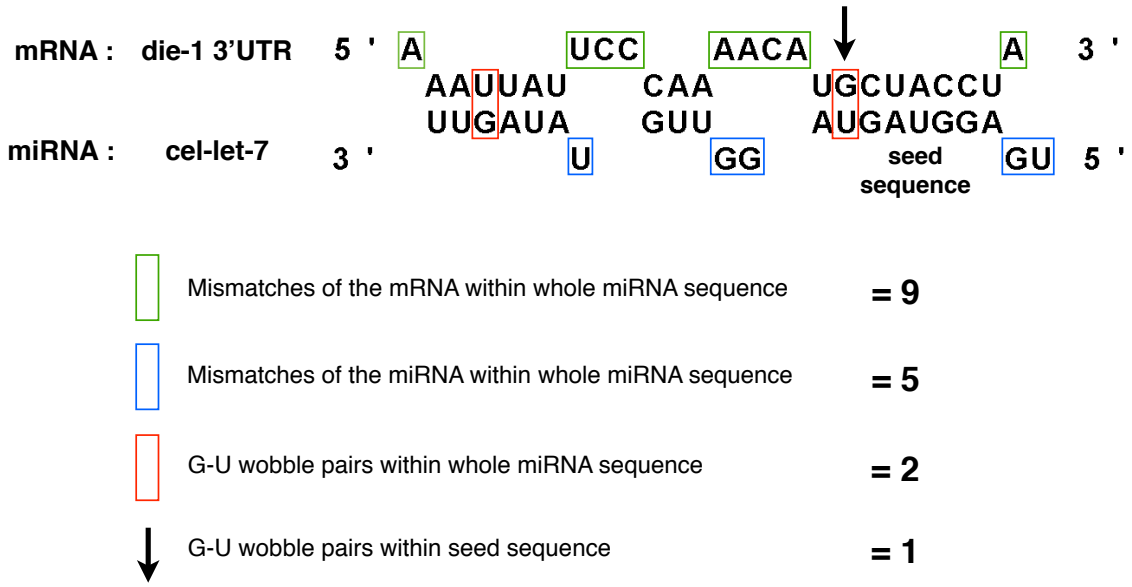


Figure 2.2 Basic concept of miRNA/mRNA duplex formation

An example of the binding pattern of miRNA (bottom)/mRNA (top) duplexes is shown using cel-let-7 and die-1 3′-UTR sequences from *C. elegans*. The green and blue squares depict mismatched nucleotide sequences of the mRNA and miRNA, respectively. The red square depicts G-U wobble pairs within the whole miRNA sequence and the black arrow pinpoints a G-U wobble pair within the seed sequence.

2.2.4 Expression vectors

To construct target-site reporter plasmids, each DNA fragment (3′-UTR sequence of the delta-like protein 1 precursor *(DLL1*) gene (668 nucleotides (nt); accession no. AF003522), ETS domain-containing protein Elk-3 (*ELK3*) gene (519 nt; accession no. BC017371), eukaryotic translation initiation factor 2C4 (*EIF2C4*) gene (2148 nt; accession no. AB046787), transgelin-2 (*TAGLN2*) gene (1391 nt; accession no. D21261), La-related protein 4 (*LARP4*) gene (1678 nt; accession no. AY004310), calponin-3 (*CNN3*) gene (1391 nt; accession no. BC025372), and V-type proton ATPase subunit B brain isoform (*ATP6V1B2*) gene (1208 nt; accession no. L35249)) was amplified from HeLa genomic DNA *via* polymerase chain reaction using site-specific primers and was inserted into the *Xho*I/*Not*I sites of the psiCHECK-2 plasmid vector (which encodes both firefly and *Renilla* luciferases; Promega, Madison, WI, USA). The oligonucleotide was designed to introduce *Xho*I and *Not*I sites at the 5′ and 3′ termini, respectively. The resulting plasmids were termed pLuc-DLL1, pLuc-CNN3, pLuc-LARP4, pLuc-ELK3, pLuc-EIF2C4, pLuc-TAGLN2, and pLuc-ATP6V1B2, respectively. miRIDIAN™ miRNA Mimic for hsa-let-7a, hsa-miR-1, hsa-miR-124, hsa-miR-34a, and negative control (miRIDIAN microRNA Hairpin Inhibitor Negative Control #1) were purchased from Dharmacon. miRNA Mimic molecules are chemically modified double-stranded RNA oligonucleotides.

The sequences of the oligodeoxyribonucleotides used for PCR were as follows:

(1) DLL1_S: 5′–TTACTCGAGAATGGAAGTGAGATGGCAAGAC–3′

(2) DLLl_A: 5′–TTAGCGGCCGCTTGTCATTCATAAAATTTATTT–3′

(3) CNN3_S: 5′–TTACTCGAGTCCACACAGAAGGAGCTCAG–3′

(4) CNN3_A: 5′–TTAGCGGCCGCCAGGAAGAGCAAATGCATCA–3′

(5) LARP4_S: 5′–TTACTCGAGGTTCCCATTTGATGGCATGT–3′

(6) LARP4_A: 5′–TTAGCGGCCGCCATAGCACCTTGGCGATGTT–3′

(7) ELK3_S: 5′–TTACTCGAGCGTCTGGCCACAATTAAGGA–3′

(8) ELK3_A: 5′–TTAGCGGCCGCTGCTTTCATATTGCCCACTG–3′

(9) EIF2C4_S: 5′–TTACTCGAGTTCTACCAGCAGCTCGGAAT–3′

(10) EIF2C4_A: 5′–TTAGCGGCCGCTTGGATTCCAGCAAGTCCTC–3′

(11) TAGLN2_S: 5′–TTACTCGAGCCCTCCCACGAATGGTTAAT–3′

(12) TAGLN2_A: 5′–TTAGCGGCCGCATGGAAAATGAGAAGCCACG–3′

(13) ATP6V1B2_S: 5′–TTACTCGAGTCCGCGCTCTTGTGAAATAC–3′

(14) ATP6V1B2_A: 5′–TTAGCGGCCGCATAATCATGCTGACTCCCCC–3′

2.2.5 Transfection and luciferase reporter assay

Transient transfection and luciferase assays were performed as described previously, with slight modifications (Lewis et al. 2003). Briefly, HeLa cells were grown in 10% FBS in DMEM and seeded in 24-well plates 24 h before transfection. Cells were transfected with the indicated amounts of reporter and miRNA Mimic (100 ng of target reporter and 5, 20, and 60 pmol of miRNA Mimic) in the presence of Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Firefly and *Renilla* luciferase activities were measured consecutively using the Dual-luciferase assay system (Promega) 24 h after

transfection, according to the manufacturer's instructions.

2.3 Results and discussion

2.3.1 Extraction of evolutionarily conserved miRNAs among five bilaterian animals

To extract conserved miRNA/target-gene pairs, I chose five model species (*H. sapiens*, *M. musculus*, *G. gallus*, *D. melanogaster*, and *C. elegans*) among bilaterian animals, for which there exists a vast array of data on both miRNAs and mRNAs (Griffiths-Jones et al. 2008). Previously, several important features were described to classify miRNAs into families. It is well known that the seed sequence (the 5′ side of the miRNA sequence) is important for interaction with the target mRNAs (Lewis et al. 2003). Many miRNA target prediction software programs were developed using the features of seed sequences (John et al. 2004; Rehmsmeier et al. 2004; Krek et al. 2005; Thadani and Tammi 2006; Kertesz et al. 2007; Friedman et al. 2009). Moreover, several features have been proposed to identify conserved miRNA families, such as conservation of the mature miRNA sequence (features of the earliest miRNA classification in miRBase) (Griffiths-Jones 2004) and information on the phylogenetic relationship among miRNAs (Huang and Gu 2007). By focusing on these features, I proposed the following criteria for extracting well-conserved miRNA families among five species: (1) complete seed sequence matching, (2) mature miRNA sequence identity exceeding 75%, and (3) high conservation among miRNA families, considering the phylogenetic relationship among miRNAs (category I) (Huang and Gu 2007). Consequently, from 2,404 mature miRNA sequences, I extracted five miRNA families (let-7, miR-1, miR-124,

miR-125/lin-4, and miR-34) conserved evolutionarily among the five bilaterian animals (Table 2.1). The sequence identity among most of the conserved miRNA families was over 80%. In particular, the sequences of let-7 and miR-1 family members showed very high mature miRNA sequence identity (exceeding 90%) among all five bilaterian species, which suggests that a strong selective pressure exists for the nucleotide sequence. Huang et al. described the extraction of 15 conserved miRNA families among six bilaterian animals (*H. sapiens*, *M. musculus*, *G. gallus*, *D. rerio*, *D. melanogaster*, and *C. elegans*) based on their original classification method (Huang and Gu 2007), which included five miRNA families I defined. In this chapter, I devised more stringent criteria based on nucleotide conservation to extract highly conserved miRNA families. *Nematostella vectensis* and *Amphimedon queenslandica*, which diverged before the emergence of bilaterian animals, reportedly express various types of miRNAs (Grimson et al. 2008); however, none of these miRNA sequences are similar to the five evolutionarily conserved miRNA families found in the current study (data not shown), which suggests that these evolutionarily conserved miRNAs appeared after the divergence of bilaterian animals or were lost in *N. vectensis* and *A. queenslandica.*

2.3.2 Filtering and enrichment of the evolutionarily conserved miRNA/target-gene pairs

To extract genes targeted through evolution by the five conserved miRNAs, I designed a procedure that comprised three screening steps (Figure 2.1 and Table 2.2). In step 1, I extracted potential target genes based on optimal free-energy information with a

Table 2.1 List of miRNAs conserved among various bilaterian animals

	Species					Identity (%)
miRNA family	***H. sapiens***	***M. musculus***	***G. gallus***	***D. melanogaster***	***C. elegans***	
let-7	let-7a	let-7a	let-7a	let-7	let-7	95.5
miR-1	miR-1	miR-1	miR-1a	miR-1	miR-1	90.9
miR-124	miR-124	miR-124	miR-124a	miR-124	miR-124	82.6
miR-125/lin-4	miR-125b	miR-125b-5p	miR-125b	miR-125	lin-4	81.8
miR-34	miR-34a	miR-34a	miR-34a	miR-34	miR-34	79.2

Table 2.2 Summary of target gene extraction after each screening step

	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*	Ratio of predicted miRNA/target-gene pairs (%)	Ratio of experimentally verified miRNA/target-gene pairs (%)	Enrichment
All possible miRNA/gene pairs	97820	98790	53995	52850	53975	357430/357430 (100)	145/145 (100)	1
STEP 1 (Free energy and seed region)	56666	53420	20410	14098	8793	153387/357430 (42.9)	112/145 (77.2)	1.8
STEP 2 (Optimal parameter)	12405	11256	2524	1118	898	28201/357430 (7.9)	76/145 (52.4)	6.6
STEP 3			31			31/10356* (0.3)	4/52 (7.7)	25.7

The total number of possible target genes for the five miRNAs (let-7, miR-1, miR-124, miR-125/lin-4, and miR-34) is represented for the five model species at each screening step (see Figure 2.1). Asterisks indicate the number of orthologous genes conserved among at least four species. Enrichment was calculated as the "ratio of experimentally verified miRNA/target-gene pairs" divided by the "ratio of predicted miRNA/target-gene pairs".

requirement of complete seed-sequence-matching using the RNAhybrid software (Rehmsmeier et al. 2004), which predicts potential binding sites of short RNAs among target sequences. For this purpose, the optimal free energy was determined beforehand to efficiently cover experimentally validated miRNA/target-gene pairs. I calculated the free energy of 139 experimentally validated miRNA/target-gene pairs with a complete match of seed sequence, as assessed using RNAhybrid; subsequently, I defined the optimal free energy as <–17 kcal/mole of potential miRNA/target-gene pairs (data not shown). The evaluation of target-gene extraction during each step was carried out using an index termed "Enrichment" (see Materials and Methods). As a result, in step 1, the number of potential target-gene candidates decreased from 357,430 to 153,387, with an enrichment index of 1.8. This prediction set contained 112 of the 145 experimentally validated target-gene pairs.

In step 2, I considered the 3′-UTR binding pattern of the miRNA/target mRNA. I defined four binding parameters (i.e., number of mismatches of mRNA within the whole miRNA sequence, number of mismatches of miRNA within the whole miRNA sequence, number of G-U wobble pairs within the whole miRNA sequence, and number of G-U wobble pairs within the seed sequence) of the hybridization pattern (Figure 2.2) for optimisation of the thresholds for each of the features used to predict reliable miRNA/mRNA pairs. The ranges of the four binding parameters were determined by calculating the coverage of 112 miRNA–mRNA pairs verified experimentally and of 153,387 miRNA–mRNA pairs predicted for each binding feature (Figure 2.3A–D). Five hundred parameter combinations were plotted on a 2D graph,

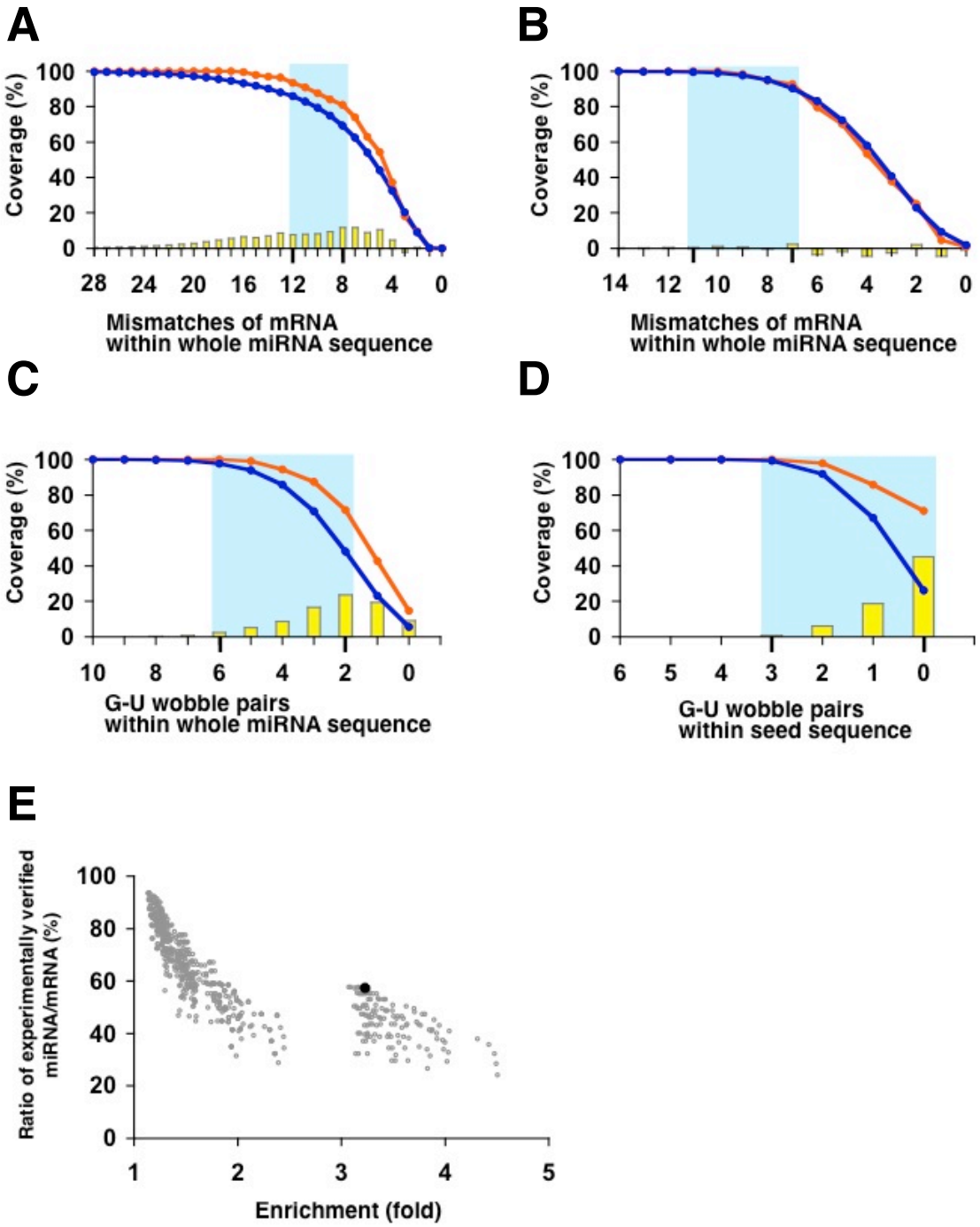


Figure 2.3 Parameters used for the prediction of miRNA/mRNA pairs and their coverage

To optimize the binding parameters of miRNA/mRNA duplexes, I determined the coverage of four binding parameters (mismatch of mRNA within the whole miRNA sequence (A), mismatch of miRNA within the whole miRNA sequence (B), G-U wobble pairs within the whole miRNA sequence (C), and G-U wobble pairs within the seed sequence (D)). Calculation of the coverage was performed using 112 experimentally verified miRNA/mRNA pairs (orange line) and 153,387 predicted miRNA/mRNA pairs (blue line). The yellow bar indicates

(Legend continued on next page)

differences in coverage between experimentally verified and computationally predicted miRNA/mRNA pairs. Four or five points chosen from the highest yellow bar were used as the range of each of the four binding parameters (blue rectangles) used in this study. I determined the parameter space using the binding patterns of the miRNA/mRNA pairs based on four features (E). Five hundred parameter combinations were plotted on a 2D graph using "ratio of experimentally verified miRNA/mRNA" on the Y-axis and "Enrichment" on the X-axis. The black circle (57.3% of the coverage and 3.2-fold of the Enrichment) indicates the point that corresponded to optimized parameters for the prediction of final conserved miRNA/target pairs: 12 for the number of mismatches in the mRNA, 10 for the number of mismatches in the miRNA, 4 for the number of G-U wobble pairs within the whole miRNA sequence, and 0 for the number of G-U wobble pairs within the seed sequence (see Materials and Methods).

using “Enrichment” on the X-axis and “Ratio of experimentally verified miRNA/mRNA” on the Y-axis (Figure 2.3E). From these parameter combinations, I defined the optimal combination of binding parameters for efficient screening based on a maximum EC value (multiplied value of Enrichment and Coverage) of 184.7 (number of mismatches of mRNA within the whole miRNA sequence: 12; number of mismatches of miRNA within the whole miRNA sequence: 10; number of G-U wobble pairs within the whole miRNA sequence: 4; and number of G-U wobble pairs within the seed sequence: 0). Accordingly, the number of potential target-gene candidates was reduced from 153,387 to 28,201 and the experimentally validated target genes decreased from 112 to 76 after introduction of the criterion of optimal hybridization pattern (Enrichment index, 6.6) (Table 2.2).

Finally, in step 3, I incorporated orthologous gene information and extracted genes that were evolutionarily conserved among four or five diverse bilaterian animals, including *H. sapiens*. As a result, the number of predicted miRNA/target-gene pairs was minimized substantially, from 10,356 to 31, using a significantly high Enrichment index of 25.7 (Table 2.2). The number of predicted miRNA/target-gene pairs was especially high for the three miRNAs let-7 (eight targets), miR-1 (seven targets), and miR-124 (eleven targets) compared with miR-125/lin-4 (three targets) and miR-34 (two targets) (Table 2.3). This suggests that let-7, miR-1, and miR-124 may have played a major role in primordial miRNA gene regulation in the common bilaterian ancestor. To verify the significance of conserved miRNA/target-gene pairs, I performed same sequence analysis (from step 1 to step 3) against total 25 species-specific miRNAs (5 miRNAs

Table 2.3 Summary of the number of target genes in each extraction step

A. let-7

	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*
Step 1	14099	13522	5353	3168	2416
Step 2	2064	1870	360	207	161
Step 3			8		

B. miR-1

	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*
Step 1	6958	6006	1811	2658	1547
Step 2	1807	1549	395	378	333
Step 3			7		

C. miR-124

	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*
Step 1	10813	10534	4248	2247	1611
Step 2	2881	2591	667	222	237
Step 3			11		

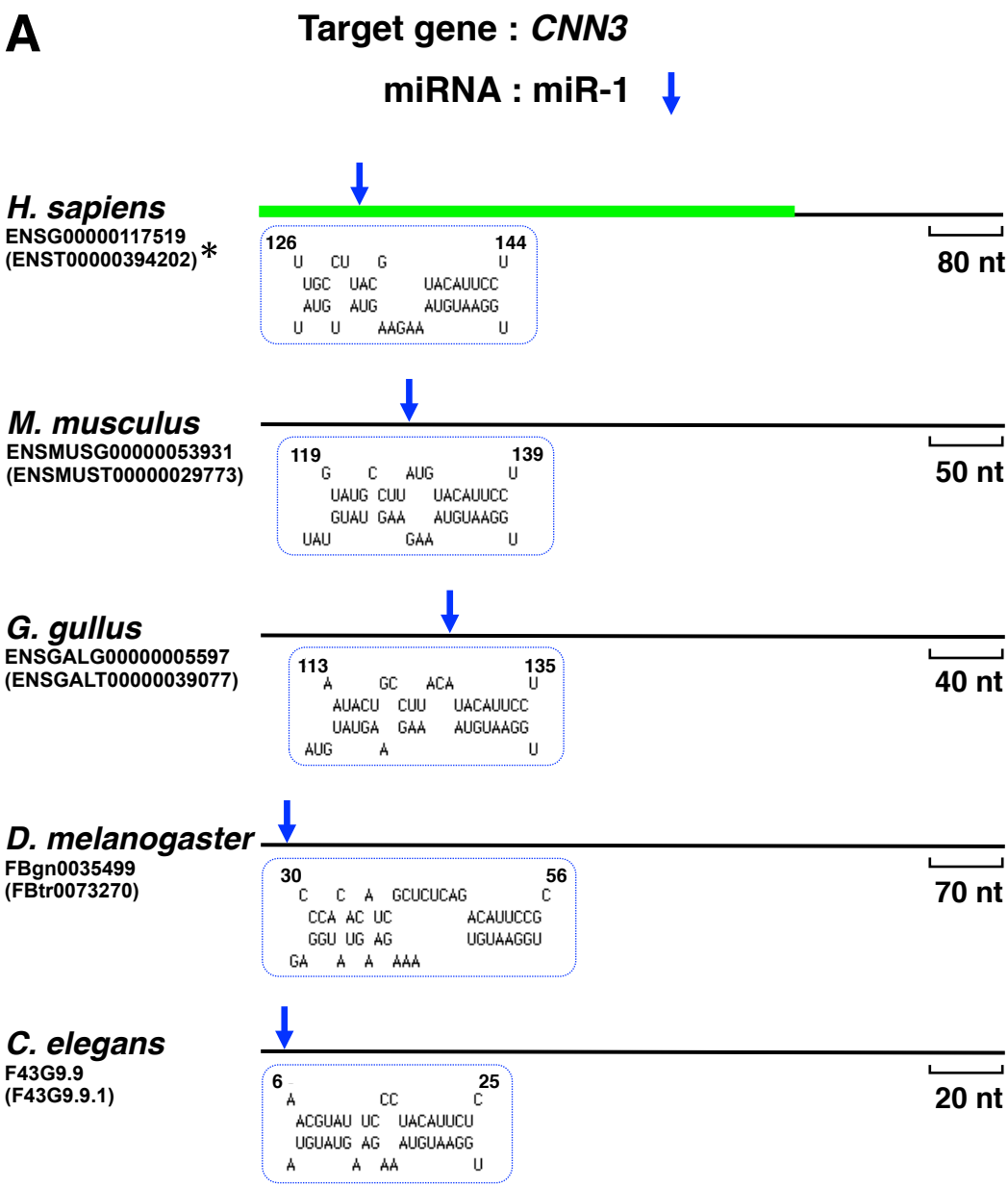
D. miR-125/lin-4

	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*
Step 1	9854	9013	2282	917	408
Step 2	2855	2557	436	82	61
Step 3			3		

E. miR-34

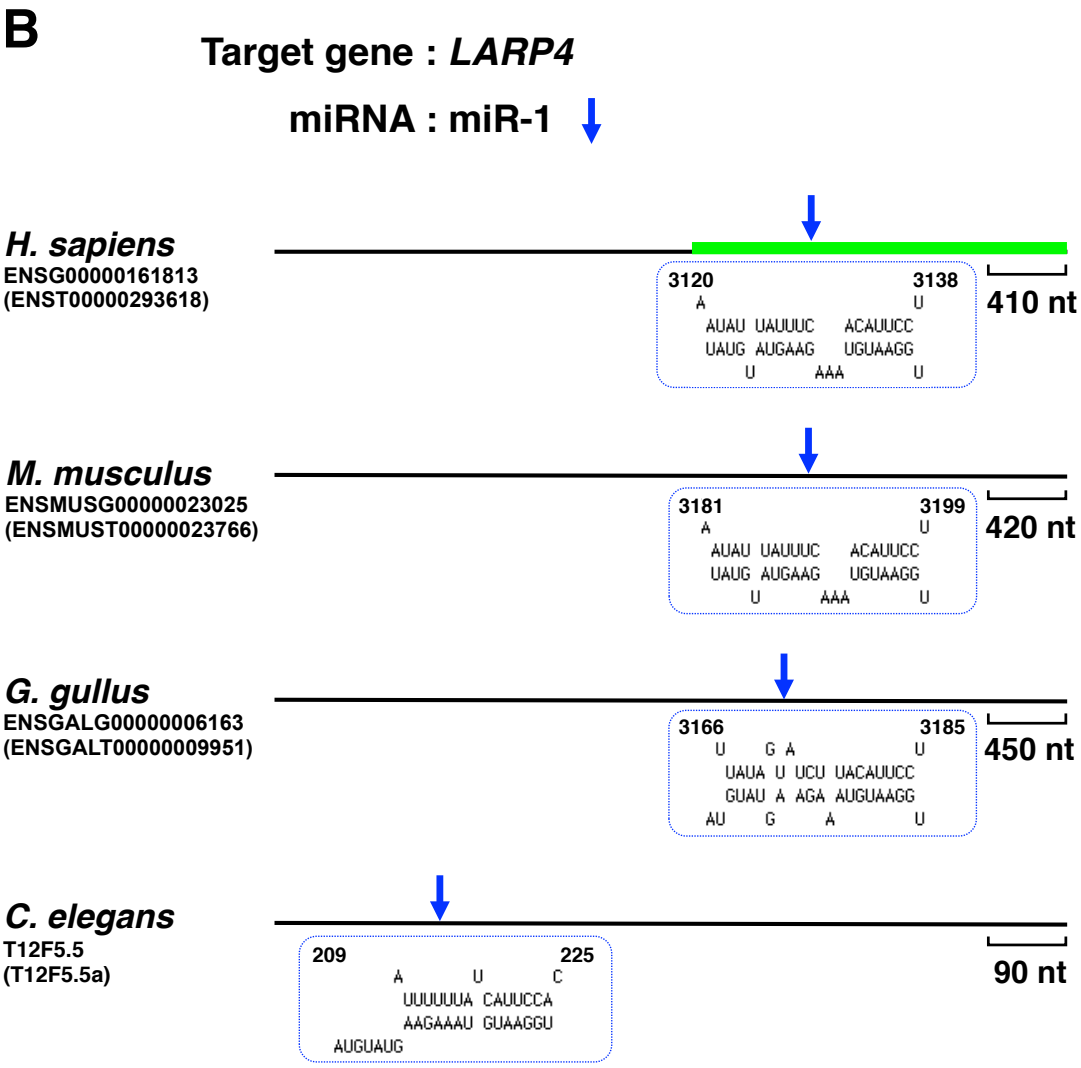
	H. sapiens	*M. musculus*	*G. gallus*	*D. melanogaster*	*C. elegans*
Step 1	14942	14345	6716	5108	2811
Step 2	2798	2689	666	229	106
Step 3			2		

each from 5 species) as a control experiment supposing that these miRNAs are also conserved in other bilaterians. For example, target prediction of hsa-miR-2277, a species-specific miRNA in human was performed in all 5 species (step 1 and step 2) and conserved targets were extracted (step 3). As a result, 11 out of 25 non-conserved miRNAs did not show any conserved miRNA/target-gene pair. Furthermore, number of the miRNA/target-gene pairs of the negative control is statistically lower than that of conserved miRNA/target-gene pairs based on the Welch's t-test ($p<0.05$). These results support that number of genes achieved from the prediction of conserved miRNAs target genes in this study is indeed significant. In summary, I developed a new filtering method for extracting evolutionarily conserved miRNA/target-gene pairs, which was used to extract 31 reliable miRNA/target-gene pairs among the five families of miRNAs. I discovered that only one orthologous target gene, *CNN3*, was conserved completely among the five bilaterian animals (Figure 2.4A). As for target genes conserved in four species, I found, for example, the *LARP4, ELK3*, *EIF2C4, TAGLN2*, and *ATP6V1B2* genes (Figure 2.4B–F). Of note, the same approximate position of the predicted target site was observed in the orthologous 3′-UTR of *CNN3* (120 nt)*, LARP4* (3,100 nt), *EIF2C4* (220 nt), and *TAGLN2* (50 nt) among vertebrates (Figure 2.4A–B, D–E). According to Grimson et al., the distribution of miRNA target sites within the 3′-UTR is biased near the mRNA stop codon or poly-A tail compared with the middle portion of 3′-UTR (Grimson et al. 2007). These results show that target site distribution varied according to the type of miRNA target gene. The target sites on 3′-UTR of *CNN3* and *TAGLN2* were biased near the stop codon, from *H. sapiens* to *C. elegans*



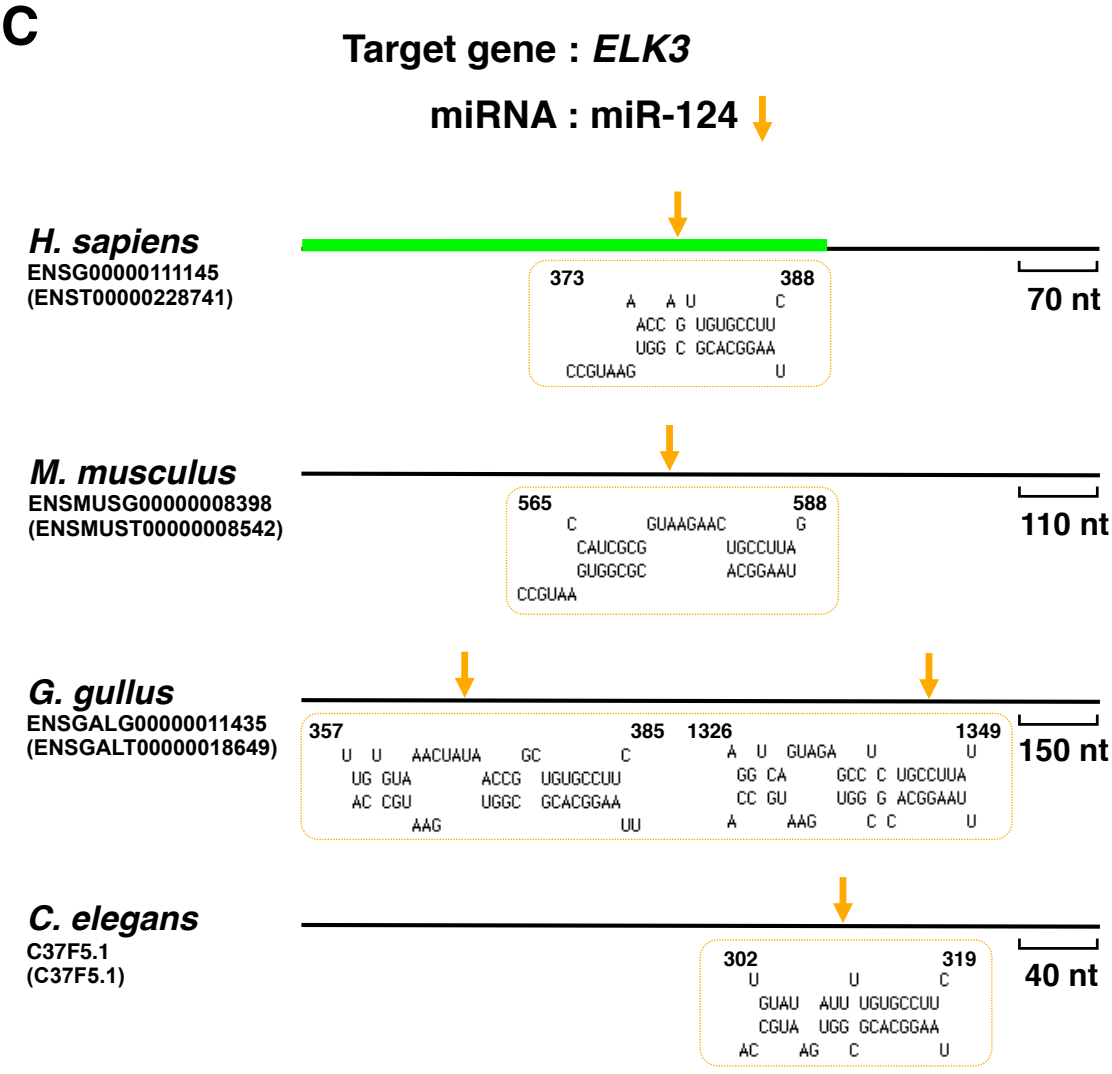
(Legend on P41)

B

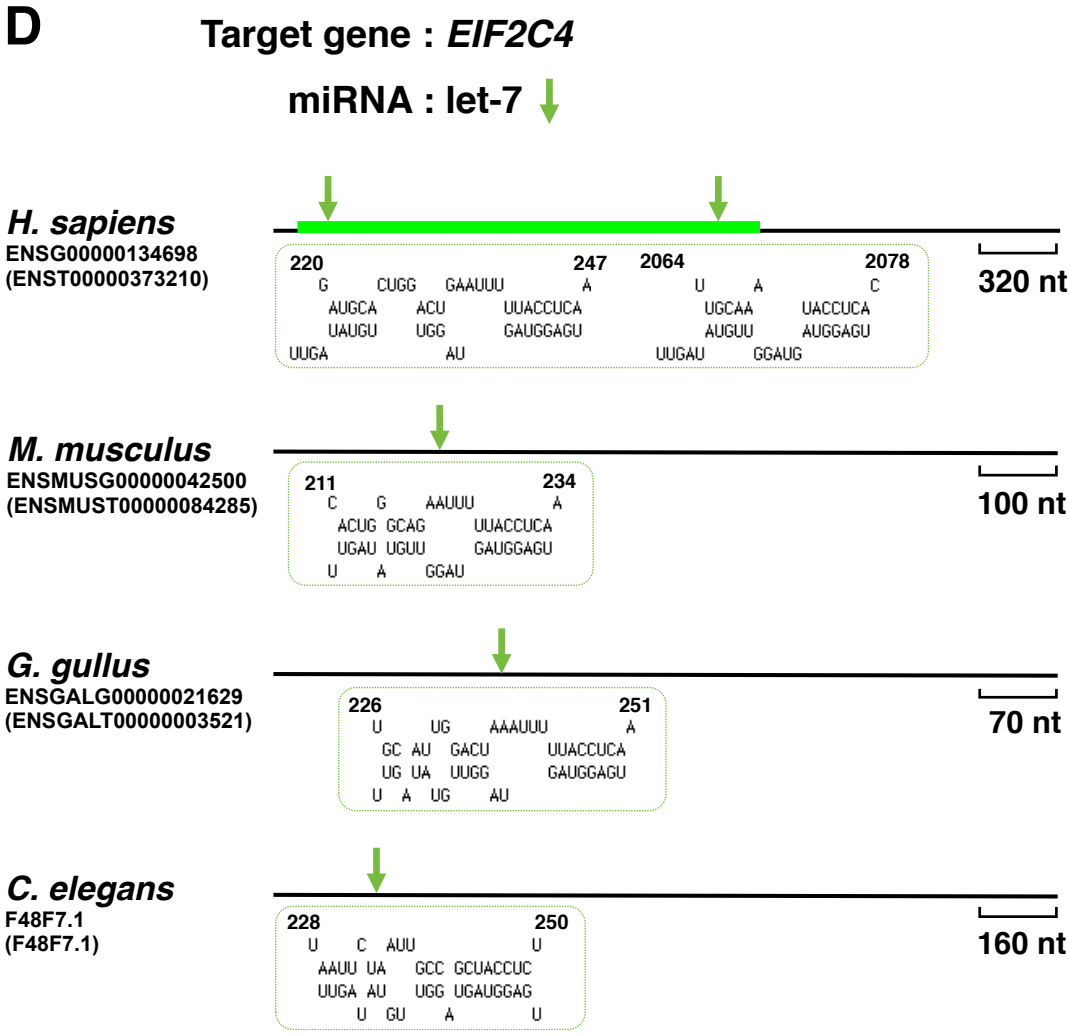


(Legend on P41)

C

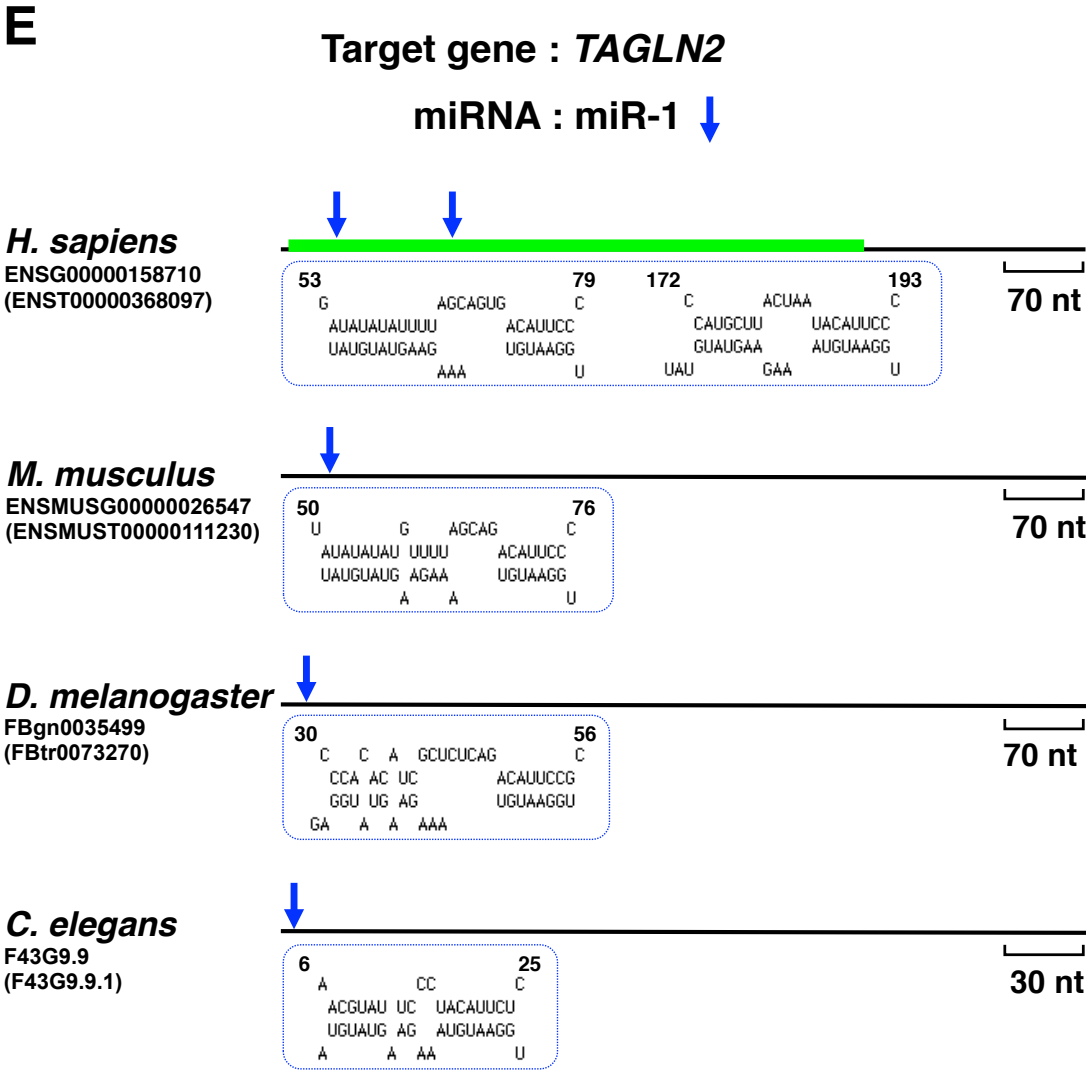


(Legend on P41)



(Legend on P41)

E



(Legend on P41)

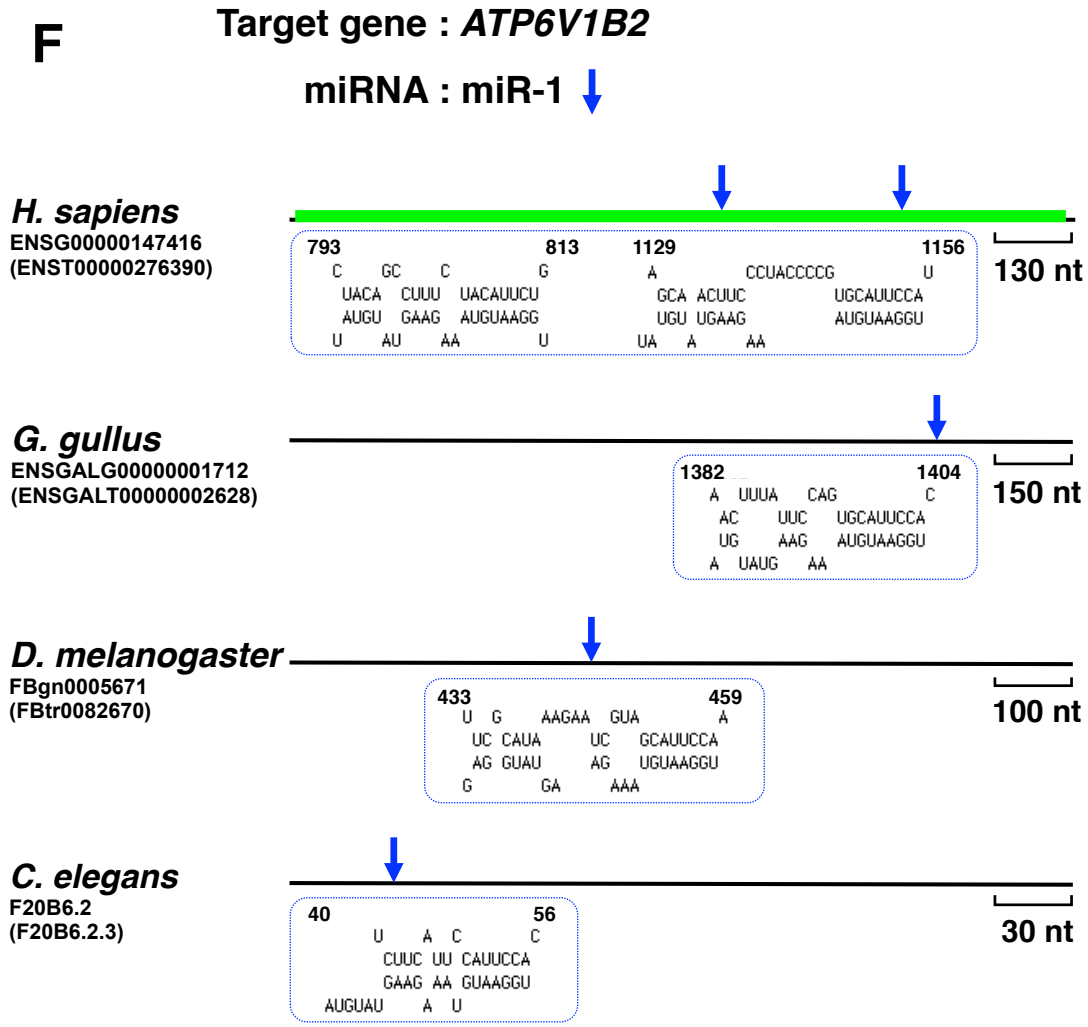


Figure 2.4 Six examples of miRNA target sites in orthologous gene transcripts used for experimental verification

Potential target sites of miRNAs in the 3′-UTR sequences of the orthologous transcripts; (A) miR-1 (blue arrows) for *CNN3,* (B) miR-1 for *LARP4,* (C) miR-124 (orange arrows) for *ELK3* (D) let-7 (green arrows) for *EIF2C4*, (E) miR-1 for *TAGLN2*, (F) miR-1 for *ATP6V1B2.* Predicted duplexes formed by the 3′-UTR sequences (top) and miRNAs (bottom) are shown in dotted boxes for each potential target site. The green bar on the *H. sapiens* 3′-UTR sequence indicates a DNA region used for the construction of the reporter plasmid. (*) The length of the CNN3 3′-UTR is currently registered as a little shorter than that indicated (527 nt in size) and contains the miR-1 binding site (Ensembl release 53).

(Figure 2.4A, E). Regarding the other candidates, I observed all types of target site distribution on 3'-UTR. A future statistical analysis of miRNA target-site distribution among conserved miRNA/target-gene pairs is required to substantiate this view. With the exception of 3′-UTR of the *LARP4* gene, most of the binding patterns of evolutionarily conserved target sites were different in sequence, without taking the seed region into consideration. The target-site binding patterns within 3′-UTR of the *LARP4* gene were identical between *H. sapiens* and *M. musculus* (Figure 2.4B), although the similarity of the two 3′-UTR sequences was ~70% (data not shown). A recent study reported on cooperative regulation by multiple miRNAs (Krek et al. 2005). Likewise, the erythrocyte membrane protein band 4.1 like 4B (*EPB41L4B*) gene was an orthologous target of two different types of miRNAs: miR-1 and miR-124 (Table 2.4). These analyses suggest that multiple miRNA regulation may have already existed in the era of ancestral bilaterian species.

2.3.3 Experimental validation of miRNA target genes

To validate the evolutionarily conserved miRNA/target-gene candidates, I performed transfection and luciferase assays on 6 of the 31 identified evolutionarily conserved miRNA/target-gene candidates. Four (*CNN3, LARP4, TAGLN2*, and *ATP6V1B2*) of the six candidates were predicted to be regulated by miR-1, while one candidate (*ELK3*) was predicted to be targeted by miR-124, and the final candidate (*EIF2C4*) was predicted to be the target of let-7. I used the well-established downregulation of the *DLL1* gene by miR-34 as a positive control (Lewis et al. 2003), and the let-7/*LARP4*,

Table 2.4 Evolutionarily conserved genes regulated by miRNAs

Target gene			let-7	miR-1	miR-124	miR-125 /lin4	miR-34
ID	Name	Function					
Candidate targeted by multiple miRNAs							
ENSG00000095203	EPB41L4B	Band 4.1-like protein 4B		+**	+		
Candidate targeted by single miRNAs							
ENSG00000187772	LIN28B	Lin-28 homolog B	+				
ENSG00000198799	LRIG2	Leucine-rich repeats and immunoglobulin-like domains protein 2 precursor	+				
ENSG00000086544	ITPKC	Inositol-trisphosphate 3-kinase C	+				
ENSG00000196233	LCOR	Ligand-dependent corepressor	+				
ENSG00000139263	LRIG3	Leucine-rich repeats and immunoglobulin-like domains protein 3 precursor	+				
ENSG00000134698	EIF2C4	Eukaryotic translation initiation factor 2C 4	+				
ENSG00000170456	DENND5B	MGC24039 protein	+				
ENSG00000136231	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	+				
ENSG00000158710	TAGLN2	Transgelin-2		+*			
ENSG00000143549	TPM3	Tropomyosin alpha-3 chain		+*			
ENSG00000117519	CNN3	Calponin-3		+			
ENSG00000071073	MGAT4A	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A		+			
ENSG00000161813	LARP4	La-related protein 4		+			
ENSG00000147416	ATP6V1B2	Vacuolar ATP synthase subunit B, brain isoform		+			
ENSG00000135862	LAMC1	Laminin subunit gamma-1 precursor			+*		
ENSG00000125695	STRADA	STE20-related adapter protein			+		
ENSG00000131459	GFPT2	Glucosamine--fructose-6-phosphate aminotransferase			+		
ENSG00000111145	ELK3	ETS domain-containing protein Elk-3			+		
ENSG00000151726	ACSL1	Long-chain-fatty-acid--CoA ligase 1			+		
ENSG00000164144	ARFIP1	Arfaptin-1			+**		
ENSG00000080819	CPOX	Coproporphyrinogen III oxidase, mitochondrial precursor			+		
ENSG00000138279	ANXA7	Annexin A7			+		
ENSG00000150093	ITGB1	Integrin beta-1 precursor			+**		
ENSG00000093167	LRRFIP2	Leucine-rich repeat flightless-interacting protein 2			+		
ENSG00000116141	MARK1	Serine/threonine-protein kinase MARK1				+	
ENSG00000166797	FAM96A	Protein FAM96A				+	
ENSG00000131914	LIN28	Lin-28 homolog A				+*	
ENSG00000113282	CLINT1	EPN4_HUMAN Isoform 2 of Q14677 - Homo sapiens					+
ENSG00000137872	SEMA6D	Semaphorin-6D precursor					+

List of conserved targets regulated by conserved miRNAs. Transcript of Band 4.1-like protein 4B is only regulated by two miRNAs (miR-1 and miR-124). "*" and "**" indicate experimentally verified miRNA–mRNA and possible miRNA–mRNA candidates revealed by microarray data, respectively.

which was extracted up to step 2 in this analysis, was chosen as a non-evolutionarily conserved pair. I subcloned the 3′-UTR sequence downstream from the *Renilla* luciferase gene (Figure 2.5A) and co-transfected 100 ng of the 3′-UTR reporter construct into HeLa cells using 5, 20, and 60 pmol of miRNA Mimics (hsa-let-7a, hsa-miR-1, hsa-miR-124, hsa-miR-34a, and miRIDIAN microRNA Hairpin Inhibitor Negative Control #1) (see Materials and Methods). I observed the downregulation of six out of six candidates and of the positive control compared with the negative controls (Figure 2.5B, a–g). Typical results of the reporter gene assay are shown in Figure 2.5B for the indicated amounts of miRNAs (5, 20, and 60 pmol). The downregulation of these candidates was significant (t test, $p<0.01$), although some of these pairs represented an inhibition of only 30% under the current conditions. Among these candidates, *TAGLN2* was previously suggested to be downregulated by miR-1, as assessed by microarray analysis (Lim et al. 2005). This feature was recently confirmed using the "pulsed stable isotope labelling with amino acids in cell culture" (pSILAC) method and a reporter gene assay (Selbach et al. 2008). Regarding the let-7/*LARP4* combination, the expression of *LARP4* was not downregulated after let-7 transfection, which was supported statistically (Figure 2.5B, h). These experimental results suggest that this new method has the potential for efficiently extracting reliable miRNA/target-gene pairs and may be effective in the elucidation of the primordial regulatory relationships between miRNAs and their target genes during the early stage of bilaterian evolution.

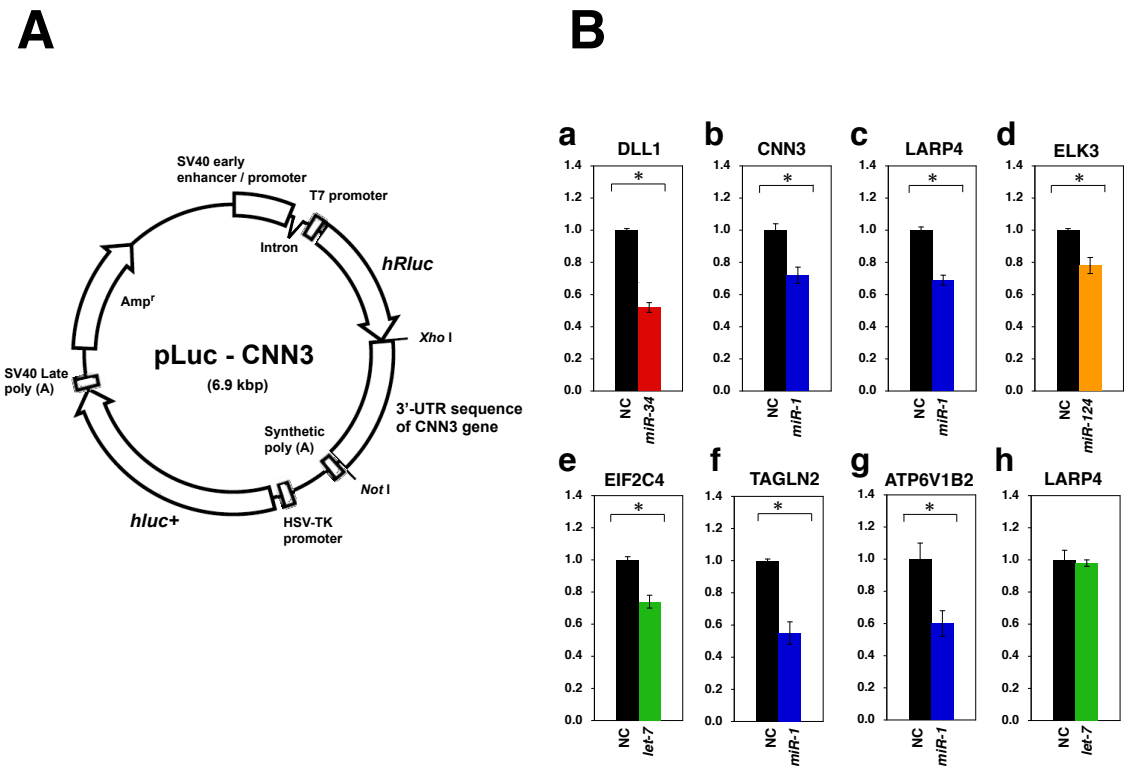


Figure 2.5 Example of the 3'-UTR reporter plasmid and experimental validation

The 3′-UTR sequences of *DLL1, CNN3, LARP4, ELK3, EIF2C4, TAGLN2,* and *ATP6V1B2* were subcloned into the *Xho*I/*Not*I site of the psiCHECK^{TM}-2 vector. *CNN3* was chosen as representative of the eight candidates listed above (see Materials and Methods). (B) HeLa cells were cotransfected with each combination of 100 ng of reporter plasmid and the indicated amounts of each miRNA ((a) *DLL1*, 5 pmol of miR-34; (b) *CNN3*, 60 pmol of miR-1; (c) *LARP4*, 20 pmol of miR-1; (d) *ELK3*, 60 pmol of miR-124; (e) *EIF2C4*, 60 pmol of let-7; (f) *TAGLN2*, 5 pmol of miR-1; (g) *ATP6V1V2*, 60 pmol of miR-1; (h) *LARP4*, 60 pmol of let-7). Colours depict each miRNA: miR-34 (red), miR-1 (blue), miR-124 (orange), let-7 (green), and negative control (black). The relative expression of the luciferase gene was measured 24 h after transfection. The normalized luciferase activity of the control vector was set as 1.0. The data represent the average of three experiments and SDs (standard deviations). * $p<0.01$.

2.3.4 Possible regulation of evolutionarily conserved miRNA targets in bilaterian animals

To provide further insight into the primary functions of evolutionarily conserved miRNAs (Table 2.4), I next focused on the functions of the target genes and found that evolutionarily conserved miRNA/target genes could be largely classified into four functional categories: development, differentiation, muscle movement, and gene regulation. First, I describe the function of evolutionarily conserved genes involved in development and differentiation. The laminin subunit gamma-1 precursor (*LAMC1*) gene, which was possibly regulated by miR-124, is one of the major components of the basement membrane. According to Smyth et al. (Smyth et al. 1999), null mutation of *LAMC1* causes embryonic lethality because of the absence of the basement membrane and failure to differentiate the endoderm. Among other candidates regulated by miR-124, the expression of the leucine-rich repeat flightless-interacting protein 2 (*LRRFIP2*) gene induces an extra axis in *Xenopus laevis* embryos (Liu et al. 2005). Moreover, the semaphorin-6D precursor (*SEMA6D*) gene, plays an important role in cardiac morphogenesis during chick embryonic development (Toyofuku et al. 2004), which was predicted as a candidate of miR-34 targeting. The *lin-28* gene, which regulates developmental timing in *C. elegans*, is reportedly controlled by lin-4, as assessed using *in vivo* experiments (Moss et al. 1997). Subsequently, the *lin-28* gene was also found to be regulated by miR-125, which is an orthologous miRNA of lin-4, in *H. sapiens* and *M. musculus* (Wu and Belasco 2005). This prediction confirmed the regulation of orthologous *lin-28* genes by lin-4/miR-125 miRNA in *H. sapiens*, *M.*

musculus, and *C. elegans* and further suggested that a similar regulatory relationship was conserved in *G. gallus*.

Next, I focused on the tissue-specific miRNA/target genes. Among the 31 evolutionarily conserved target candidates, approximately one-half were lowly expressed in a tissue-specific manner in humans, according to the BioGPS portal (http://biogps.gnf.org). miR-1 is highly expressed in muscle tissues (Chen et al. 2006). Here, three candidate genes regulated by miR-1 (i.e., *TAGLN2* and *CNN3*) are also lowly expressed in muscle tissues, according to BioGPS. *TAGLN2* is a homolog of *TAGLN*, which encodes an actin-binding protein (Shields et al. 2002). The *CNN3* gene also encodes an actin-binding protein that represses bone morphogenetic protein (BMP) signalling in chondrocytes, which is important for bone formation (Haag and Aigner 2007). The other miRNA, miR-124, is expressed in the nervous system (Yu et al. 2008). Similarly, the annexin A7 (*ANXA7*) gene, which is another candidate target of miR-124, is involved in the development of the murine brain (Rick et al. 2005). The striking overlap between the tissue specificity of evolutionarily conserved miRNA and that of their target genes suggests that one of the main functions of primordial miRNAs may have been the regulation of genes implicated in the temporary control of the development of muscle and of the nervous system, in a tissue-specific manner.

Finally, I found two interesting candidate genes, *EIF2C4* and *LARP4*, which encode translation-related proteins. It is well accepted that miRNAs are regulators of gene expression, mostly at the translational level (Bartel 2004). *EIF2C4* is also known as Argonaute 4 (*AGO4*). Although the function of AGO4 is unknown, other AGO

protein family members are involved in the miRNA-induced silencing complex (miRISC), which is essential for the miRNA or siRNA pathways. A previous microarray analysis performed in HepG2 cells revealed that the *EIF2C4* gene was affected by let-7 (Johnson et al. 2007). In the present study, I demonstrated for the first time the direct downregulation of *EIF2C4* by let-7, as assessed using a reporter gene assay in HeLa cells (Figure 2.5B, e); therefore, I speculate that negative-feedback regulation of *EIF2C4* by let-7 exists in the miRNA pathway. It has been reported that regulation of the *AGO1* mRNA, which is a major component of the miRISC, in the miRNA pathway by miR-168 controls plant development in *Arabidopsis thaliana* (Vaucheret et al. 2004). Another candidate, *LARP4*, encodes a member of La-motif protein family that controls translational efficiency (Bousquet-Antonelli and Deragon 2009). I also demonstrated the downregulation of the *LARP4* gene *via* miR-1 using a reporter gene assay (Figure 2.5B, c), which further supports my contention that some of the evolutionarily conserved miRNAs may play an important role in the regulation of translation by controlling the expression levels of translation factors and by negatively regulating their own miRNA pathway.

Chapter 3

Identification, expression, and molecular evolution of microRNAs in the "living fossil" *Triops cancriformis* (tadpole shrimp)

3.1 Introduction

Some microRNA (miRNA) genes are conserved in various species. I have previously shown that five miRNAs (let-7, miR-1, miR-124, miR-125/lin-4, and miR-34) are evolutionarily conserved in bilaterian animals, and that these conserved miRNAs target orthologous mRNAs in corresponding species (Chapter 2) (Takane et al. 2010). It has also been reported that *Nematostella vectensis* possesses almost no miRNAs with sequences similar to those of bilaterian animals (Grimson et al. 2008). Because the number of *N. vectensis* miRNAs is small compared to the number in most bilaterian animals, and because bilaterian animals are morphologically more complex than nonbilaterians, miRNAs are considered to be strongly related to the morphological evolution of animals (Grimson et al. 2008). Based on these observations, miRNAs have been studied not only in typical model species but also in non-model species with interesting characteristics. For instance, an evolutionary analysis of miRNAs was performed in the amphioxus *Branchiostoma belcheri,* which is a key animal in the evolution of chordates, and a phylogenetic analysis of miRNAs demonstrated that *B. belcheri* is more similar to vertebrates than to tunicates (Chen et al. 2009). In addition to analyses of miRNA sequences, it is also important to focus on the components of the RNA interference (RNAi) machinery. For example, *Ascaris suum* reportedly lacks the piRNAs, PIWI-clade Argonautes, and other proteins associated with the piRNA pathway, indicating that the piRNA pathway was lost in *A. suum* (Wang et al. 2011). These findings indicate that the investigation of miRNAs and components of RNAi

machinery in non-model species is important for elucidating their evolutionary pathways.

In this study, I focused on the tadpole shrimp *Triops cancriformis*. The morphological form of *T. cancriformis* is considered to have remained unchanged for 200 million years, making it a so-called "living fossil". In contrast, recent reports have indicated that *T. cancriformis* appeared less than 50 million years ago (Korn et al. 2013; Mathers et al. 2013), so the use of the term "living fossil" for *T. cancriformis* is controversial. The larvae of *T. cancriformis* change dramatically during development (Figure 3.1), progressing from the 1st to the 4th instar in only approximately 26 h, and at the same time doubling in size. Important morphological changes occur during that time, including the appearance of compound eyes and an increase in the number of body segments (Igarashi 1971). Despite these interesting characteristics of *T. cancriformis*, almost no genome or transcriptome data are currently available for this organism. To investigate the evolutionary history of miRNAs, I conducted a comparative analysis of miRNAs and the components of RNAi machinery. I identified conserved miRNAs and novel candidate miRNAs and deduced the components of the RNAi machinery in *T. cancriformis*. Most of the conserved miRNAs of *T. cancriformis* share sequence similarities with those of the arthropods, although *T. cancriformis* is called a "living fossil". However, a comparative analysis revealed that *T. cancriformis* let-7 and DICER1 are more similar to those of the vertebrates than to those of the arthropods, suggesting that the evolution of miRNA system in *T. cancriformis* has been unique, differing from those of other model species.

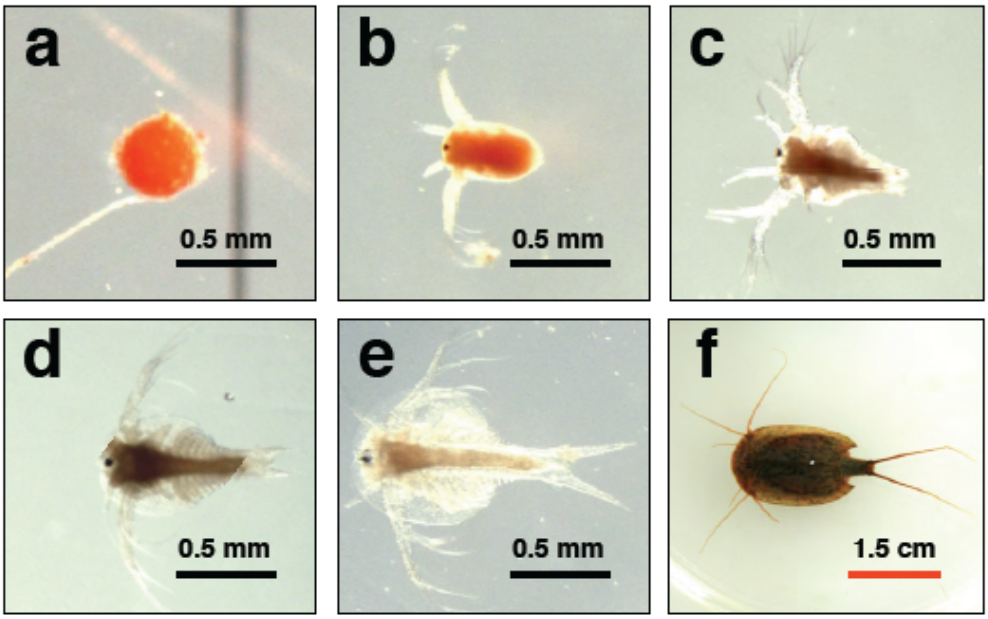


Figure 3.1 Morphological changes during *T. cancriformis* development

Scale bars represent 0.5 mm for egg and larvae from the 1st to 4th instar (black bars), and 1.5 cm for adult (red bar). a. Egg; b. 1st instar larva; c. 2nd instar larva; d. 3rd instar larva; e. 4st instar larva; f. adult.

3.2 Materials and methods

3.2.1 *T. cancriformis* culture

Triops cancriformis (adults and eggs) were obtained from two rice fields (Sakata, Yamagata, Japan and Higashitagawa-gun, Yamagata, Japan). Hundreds of eggs were placed in water and exposed to light for 24 h to enhance the hatching efficiency (Takahashi 1975). After hatching, the larvae were incubated at 20 °C and kept under a 14.5 h light: 9.5 h dark cycle. According to a previous study (Igarashi 1971), the developmental stages of *T. cancriformis* are defined based on average body length (1st instar: 530 μm; 2nd instar: 590 μm; 3rd instar: 800 μm; and 4th instar: 1180 μm). More concretely, the *T. cancriformis* larvae at each developmental stage were harvested after the appropriate incubation period after hatching (0–0.5 h for 1st instar, 3–8 h for 2nd instar, 13–22 h for 3rd instar, and 26–37 h for 4th instar larvae). Cultured *T. cancriformis* specimens (one month after hatching; body length: 1–3 cm) were used as the adult samples.

3.2.2 Deep-sequencing of *T. cancriformis* small RNA and genomic DNA

To construct the small RNA libraries, 500 eggs, 500 1st instar larvae, 500 2nd instar larvae, 500 3rd instar larvae, 500 4th instar larvae, and eight adults were used. Each sample was ground with a mortar and pestle, and the total RNAs were extracted with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to manufacturer's instructions. For the purpose of normalizing the reads among the developmental stages,

two small RNA spikes (spike1: p-AACUGUGUCUUUUCUGAAUAGA; and spike2: p-UAUUUAGAAUGGCGCUGAUCUG) corresponding to mammal-specific small RNAs (Wang et al. 2011) were added at different concentrations (spike 1, 20 fmol; spike 2, 0.5 fmol) to 20 μg of total RNA (or 30 μg for egg RNA only). The small RNA fraction (approximately 12–45 nucleotides (nt)), including miRNAs, was isolated and purified from these total RNA samples by gel electrophoresis. Takara Bio Incorporated (Shiga, Japan) constructed the small RNA libraries and performed the deep-sequencing analysis of the libraries. A cDNA library was constructed as described in a previous study (Pfeffer et al. 2005). Briefly, each small RNA was directly joined to the 3′ adaptor and 5′ adaptor with T4 RNA ligase. The ligation product was reverse transcribed and amplified with polymerase chain reaction (PCR). The nucleotide sequences were determined with the Illumina HiSeq 2000 (Illumina, San Diego, CA, USA).

To construct the genomic DNA library, two adult specimens were used. After the specimens were ground, the genomic DNA was extracted with the GNOME® DNA Isolation Kit (BIO101, La Jolla, CA, USA), according to the manufacturer's protocol, and treated with ribonuclease mix solution (Wako Pure Chemical Industries, Osaka, Japan) for 30 min at 37 °C. Takara Bio Incorporated constructed the genomic DNA library and performed the deep-sequencing analysis with Illumina HiSeq 2000.

3.2.3 Computational extraction of conserved *T. cancriformis* miRNAs

A six-step filtering approach was used to extract the reliable *T. cancriformis* candidate miRNAs from the small RNA deep-sequencing data. In step 1, unique sequences with

their associated count numbers were obtained and low-quality reads, sequence errors containing the character "N", low-quality base calls, and low count reads (<5) were discarded. In step 2, the small RNA reads that completely matched the genomic DNA contigs obtained from the deep sequences were retained. In step 3, sequence reads whose lengths corresponded to the miRNA fraction (18–24 nt) were extracted. In step 4, the sequence reads that mapped to the *T. cancriformis* cDNAs (3,981 expressed sequence tags and 579 genes) from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov, October 2013), or to the internal transcribed spacer (ITS) region (AB930494) between 18S ribosomal RNA (rRNA) and 28S rRNA were removed. In step 5, the BLASTN program (Camacho et al. 2009) was used to search the *T. cancriformis* miRNAs, and sequence reads were compared based on sequence similarities to already reported miRNA sequences from other species registered in miRBase Release 20.0 (June, 2013) (Kozomara and Griffiths-Jones 2013). Specifically, sequence reads with ≧80% sequence similarity to other reported miRNAs and with a complete seed match with at least two species (one of which was *T. cancriformis*) were extracted. In this study, I defined the seed sequence as occurring at either nucleotide positions 1–7, 2–8, or 3–9 from the 5′ side of the miRNA. In step 6, each genomic contig corresponding to the sequence read was analyzed with an RNA-folding program (Hofacker 2003) to determine whether the nucleotide sequence could form the potential secondary structure typical of precursor miRNAs. In this process, small RNA reads were mapped to the genomic DNA sequences corresponding to the miRNA precursor sequences. I confirmed that most abundant small RNA reads

were identical to the candidate miRNAs, and that the small RNA mapping patterns against the genomic DNA sequences did not resemble the degradation products of mRNAs. Finally, the conserved *T. cancriformis* miRNAs were designated "tcf-miR-XX", with identifying numbers (XX; e.g., tcf-miR-1) according to Ambros (Ambros et al. 2003).

3.2.4 Prediction of novel *T. cancriformis* candidate miRNAs

Novel *T. cancriformis* candidate miRNAs were predicted with miRDeep2 (Friedlander et al. 2012) using the sequence reads that remained after the removal of reads that match the *T. cancriformis* cDNAs. To improve the performance of miRDeep2, mature and precursor *T. cancriformis* candidate miRNAs and previously reported miRNAs were used as the positive input dataset. After prediction, I used three thresholds to detect the reliable novel candidate miRNAs: 1. the lowest miRDeep2 score cutoff (5.0) that yielded the highest signal-to-noise ratio (5.3); 2. a significant Randfold p value (equal to or lower than 0.05) for the potential miRNA precursor; and 3. novel nonredundant miRNA precursor candidates were permitted no mismatches between the small RNA reads and genomic DNA contigs. Novel *T. cancriformis* candidate miRNAs were designated "tcf-miR-n5XX" with identifying numbers (XX; e.g., tcf-miR-n501).

3.2.5 Expression profiles of *T. cancriformis* candidate miRNAs

To compare the relative levels of small RNA expression in the six different developmental stages, read normalization was performed using two spike RNAs, as

described previously (Wang et al. 2011). The read number obtained from the 2nd instar larvae was defined as the standard by setting it to 20 million, and the read numbers for the other stages were normalized to it. Read normalization was as described previously (Wang et al. 2011). The normalized expression profiles of *T. cancriformis* miRNAs are listed with their total normalized expression. The miRNA profiles were clustered using Cluster 3.0 (Eisen et al. 1998), followed by manual adjustment, and visualized with Java Treeview using a pixel setting value of 2.0 (Saldanha 2004).

3.2.6 Northern blot analysis

Total RNA was extracted from each developmental stage of *T. cancriformis* with TRIzol Reagent (Invitrogen), according to the manufacturer's instructions. The total RNA (0.05–20 μg) was separated on denaturing 12.5% polyacrylamide gel containing 8 M urea, and transferred to Hybond-N+ membrane (GE Healthcare, Piscataway, NJ, USA) by electroblotting. After UV crosslinking, the membrane was prehybridized in DIG Easy Hyb buffer (Roche Diagnostics, Indianapolis, IN, USA) for 30 min at 30 °C or 37 °C, depending on the oligodeoxynucleotide used. In some cases, hybridization buffer made in-house containing 0.6 M sodium citrate, 0.06 M NaCl, Denhardt's solution (1% Ficoll, 1% polyvinylpyrrolidone, and 1% bovine serum albumin), 0.1 mg/ml UltraPure™ Salmon Sperm DNA Solution (Invitrogen), and 0.5% SDS, was used. A biotin-labeled antisense oligodeoxynucleotide was prepared using the Biotin 3′ End DNA Labeling Kit (Pierce Biotechnology, Rockford, IL, USA), and hybridization was performed in the same buffer with the labeled antisense oligodeoxynucleotide

overnight at 30 °C or 37 °C. The membrane was then washed with buffer containing 0.6 M sodium citrate, 0.06 M NaCl, and 0.5% SDS at either 30 °C or 37 °C. The nonisotopic blots were visualized with ECF Substrate (GE Healthcare) and the images were captured with a Molecular Imager FX Pro (Bio-Rad Laboratories, Hercules, CA, USA). The band intensities were also analyzed with Molecular Imager FX Pro. Using the quantified band intensities and read counts from the deep-sequencing data, I calculated Pearson's correlation with the StatPlus:mac LE.2009 software (AnalystSoft, Inc., Alexandria, VA, USA).

The probes used to detect *T. cancriformis* miRNAs were:

tcf-let-7-5p_comp	5′-AACCATACAACCTACTACCTCA-3′
tcf-miR-87_comp	5′-ACGCACCTGAAGCTTTGCTCAA-3′
tcf-miR-125_comp	5′-TCACAAGTTAGGGTCTCAGGGA-3′
tcf-miR-1_comp	5′-CTCCATACTTCTTTACATTCCA-3′
tcf-miR-2b_comp	5′-CTCGTCAAAGCTGGCTGTGATA-3′
tcf-miR-12_comp	5′-CCAGTACCTGATGTAATACTCA-3′
tcf-miR-34_comp	5′-CAACCAGCTAACCACACTGCCA-3′
tcf-miR-133_comp	5′-ACAGCTGGTTGAAGGGGACCAA-3′
tcf-miR-184-3p_comp	5′-GCCCTTATCAGTTCTCCGTCCA-3′
tcf-miR-276-3p_comp	5′-AGAGCACGGTATGAAGTTCCTA-3′
tcf-miR-279a_comp	5′-TGGATGAGTGTGGATCTAGTCA-3′
tcf-miR-375_comp	5′-TAACTCGAGCCGAACGAACAAA-3′

tcf-miR-750_comp	5′-TGAGCTGGAAGAGATAGATCTGG-3′
tcf-miR-n501_comp	5′-AGCTGTCAATCATATAACCAAGT-3′
tcf-miR-n502_comp	5′-CAGGATGAACCGCACCCAGTGA-3′
tcf-miR-n503_comp	5′-TCGCCTCGAACCATACAGTGCAA-3′
tcf-miR-n504_comp	5′-AAGCCCACTACCGGTTAGTGCAA-3′
tcf-miR-n505_comp	5′-ACGCACCTGATGATTTGCTCAC-3′
tcf_5.8S_comp	5′-CAGCGTTCTTCATCGATCCACGAGCCGAGTGATCC-3′

3.2.7 Gene prediction and miRNA target prediction

I first searched for genes orthologous to *grim*, *cos*, and *Eip74EF* in *T. cancriformis* in the *T. cancriformis* genomic DNA contigs using TBLASTN with the default parameters (Camacho et al. 2009). The amino acid sequences of grim (NP_524137.2), cos (NP_477092.1), and Eip74EF (NP_001014590) were used as the query sequences. Because contig sequences were identified that were partly similar to the deduced *D. melanogaster* grim, cos and Eip74EF proteins, the exon regions in the *T. cancriformis* contigs were predicted with GENSCAN 1.0 (Burge and Karlin 1997). Because the 3′-UTR sequences of the *cos* and *Eip74EF* genes were inadequately predicted, the downstream sequences of their coding regions (1,000 bp for *cos* and 3,000 bp for *Eip74EF*) were also identified according to the 3′-UTR lengths of *cos* and *Eip74EF* in *D. melanogaster*. miRNA target prediction was performed with two software programs, miRanda v3.3a and RNAhybrid version 2.1 (John et al. 2004; Rehmsmeier et al. 2004) on the putative 3′-UTR sequences that corresponded to the three *T. cancriformis* genes

(*grim, cos*, and *Eip74EF*). For the miRanda predictions, the threshold was set so that the miRanda score was ≥100 and the free energy was ≤–15 kcal/mol, with complete seed matching. For the RNAhybrid predictions, the threshold was set so that the free energy was ≤–15 kcal/mol, with complete seed matching. I deemed a gene to be an miRNA target when both software programs predicted it.

3.2.8 Evolutionary conservation of *T. cancriformis* miRNAs

To identify the conserved *T. cancriformis* miRNAs, I first collected the miRNA sequences from miRBase release 20.0 for 12 model species: *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Apis mellifera*, *Bombyx mori*, *D. melanogaster*, *Daphnia pulex*, *C. elegans*, *N. vectensis*, and *Amphimedon queenslandica*. The BLASTN program was used to compare the distributions of the miRNAs across the species. The conservation criteria were defined as ≧80% sequence identity shared across the miRNAs of all 12 species, with complete seed matching.

For novel *T. cancriformis* candidate miRNAs, I examined the conservation of novel *T. cancriformis* candidate miRNAs in model species. The *T. cancriformis* miRNA sequences of novel candidates were compared with the genomic DNA sequences of model species (*D. melanogaste*r, *Daphnia pulex*, and *C. elegans*) using BLASTN. The *D. melanogaster* and *C. elegans* genomes were downloaded from the University of California at Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/), and the *Daphnia pulex* genome was obtained from Joint Genome Institute (JGI) (http://www.jgi.doe.gov/). If small RNAs in related species shared ≧80% sequence

similarity with the novel miRNA sequences of *T. cancriformis* with complete seed matching, were predicted to form the appropriate secondary structure, and met the reliability criteria (2), (4), and (5), these small RNA sequences were considered potential candidate miRNAs in the corresponding species (see Results and Discussion).

3.2.9 Construction of a phylogenetic tree

A maximum likelihood tree of 13 animal species, including *T. cancriformis*, was constructed based on 18S rRNA sequences (including partial 18S rRNA sequences) obtained from NCBI. Seaview version 4.4.0 (Gouy et al. 2010) was used to align the sequences and construct the phylogenetic tree. The multiple alignment of these 18S rRNAs was generated with MUSCLE version 3.8.31 (Edgar 2004) and the phylogenetic analysis was performed with phyML 3.0 (Guindon et al. 2010) with the GTR model. Support values were calculated with 1,000 bootstrap replications. The amino acid sequences of the DICER proteins and AGO family proteins (including the predicted DICER and AGO family proteins) were obtained from either previous studies (Grimson et al. 2008; Schurko et al. 2009; Mukherjee et al. 2013), JGI, or UniProt (http://www.uniprot.org, October 2013), and multiple alignments of these amino acid sequences were generated with MUSCLE. The aligned sequences were edited manually and the gaps were trimmed. The phylogenetic trees were constructed using the maximum likelihood method with phyML 3.0 with the LG model for the DICER proteins, and with the distance neighbor-joining method for the AGO family proteins. Support values were calculated with 1,000 bootstrap replications.

3.2.10 Prediction of the components of the RNAi machinery

The *T. cancriformis* genomic DNA contigs were first searched for sequences encoding the RNAi machinery components (DICER and AGO family proteins) with TBLASTN using the default parameters. As the query sequences, I used the amino acid sequences of the DICER and AGO family proteins and their functional domains from *H. sapiens*, *D. melanogaster*, *M. japonicus,* and *Daphnia pulex*. For *H. sapiens*, *D. melanogaster*, and *M. japonicus*, these query sequences were obtained from UniProt (*H. sapiens* DICER1, Q9UPY3; AGO1, Q9UL18; AGO2, Q9UKV8; AGO3, Q9H9G7; AGO4, Q9HCK5; PIWIL1, Q96J94; PIWIL2, Q8TC59; PIWIL3, Q7Z3Z3; PIWIL4, Q7Z3Z4; for *D. melanogaster* DCR1, Q9VCU9; DCR2, A1ZAW0; AGO1, Q27IR0; AGO2, Q9VUQ5; AGO3, Q7PLK0; PIWI, Q9VKM1; AUB, O76922; and for *M. japonicus* DICER1, D2XYX5; DCR2, H6WZT1; AGO1, J7I7H1; AGO2, J7MCI3). For *Daphnia pulex*, the amino acid sequences of these proteins were obtained from previous studies (Schurko et al. 2009; Mukherjee et al. 2013) and JGI (DICER1, EFX72380; DICER B, EFX69538; DICER C, EFX86072; AGO1, 305022; AGO2, 311791; AGO3, 442510; AUB A, 239845; AUB B, 220987; AUB C, 308681; AUB F, 195225). The domain sequences of these proteins were predicted with SMART version 7.0 (Letunic et al. 2012). When DNA contigs that were partly similar to the deduced DICER or AGO proteins were identified, the exonic regions were predicted with AUGUSTUS ver. 2.7 (Stanke et al. 2006) and GENSCAN 1.0 (Burge and Karlin 1997). The domain information for each protein was checked with SMART version 7.0. I chose the candidates that contained PAZ, RNase III 1, and RNase III 2 domains for the DICER

proteins, and the PAZ and Piwi domains for the AGO family proteins. In the DICER2 search, I looked for further homologous sequences. A TBLASTN search was performed against the *T. cancriformis* genomic DNA contigs with the default parameters. As query sequences, the amino acid sequences of the proteins and the functional domains of DICER2 were used in four species: *B. mori* (D7UT11)*, Schmidtea mediterranea* (XP_002574802), *Litopenaeus vannamei* (F5AW47), and *Tribolium castaneum* (NP_001107840). The amino acid sequences of these proteins were obtained from a previous study (Mukherjee et al. 2013) or UniProt, and their domain sequences were predicted with SMART version 7.0. The amino acid sequences of the proteins and the functional domains of *Daphnia pulex* DICER B and DICER C*,* and tcf DICER1 were also used to extract any other DICER proteins containing their functional domains.

3.3 Results and discussion

3.3.1 Identification of 87 evolutionarily conserved miRNAs and 93 novel candidate miRNAs in *T. cancriformis*

To clarify the relationships between miRNA expression and the dramatic morphological changes that occur in *T. cancriformis* (Figure 3.1), I performed a deep-sequencing analysis of small RNA libraries from the six different developmental stages (egg, 1st, 2nd, 3rd, and 4th instar, and adult) of *T. cancriformis*. Illumina deep-sequencing yielded 151 million small RNA reads, predominantly 18–35 nt in length, from the six developmental stages. Reads with identical sequences were collapsed into a unique read to calculate the variation in the unique reads per stage, but information on the read numbers of the identical sequences was retained. The numbers of unique reads from which conserved miRNAs were extracted are summarized in Table 3.1. First, small RNA reads of poor quality or with <5 counts were discarded. Because the genomic DNA sequence was required to reliably identify both the miRNAs and the components of the RNAi machinery, deep-sequencing of the genomic DNA was performed. I obtained 133 million genomic DNA reads (100 nt) in total, and discarded the low-quality reads. Reads with identical sequences were collapsed into a unique read to reduce the repeated sequences, and the number of unique reads was 86,279,282. These genomic DNA sequences were assembled with Velvet v. 1.2.10 (Zerbino and Birney 2008), generating 60,629 contigs with N50 of 12,784 bp, a largest contig of 133,027 bp, and approximately 109 Mb of assembled sequence. I then discarded the small RNA reads

Table 3.1 Summary of the small RNA reads in this study

	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
Illumina sequencing reads	5,633,241	3,527,456	4,041,614	3,857,810	3,695,978	3,943,449
Reads counts of ≥ 5	579,881	447,023	536,984	455,710	462,298	476,686
Reads mapped to *Triops* genomic sequences	388,360	405,810	484,109	412,425	414,983	432,126
miRNA fraction reads (18-24 nt)	72,071	66,949	79,489	62,687	64,709	52,270
Reads that did not match known *Triops* cDNA sequences	46,989	58,533	69,924	57,811	55,819	40,301
Conserved miRNAs	68	85	86	86	87	83
Novel candidate miRNAs	65	74	83	81	82	71
Unannotated small RNA reads	46,856	58,374	69,755	57,644	55,650	40,147

that did not match the *T. cancriformis* genomic contigs. In this step, the lengths of these small RNA had a trimodal distribution, with distinct peaks at 22 nt, at approximately 27 nt, and at 32 nt (Figure 3.2A). Based on previous reports (Kim et al. 2009; Wei et al. 2012), I presumed that the peak at 22 nt corresponded predominantly to miRNAs, whereas the peak at approximately 27 nt primarily corresponded to piRNAs in *T. cancriformis*. The third peak at 32 nt consisted of many transfer RNA (tRNA) fragments (to be published separately). In the egg stage, the number of reads corresponding to small RNAs of 18–20 nt was larger (approximately 3% of small RNAs matched the *T. cancriformis* genomic contigs) than the numbers in the other stages (<1%) (Figure 3.2A). More than 10% of these reads in the egg were identical to the 3′ end of *T. cancriformis* 28S rRNA, suggesting that 28S rRNA was specifically degraded to small fractions in the egg stage. Small RNA reads from the 18–24 nt miRNA fraction were then extracted, and those that matched known *T. cancriformis* cDNA sequences were discarded. To extract candidate miRNAs, I predicted the *T. cancriformis* miRNA candidates would share high sequence similarity with those of other species, and I selected major sequences as conserved candidate miRNAs from several miRNA isoforms. After extracting the genomic DNA sequences corresponding to conserved candidate miRNAs, I examined whether these genomic DNA sequences folded into secondary structures commensurate with precursor miRNAs. In this way, I identified 87 conserved mature miRNAs and 71 putative miRNA precursors (Tables 3.2–3.3). Among the conserved miRNAs, two different putative precursor sequences of the mature tcf-miR-2a miRNA were found, and I designated them tcf-miR-2a and

Table 3.2 Nucleotide sequences of conserved miRNAs in *Triops cancriformis* and their read numbers for each developmental stage

miRNA name	miRNA sequence	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult	Average read	[a]Reliability level
tcf-bantam	UGAGAUCAUUGUGAAAGCUGAUU	40083	91300	122780	297010	357576	331433	206697.0	4
tcf-let-7-3p	CUGUACAACUUGCUAACUUUCC	0	22	54	173	151	649	174.8	4
tcf-let-7-5p	UGAGGUAGUAGGUUGUAUGGUU	23	174	687	2627	4568	29195	6212.3	4
tcf-miR-1	UGGAAUGUAAAGAAGUAUGGAG	238	3758	6365	12373	15609	17187	9255.0	4
tcf-miR-2a-5p	CUCACAAAGUGGCUGUCAUGUG	5	265	447	718	887	110	405.3	4
tcf-miR-2a-3p	UAUCACAGCCAGCUUUGAUGAG	431	1471	2672	5055	5483	864	2662.7	4
tcf-miR-2a-2	UAUCACAGCCAGCUUUGAUGAG	431	1471	2672	5055	5483	864	2662.7	4
tcf-miR-2b	UAUCACAGCCAGCUUUGACGAG	1456	3382	6701	13389	14353	2720	7000.2	3
tcf-miR-7	UGGAAGACUAGUGAUUUUGUUGUU	36	121	260	668	965	225	379.2	3
tcf-miR-8-5p	CAUCUUACCGGGCAGCAUUAGA	254	1209	2957	3799	4084	6348	3108.5	4
tcf-miR-8-3p	UAAUACUGUCAGGUAAAGAUGUC	9057	15144	36306	105985	117144	244410	88007.7	4
tcf-miR-9a-5p	UCUUUGGUUAUCUAGCUGUAUGA	3833	5753	10456	19403	19433	19321	13033.2	4
tcf-miR-9a-3p	AUAAAGCUAGGUUACCAAAGUU	20	18	44	165	79	127	75.5	4
tcf-miR-9b-5p	UCUUUGGUGGUCUAGCUGUAUGA	2788	880	1291	843	394	220	1069.3	4
tcf-miR-9b-3p	AUAAAGCUAGAUCAGCAAGGC	1578	2687	1629	1978	1533	125	1588.3	4
tcf-miR-10-5p	UACCCUGUAGAUCCGAAUUUUGU	13950	147193	584564	499716	1127757	379788	458828.0	4
tcf-miR-10-3p	CAAAUUCGGUUCUAGAGAGGUUC	206	1763	3310	4235	4826	4303	3107.2	3
tcf-miR-12	UGAGUAUUACAUCAGGUACUGGU	11076	66920	53056	141652	85565	104791	77176.7	4
tcf-miR-13	UAUCACAGCCAUUCUUGAUGAG	342	1417	2516	4388	5270	435	2394.7	4
tcf-miR-31	AGGCAAGAUGUCGGCAUAGCU	153	2625	5047	8976	15182	32874	10809.5	4
tcf-miR-33	GUGCAUUGUAGUUGCAUUGCA	8	23	52	69	162	86	66.7	4
tcf-miR-34	UGGCAGUGUGGUUAGCUGGUUG	22323	75	63	64	88	389	3833.7	4
tcf-miR-61	UGACUAGAUCCAUACUCACCAG	5578	52820	94316	121020	140716	31063	74252.2	4
tcf-miR-71-5p	UGAAAGACAUGGGUAGUGAGAU	340	1719	3143	4203	5661	1051	2686.2	4
tcf-miR-71-3p	UCUCACUAUCUUGUCGUUCAUG	1129	3416	4578	9529	10295	2344	5215.2	4
tcf-miR-87	UUGAGCAAAGCUUCAGGUGCGU	1787	8427	11635	23019	25062	37880	17968.3	4
tcf-miR-92a	UAUUGCACUCGUCCCGGCCUGU	11090	119532	112208	98046	74381	23387	73107.3	4
tcf-miR-92b	AAUUGCACUUGUCCCGGCCUGC	6898	74004	75245	65076	55903	19824	49491.7	4
tcf-miR-96	UUUGGCACUAGCACAUUUUUGU	140	329	1115	1409	3758	1216	1327.8	4
tcf-miR-100	AACCCGUAGAUCCGAACUUGUG	14	800	4197	6453	15434	21586	8080.7	4
tcf-miR-124-5p	CGUGUUCACUGUUGGCCUUGAUG	0	334	811	1507	1770	29	741.8	3
tcf-miR-124-3p	UAAGGCACGCGGUGAAUGCCAA	0	37	77	219	366	19	119.7	3
tcf-miR-125	UCCCUGAGACCCUAACUUGUGA	27	440	1300	2626	5340	10645	3396.3	4
tcf-miR-133	UUGGUCCCCUUCAACCAGCUGU	13	1538	2612	6232	6842	4411	3608.0	4
tcf-miR-137	UUAUUGCUUGAGAAUACACGU	0	15	46	100	177	91	71.5	4
tcf-miR-153-5p	UCAUUUUUGUGAUUUUUGCAAUU	0	0	0	0	8	11	3.2	4
tcf-miR-153-3p	UUGCAUAGUCACAAAAGUGAUG	29	87	239	641	1470	4698	1194.0	4
tcf-miR-184-5p	CCUUAUCAUUCACCAGUCCGG	0	10	13	26	31	17	16.2	4
tcf-miR-184-3p	UGGACGGAGAACUGAUAAGGGC	125505	780841	746145	838107	975769	838326	717448.8	4
tcf-miR-190	AGAUAUGUUUGAUAUUCUUGGUUG	122	176	297	898	1056	1300	641.5	4
tcf-miR-193	UACUGGCCUGCUAAGUCCCAAG	0	226	864	2754	3720	2760	1720.7	4
tcf-miR-210-5p	AGCUGCUGGACACUGCUCAAGAU	0	38	48	73	112	190	76.8	4
tcf-miR-210-3p	CUUGUGCGUGUGACAGCGGCU	18	20	70	149	363	671	215.2	4
tcf-miR-219	UGAUUGUCCAAACGCAAUUCUUG	0	50	53	93	79	0	45.8	4
tcf-miR-252a-5p	UAAGUACUAGUGCCGCAGGAG	23	103	245	766	945	3662	957.3	4
tcf-miR-252a-3p	UCCUGCAGCUAUAGUGCUUACC	0	5	13	29	22	44	18.8	4
tcf-miR-252b	CUAAGUAGUUGUGCCGCAGGUAA	0	110	309	677	774	1049	486.5	4
tcf-miR-263a	AAUGGCACUGGAAGAAUUCACGGG	3021	7125	35687	27734	65163	16552	25880.3	4
tcf-miR-263b	CUUGGCACUGGAAGAAUUCACAGA	186	554	2538	2841	6824	2255	2533.0	4
tcf-miR-275	UCAGGUACCUGAAGUAGCGCGC	8559	28360	46715	45526	32406	12538	29017.3	4
tcf-miR-276-5p	AGCGAGGUAUAGAGUUCCUACG	0	10	33	78	125	49	49.2	4
tcf-miR-276-3p	UAGGAACUUCAUACCGUGCUCU	2225	20586	67272	219025	301011	275009	147521.3	4
tcf-miR-277	UAAAUGCAUUUUCUGGUAUGUC	1155	231	516	892	1356	1812	993.7	4
tcf-miR-278	UCGGUGGGAUUUUCGUCCGUU	0	33	65	131	160	166	92.5	4
tcf-miR-279a	UGACUAGAUCCACACUCAUCCA	5306	29898	54580	90741	88716	20710	48325.2	4
tcf-miR-279b	UGACUAGAUCCAUACUCAUCU	5604	9781	21739	36678	51644	46649	28682.5	4
tcf-miR-279c	UGACUAGAUCCCACACUCGUCCGG	4836	886	713	404	188	146	1195.5	4
tcf-miR-279d	UGACUAGAUUCACACUCAUCCA	123	236	445	1283	852	628	594.5	4
tcf-miR-279e	UGACUAGAUCCUACACUCGUCC	118	99	101	105	80	0	83.8	4
tcf-miR-281-5p	AAGAGAGCUAUCCGUCGACAGU	3854	43816	52395	85978	124278	54417	60789.7	4
tcf-miR-281-3p	UGUCAUGGAGCUGCUCUCUUUA	78	512	781	2022	2048	780	1036.8	4
tcf-miR-282	UAGCCUCUCCUAGGCUUUGUCU	6	17	31	65	152	173	74.0	3
tcf-miR-283	AAAUAUCAGCAGGUAAUUCUGGGC	461	1849	4438	10030	9893	6060	5455.2	4
tcf-miR-285	UAGCACCAUUGGAAUUCAGUUUA	17	65	186	279	677	185	234.8	4
tcf-miR-305	AUUGUACUUCAUCAGGUGCUCUGG	522	1743	2053	3881	1795	1312	1884.3	4
tcf-miR-307	UCACAACCUCCUUGAGUGAGUG	29	2117	3427	3742	3555	2061	2488.5	4
tcf-miR-315-5p	UUUUGAUUGUUGCUCAGAAAGCC	325	18356	42306	113690	159939	77117	68622.2	4
tcf-miR-315-3p	CUUUCGAGUAACAAUCAGAGUC	0	14	37	46	38	0	22.5	4
tcf-miR-317	UGAACACAGCUGGUGGUAUCU	18	564	1631	1620	3326	5022	2030.2	3
tcf-miR-375	UUUGUUCGUUCGGCUCGAGUUA	115	3366	6503	10298	10372	20126	8463.3	4
tcf-miR-745	GAGCUGCCCAGUGAAGGGCUUUC	0	26	30	67	74	228	70.8	4
tcf-miR-750	CCAGAUCUAUCUCUUCCAGCUCA	472	6053	10705	16024	14614	13494	10227.0	4
tcf-miR-965	UAAGCGUAUGGCUUUUCCCCUG	561	718	1344	2240	2197	250	1218.3	4
tcf-miR-981	UUCGUUGUCGACGAAACCUGCAC	112	662	1076	2805	3576	2337	1761.3	4
tcf-miR-993-5p	CUACCCUGUGGAUCCGGGCUUUU	174	1048	887	1633	1975	2047	1294.0	3
tcf-miR-993-3p	GAAGCUCGUCUCUACAGGUAUCU	617	4059	6597	11014	12086	11759	7688.7	3
tcf-miR-995	UAGCACCACAGGAUUCAGCUU	13	67	179	109	175	53	99.3	4
tcf-miR-996	UGACUAGAGUUACACUCAUCU	161	115	280	463	576	1595	531.7	4
tcf-miR-998	UAGCACCACGGGAUUCAGCCGC	131	22	74	25	42	29	53.8	3
tcf-miR-1175-5p	AAGUGGAGCAGUAGUCUCGUCACU	89	15454	26966	20523	23449	12212	16448.8	4
tcf-miR-1175-3p	UGAGAUUCAACUCCUCCAACUUUG	99	270	531	1028	958	2539	904.2	4

Table 3.2 (Continued)

miRNA name	miRNA sequence	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult	Average read	[a]Reliability level
tcf-miR-2788	CAAUGCCCUGAGAAAUCCCAGA	0	50	729	507	2732	920	823.0	4
tcf-miR-2944	UAUCACAGCCGUAGUUACCUAGA	10898	6694	11530	8236	6174	962	7415.7	4
tcf-miR-3477	UAAUCUCAUGCGGUAACUCUGAGA	589	15297	22305	23198	24658	24459	18417.7	4
tcf-miR-3791	UCACCGGGUAGAAUUCAUCCAG	1825700	35480	70510	41510	48868	62798	347477.7	4
tcf-miR-5608	UUUUAUCCGACCGUGCGUACUGUA	0	7	15	15	43	0	13.3	3
tcf-miR-iab-4	ACGUAUACUGAAUGUAUCCUGA	0	371	607	824	880	247	488.2	3
tcf-miR-iab-8	UACGUAUACUGAAGGUAUACCGGA	0	0	5	47	68	68	31.3	4

[a]Reliability level; see the main text in details.

Table 3.3 Nucleotide sequences of putative conserved miRNA precursors in *T. cancriformis*

miRNA name	precursor sequence
tcf-bantam	CUAGUUUUCUCAGUGAUCUGCCAGAUAUUGUUAAAAUUUCUGAGAUCAUUGUGAAAGCUGAUU
tcf-let-7	UGAGGUAGUAGGUUGUAUGGUUCAGAAUUACACCGUUCAGGCGAAACUGUACAACUUGCUAACUUUCC
tcf-miR-1	CCAUGCUUCCUUACUUCCCAUAGCGAUAUGACAUAUGGAAUGUAAAGAAGUAUGGAG
tcf-miR-2a	CUCACAAAGUGGCUGUCAUGUGUAGGUGAUUCACGUUGGAUUGCAUAUCACAGCCAGCUUUGAUGAG
tcf-miR-2a-2	CUCGUCAGAGUGGUAGUGAUGUGUAGACGUAACUUCAUAUCACAGCCAGCUUUGAUGAG
tcf-miR-2b	AGUCGAAGCGGGUUGGGAAAUGGUGUGACGAAAACGCAUAUCACAGCCAGCUUUGACGAG
tcf-miR-7	UGGAAGACUAGUGAUUUUGUUGUUCGAUAUAUGGGGAAACAAAAAGUCACUAGUUAUCCUA
tcf-miR-8	CAUCUUACCGGGCAGCAUUAGAUGUACAAAAAAACUUCUAAUACUGUCAGGUAAAGAUGUC
tcf-miR-9a	UCUUUGGUUAUCUAGCUGUAUGAGUGAAACAGACGUCAUAAAGCUAGGUUACCAAAGUU
tcf-miR-9b	UCUUUGGUGGUCUAGCUGUAUGACUGUGUUCGACUCAUAAAGCUAGAUCAGCAAGGC
tcf-miR-10	UACCCUGUAGAUCCGAAUUUGUAUUUCCCACAAUGACAAAUUCGGUUCUAGAGAGGUUC
tcf-miR-12	UGAGUAUUACAUCAGGUACUGGUUGAACUUGGGACACCAGUACUUCUGUAACACUCUCC
tcf-miR-13	CGUCGGAAUGGCUGCGGUGUGUUGCGUUGGUUCAUAUCACAGCCAUUCUUGAUGAG
tcf-miR-31	AGGCAAGAUGUCGGCAUAGCUGAGAUAUUUUCACUUUUUGGUAUAGAAAAGUAAGCUGUGUUUACAUCGUGCCAUC
tcf-miR-33	GUGCAUUGUAGUUGCAUUGCAUGGUCGGUUUAAGAUGUGCAGUGUUUCUGCAGUGCAAACC
tcf-miR-34	UGGCAGUGUGGUUAGCUGGUUGUGUGGAUUUCAAGUUUUCACAGCCACUAUCCAUACUGCGCCC
tcf-miR-61	GGUGGGCAUGGCUCUGACGUCAUGGAGAAUUGAACGCGUGACUAGAUCCAUACUCACCAG
tcf-miR-71	UGAAAGACAUGGGUAGUGAGAUGUUGGCAUAUCGUACCUCUCACUAUCUUGUCGUUCAUG
tcf-miR-87	GGCCUGAGGCUCCUUUGCUCAACAGGGUCUUGUUGAUGUUGAGCAAAGCUUCAGGUGCGU
tcf-miR-92a	AGACUGCGACUUGGGCAAUGUUCUGUCGUCUUAAGAGUAAUAUUGCACUCGUCCCGGCCUGU
tcf-miR-92b	AGGCUGGUUCAGGGGCAAAUUGUUGGAUAACGUGAGUCAAAUUGCACUUGUCCCGGCCUGC
tcf-miR-96	UUUGGCACUAGCACAUUUUUGUGUUGAGACUGUUAACAAAACUGUGAUAGUGUCAAGU
tcf-miR-100	AACCCGUAGAUCCGAACUUGUGUGGCAACCUCGCAAUUUCUUUCGCAAGUUCGAUUCUAAGGGCGCA
tcf-miR-124	CGUGUUCACUGUUGGCCUUGAUGUGGAUAUUGUCAUAAGGCACGCGGUGAAUGCCAA
tcf-miR-125	UCCCUGAGACCCUAACUUGUGAUAGCAAUCUCACAGGCUAGCUUCUCAGGUAU
tcf-miR-133	AGCUGGUUGAAUCCGGGCCAAAUUGUUAUUCAUAGCGGAGCAUUUGGUCCCCUUCAACCAGCUGU
tcf-miR-137	GCGUAUUCUUAAGUGAUUAGCACGCAUUUUAAGUUGUUAUUGCUUGAGAAUACACGU
tcf-miR-153	UCAUUUUUGUGAUUUUUGCAAUUGUAUUAAAGAUUUUAAUUGCAUAGUCACAAAAGUGAUG
tcf-miR-184	CCUUAUCAUUCACCAGUCCGGUUGGAAUUCUGUAGACUGGACGGAGAACUGAUAAGGGC
tcf-miR-190	AGAUAUGUUUGAUAUUCUUGGUUGUGGCUUUCCAACUCAACCAGAUAUCAGACAUAUUAUUA
tcf-miR-193	AGGGACUUGGCGGAACUGUGGGUUGGGACUUGGACCUCUUAAGGAAGCCUACUGGCCUGCUAAGUCCCAAG
tcf-miR-210	AGCUGCUGGACACUGCUCAAGAUUAGAAUGGUACACAACUCUUGUGCGUGUGACAGCGGCU
tcf-miR-219	UGAUUGUCCAAACGCAAUUCUUGUAACAUAUUAUUCAACAAGAAGUGUGUGGGGACAUCAUU
tcf-miR-252a	UAAGUACUAGUGCCGCAGGAGAGGGAAUUCGAGUCCUCCUGCAGCUAUAGUGCUUACC
tcf-miR-252b	CUAAGUAGUUGUGCCGCAGGUAACCGAUCCGAGUGCUUACCUGCACACGCCUGCUUACAUC
tcf-miR-263a	AAUGGCACUGGAAGAAUUCACGGGAAGAGGCACAUUGAGACGGCCCCCGUGGCUCUCCAGUGCUGUACC
tcf-miR-263b	CUUGGCACUGGAAGAAUUCACAGAGUAUUUGAAACCAUCGUGGAUUCCCUGGUGCCAUAGA
tcf-miR-275	CUGUGCUACGGCAGGUGCUUGCGUCUGAGUCAGAUCAGUCAGGUACCUGAAGUAGCGCGC
tcf-miR-276	AGCGAGGUAUAGAGUUCCUACGUCAUUUUAUGCGGUAGGAACUUCAUACCGUGCUCU
tcf-miR-277	CGUACCAGGAAUGUGUUUACAUUAUGUUUCGUAAAAGAUGUAAAUGCAUUUUCUGGUAUGUC
tcf-miR-278	CCGGACGAAAAUCUUCUUGCCAGACCUUAAUUCCAAAUCAUGUCGUAUGAUGUAUUGGUCGGUGGGAUUUUCGUCCGUU
tcf-miR-279a	GGUGAAUGUGUAUCUGGUGCAUGUGUAUCAAUUUCCAUGACUAGAUCCACACUCAUCCA
tcf-miR-279b	GGUGGGUAUGUUUCUAGUGCAUGCAUGAGGUUGUUUAUAAGAACCAUGACUAGAUCCAUACUCAUCU
tcf-miR-279c	GGCGAGUGGGCAUCUGGUACAUGUGUUUUCACUAAGUUGGUCAUGACUAGAUCCCACACUCGUCCGG
tcf-miR-279d	GAUGAGUGCGUUUCUGGUGCAUAGUACUGUAACAUAUGACUAGAUUCACACUCAUCCA
tcf-miR-279e	UGGCGAGUGGGUUUGUAGUCCACGUGCAAUAGUUUCCGUGACUAGAUCCUACACUCGUCC
tcf-miR-281	AAGAGAGCUAUCCGUCGACAGUCAAGAAUUCACGAUACUGUCAUGGAGCUGCUCUCUUUA
tcf-miR-282	UAGCCUCUCCUAGGCUUUGUCUGUCUGUUGUGACCGGAGACACUGCCUUGAGGAGGCCAUCC
tcf-miR-283	AAAUAUCAGCAGGUAAUUCUGGGCUGUUCCCAUCCCAGACUACCCGUUGAUAUGCAA
tcf-miR-285	ACUGAAUUCUUUUGAUGCGUAGAUUGACUCCGUUUAGUAUUCUAGCACCAUUGGAAUUCAGUUUA
tcf-miR-305	AUUGUACUUCAUCAGGUGCUCUGGUACAGCUUCACCCGGCACCUGUUGGAGUGCAAUUU
tcf-miR-307	UCACUCAACUUGGGUGUGGCGCGUGAUUCAAAAAAAUGUCGUCACAACCUCCUUGAGUGAGUG
tcf-miR-315	UUUUGAUUGUUGCUCAGAAAGCCGUGCGACAAGUGAGGCUUUCGAGUAACAAUCAGAGUC
tcf-miR-317	CGGGAGCCACCCUGAGUUCACUUGGACUUGAAAAGUGAACACAGCUGGUGGUAUCU
tcf-miR-375	ACUCGAGCCCUUCGUAGCACAUCCCAUAUACGUUGGAAUUGAUUUGUUCGUUCGGCUCGAGUUA
tcf-miR-745	CAGUCCUUUCCCGGGCACUAGCUUCUGGUAUCGAAAGAGCUGCCCAGUGAAGGGCUUUC
tcf-miR-750	AGUUGGAAGUGGGGAUCUAGGCACCUGCAACUUGCUCACAAUUUCGCAAAGCUUUCCUUCAACUCGUUGCUGAAAGAGC UUUGCUGAAAUCGUGCCAGAUCUAUCUCUUCCAGCUCA
tcf-miR-965	AGGGAAAAGCUGUGACGAUUGUGCGAAUCUAAGAAUUGCAUAAGCGUAUGGCUUUUCCCCUG
tcf-miR-981	UCGGGUUUCGCGAUAUUCGAACUGUGCUUCCCGGAGCCUAUUGCUACCGAGUUCGUUGUCGACGAAACCUGCAC
tcf-miR-993	CUACCCUGUGGAUCCGGGCUUUUGUAGAAUUUAGCCAUCAGAAGCUCGUCUCUACAGGUAUCU
tcf-miR-995	ACUGAAUCUUGCGGUGGGUACCAGCUUCGCACGGCCGUACUGUACUGUACAUCAAGCUUGCUUUCCGACUCAAGGUAGC ACCACAGGAUUCAGCUU
tcf-miR-996	GAUGGGUGUUGCUUUGGUGCAUGUGAUACUAUACUCUCCAUGACUAGAGUUACACUCAUCU
tcf-miR-998	UGGAUCCAGUGGAUGGGAUGCUGGGAUAUUCACGUUCGAUAAUGCAUAGCACCACGGGAUUCAGCCGC
tcf-miR-1175	AAGUGGAGCAGUAGUCUCGUCACUUGGAUACUUAGCAAGUGAGAUUCAACUCCUCCAACUUUG
tcf-miR-2788	UGGGGGUUCUUGUGGGCAUAUGCCAUGCAAUGGGGCAAUGCCCUGAGAAAUCCCAGA
tcf-miR-2944	AAGGAACUACCGCUGUGGUAUGUGGAGCCUUCCGUAUCACAGCCGUAGUUACCUAGA
tcf-miR-3477	UAAUCUCAUGCGGUAACUCUGAGAGAGUGUAUGUCUCUCGGAGUGUCACGUGAGAUAUG
tcf-miR-3791	CGGUGAAUUUUGCUAUCGGUGAUGUUAACGUCGGUUUUUCAUCACCGGGUAGAAUUCAUCCAG
tcf-miR-5608	UUUUAUCCGACCGUGCGUACUGUAACCGGUAUUUGGAUCUGAUACUUGAACUCGUGGGUAGAGGC
tcf-miR-iab-4	ACGUAUACUGAAUGUAUCCUGAGCAUAUACAUUCCGGUAUACCUUCAGUAUACGUAAC
tcf-miR-iab-8	UACGUAUACUGAAGGUAUACCGGAAUGUAUAUGCUCAGGAUACAUUCAGUAUACGUCC

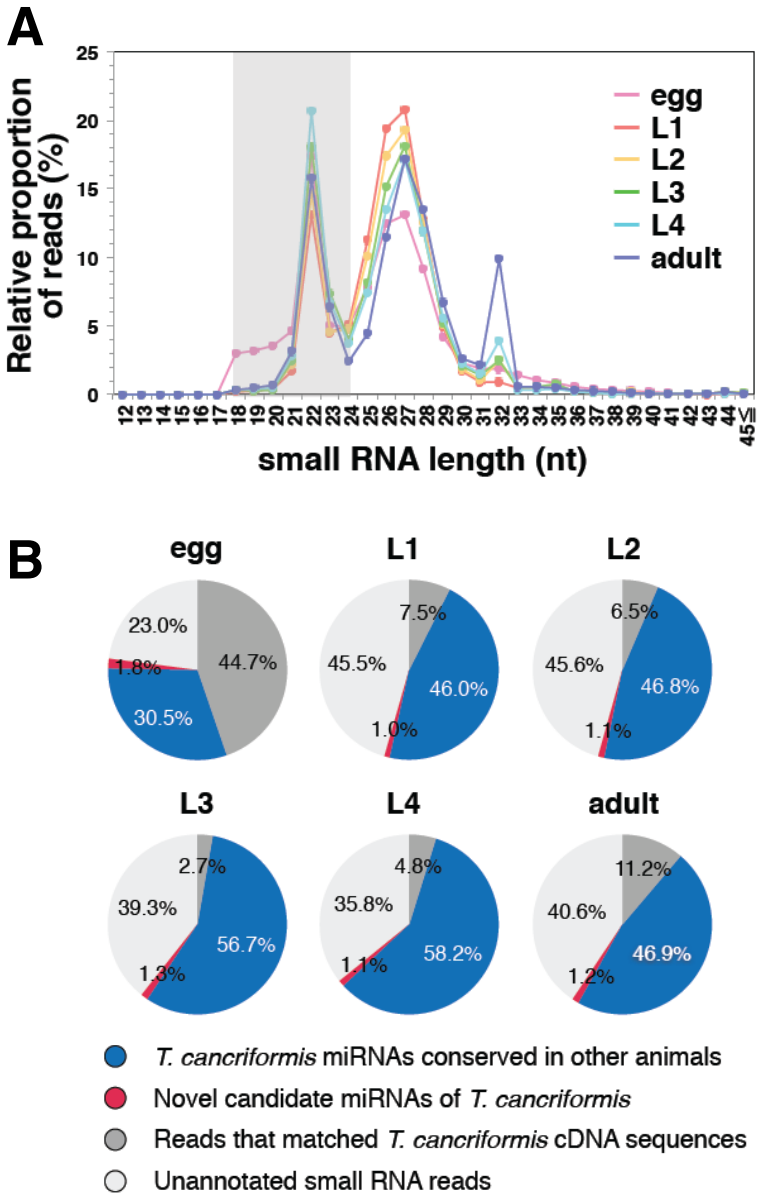


Figure 3.2 Summary of the deep-sequencing analysis in each developmental stage of *T. cancriformis*

(A) Relative numbers of reads based on small RNA lengths. The Y-axis represents the ratio of small RNA reads and 100% represents the total small RNA reads that perfectly matched *T. cancriformis* genome contigs in each stage (see Materials and Methods). The color indicates each stage (pink for egg, red for 1st instar, yellow for 2nd instar, green for 3rd instar, light blue for 4th instar, and purple for adult). The miRNA fraction (18–24 nt) is shaded in gray. (B) Pie charts summarizing the proportions of small RNA annotations in the miRNA fraction in each stage.

tcf-miR-2a-2. To clarify the miRNA sequence quality, I assigned a reliability level to individual *T. cancriformis* miRNAs based on a previous study (Kozomara and Griffiths-Jones 2013) by applying the following criteria: (1) the mature miRNA is detected in ≧10 reads, with no mismatches to the *T. cancriformis* genome contigs; (2) each strand of the miRNA precursor must pair with the 0–4 nt overhang at its 3′ end; (3) at least 50% of reads mapping to each strand of the precursor must have the same 5′ end; (4) the free energy of the predicted RNA secondary structure must be less than 0.2 kcal/mol/nt; and (5) at least 60% of the bases in the mature sequence must be paired in the predicted secondary structure. I categorized the reliability levels in *T. cancriformis* from level 1 to level 4 (i.e., level 4 represents the highest reliability), using the four criteria (2–5) described above, and all miRNAs must meet criterion 1. The miRNAs were then classified according to the number of the criteria they met (e.g., if miR-X met three criteria (3–5), it was categorized at level 3, and if miR-Y met two criteria (4–5), it was categorized at level 2). Of the 88 conserved miRNAs (including miR-2a-2), 76 were categorized into reliability level 4. Furthermore, 84 of 87 were found in the experimental data from the deep-sequencing analysis of the small RNA libraries from the six developmental stages of *T. cancriformis* (data not shown), suggesting strong reproducibility of the conserved miRNA expression. These results show that this method efficiently screened the conserved miRNAs in *T. cancriformis*.

I then predicted the novel candidate miRNAs in *T. cancriformis* using miRDeep2 (Friedlander et al. 2012), inputting the 87 conserved mature *T. cancriformis* miRNAs, 71 putative *T. cancriformis* miRNA precursors, and 30,424 known mature

miRNAs registered in miRBase release 20.0 (Kozomara and Griffiths-Jones 2013). After evaluating the scores obtained for the output of miRDeep2 (see Materials and Methods), I identified 93 novel candidate miRNAs and 98 putative precursor sequences for the novel candidate miRNAs (Tables 3.4–3.5). Sixty-six of the 98 putative precursor sequences of the novel candidate miRNAs were categorized into reliability level 4, suggesting that many reliable novel miRNAs were successfully predicted by miRDeep2.

When I compared the annotated miRNA sequences of 18–24 nt in the six developmental stages (Figure 3.2B), I found that the proportion of conserved miRNAs was higher than the proportion of novel candidate miRNAs in all stages (30.5%–58.2% conserved miRNAs and 1.0%–1.8% novel candidate miRNAs). The average read number for the conserved miRNAs was 30,823.6, whereas that of the novel miRNAs was only 756.0, supporting the previous finding that the number of miRNAs conserved among a wide range of species tends to be high (Watanabe et al. 2008). Because both the variety and proportion of conserved miRNAs increased from the egg to the 1st instar (Figure 3.2B and Table 3.1), I presumed that these conserved miRNAs are related to the morphogenesis of the 1st instar larva. In contrast, the proportion of novel candidate miRNAs decreased from the egg to the 1st instar, although their variety increased. Because four novel candidate miRNAs (tcf-miR-n501, tcf-miR-n502, tcf-miR-n503, and tcf-miR-n504) were expressed in the egg stage with high read counts (≥10,000), these novel candidate miRNAs may be related to the release from dormancy.

Table 3.4 Nucleotide sequences of novel candidate miRNAs in *T. cancriformis* and their read numbers for each developmental stage

miRNA name	miRNA sequence	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult	Average read	[a]Reliability level
tcf-miR-n501	ACUUGGUUAUAUGAUUGACAGCU	15230	23048	40512	54129	60082	33266	37711.2	4
tcf-miR-n502	UCACUGGGUGCGGUUCAUCCUG	34665	1873	2582	1750	1473	1324	7277.8	4
tcf-miR-n503	UUGCACUGUAUGGUUCGAGGCGA	29350	322	372	313	282	9934	6762.2	4
tcf-miR-n504	UUGCACUAACCGGUAGUGGGCUU	34766	918	832	601	343	94	6259.0	4
tcf-miR-n505	GUGAGCAAAUCAUCAGGUGCGU	90	1083	1994	3485	4599	4285	2589.3	4
tcf-miR-n506	UUGGUUUGAGAUGAGUGGACCCG	263	837	1347	1632	2145	5370	1932.3	3
tcf-miR-n507	AUCGGCACUGGGUAACAAUGAA	120	852	1950	1281	2877	3201	1713.5	4
tcf-miR-n508	UGUGAUGUGUUAUUUCAAACUAGU	8	0	0	0	0	9589	1599.5	4
tcf-miR-n509	AAGGAACACUUGCUGUGGUAUG	441	1032	1852	2301	1745	1473	1474.0	4
tcf-miR-n510	UGAGAUCAACCUAUGUAGCAUU	1183	714	1435	743	961	3535	1428.5	4
tcf-miR-n511	UCUUGGUUAUGGUGAAAAUGG	174	1090	1083	709	542	143	623.5	4
tcf-miR-n512	UUGCACUGGCCUGCCCGGGGGU	3639	5	22	10	28	18	620.3	4
tcf-miR-n513	UGUCUUUUCCGCUUGACUGCCG	0	105	276	657	966	1019	503.8	4
tcf-miR-n514	GUAGAAGUUUUCGCCACCCCUGAA	35	315	466	378	754	227	362.5	4
tcf-miR-n515	AUAUAUUGCGAGGUGCAGAACU	118	112	207	370	396	257	243.3	4
tcf-miR-n516	UCGGUAUUUCUAUUUCUGUCCU	33	85	216	328	410	177	208.2	4
tcf-miR-n517	UUGGUCCCAGGACGGUAGAUGACG	148	283	268	223	193	27	190.3	2
tcf-miR-n518	CUAAAGCUAAGCCACCAGAGGG	479	151	166	194	100	27	186.2	4
tcf-miR-n519	UCCAUGAGCUCUCCAACGAACCU	56	353	348	126	169	30	180.3	2
tcf-miR-n520	UCACUGGGUGUGGUUUAUCCG	744	19	45	49	47	0	150.7	4
tcf-miR-n521	UCACUGGGUAAGGUUCGCCCCCG	610	10	12	10	21	88	125.2	4
tcf-miR-n522	UGUGUAGGAUAGGUGGGAGGCU	705	0	0	0	0	0	117.5	2
tcf-miR-n523	UAAGCAGCAUCGGGCUUGGUGACC	69	234	116	88	80	49	106.0	3
tcf-miR-n524	UGCAAGGUCGUUCUGCUUACGGUC	88	145	127	130	91	30	101.8	3
tcf-miR-n525	UUGGUCUGUAACGCUUUUACCACC	51	142	171	116	118	9	101.2	2
tcf-miR-n526	UUCGUGAAUUUUAGCAUAAUGU	0	0	5	17	20	551	98.8	4
tcf-miR-n527	UCACCAGGUGAGAUUCAUCCAU	246	80	100	53	53	10	90.3	4
tcf-miR-n528	UCUGCGACAAUCUUGGCCUCGUCG	28	128	181	63	87	53	90.0	4
tcf-miR-n529	AAUGUCCAACUUUUAGAAACUU	21	29	82	93	119	191	89.2	2
tcf-miR-n530	AAGAAACGAAUCGGCGCUGAGACU	64	109	149	108	85	19	89.0	3
tcf-miR-n531	UCGUAUACGAGACCGGGGUCUC	32	60	80	102	113	26	68.8	4
tcf-miR-n531-2	UCGUAUACGAGACCGGGGUCUC	32	60	80	102	113	26	68.8	4
tcf-miR-n532	CUACAUAGGUUGAUCUCACCG	96	64	75	99	43	27	67.3	4
tcf-miR-n533	CUGGGCAACAAUGAAAAACGCU	11	34	69	61	68	44	47.8	4
tcf-miR-n534	CAUUCUUGUAGCAGUAUCACCU	20	39	45	38	47	21	35.0	4
tcf-miR-n535	UUGCACUGGCCGGUCCAGGGUU	169	5	9	7	7	12	34.8	4
tcf-miR-n536	GCAAGAAACAAGUCGGAAAUCGGA	28	52	36	40	27	24	34.5	4
tcf-miR-n537	UUGCACUGGCUAGCCCAGGGGA	196	0	0	0	0	0	32.7	4
tcf-miR-n538	UUUUUGAAACACAGCGGUACUG	19	37	36	28	32	18	28.3	4
tcf-miR-n539	UCUGAGAACGGUUCGACUCGACU	10	40	36	37	30	5	26.3	3
tcf-miR-n540	UCAACAUUCUGUAACUCAGCCU	6	36	39	17	23	33	25.7	1
tcf-miR-n541	UUUGAAAAACUCAGUGCGCACC	0	0	10	7	17	116	25.0	4
tcf-miR-n542	CUUCCAUAGGUCGUACCUGAAU	5	13	27	31	35	19	21.7	4
tcf-miR-n543	UCUGAUCUGAUCUUUUUUACACC	14	26	23	23	14	11	18.5	4
tcf-miR-n544	UUGGAUAGAGGUAAUCUCGGCCU	16	27	14	8	11	34	18.3	4
tcf-miR-n545	UGUGUCUUGUCGUGCUGCUGCU	0	17	18	36	28	6	17.5	4
tcf-miR-n546	GUACGGUUUUCAGGGAGUCGGCU	30	32	19	13	0	10	17.3	3
tcf-miR-n547	UGCACUUACUGACUUUACUGCCA	0	8	10	26	30	30	17.3	4
tcf-miR-n548	UAUGAACUCGUCCCUUGGUCACGU	6	13	22	28	26	8	17.2	2
tcf-miR-n549	AUGGCGGACAAAGGGUUUGACAUU	6	25	31	19	15	0	16.0	4
tcf-miR-n550	UAAGAACGAGAGCAACUCAGAACC	12	12	25	21	15	11	16.0	3
tcf-miR-n551	UCGACCAGUACUCGUCACGGCACC	5	22	41	20	6	0	15.7	4
tcf-miR-n552	UAUCAUCGUCUUCUAUGGAUGGCU	13	10	28	22	19	0	15.3	2
tcf-miR-n553	UGUAUAGUACUCGGCAUAAAAA	29	6	17	26	14	0	15.3	4
tcf-miR-n554	UUUUGGUGGAUUCGGAUGCAAGC	9	26	26	6	11	12	15.0	3
tcf-miR-n555	UUCUCUCCCGGACUCUUCUUGGUC	10	25	22	13	14	0	14.0	3
tcf-miR-n556	CAAGGAAAAGGUUAUAAUAACC	0	0	5	8	11	53	12.8	3
tcf-miR-n557	CUUUUGAAUUGCGCGGUCUAGC	0	9	6	0	9	50	12.3	4
tcf-miR-n558	UAGGUCUGUCGCAACGGCGAGCC	10	24	15	11	0	11	11.8	4
tcf-miR-n559	UGCGUGACAGAAUUCAUAAGAA	0	6	18	15	32	0	11.8	3
tcf-miR-n560	AGAAGUCUUAGCGAAGGAGAGCUU	9	12	11	24	13	0	11.5	4
tcf-miR-n561	CCCCGUGGAAGUCUGGAUAUUUUA	0	21	21	12	8	7	11.5	3
tcf-miR-n562	UAAAGAUUGACAAGCUGGCGUAGC	15	7	18	7	13	5	10.8	4
tcf-miR-n563	UAAAAACCACAUGUCCGAUUUG	11	13	22	17	0	0	10.5	4
tcf-miR-n564	GAAAUGCUCCCGUCGUUUGGG	0	10	13	11	16	12	10.3	4
tcf-miR-n565	UUGUCGGAGUAGAUACUCAAUGGC	12	17	0	14	12	7	10.3	4
tcf-miR-n566	AACCAGGCUGAGGUCUUGACGAGA	0	0	0	7	36	16	9.8	3
tcf-miR-n567	UAGGAUCGUUUCAAGUGACACGUC	10	15	13	9	7	5	9.8	3
tcf-miR-n568	UAAGCUCUACGUCUGGAGGCAUCC	5	9	17	10	15	0	9.3	3
tcf-miR-n569	AUUGGACCUAGAUCUCGGAAACGC	0	13	20	6	8	7	9.0	2
tcf-miR-n570	UAGGGGAUAGUGCAGAGGUCUA	7	15	8	9	8	7	9.0	4
tcf-miR-n571	UUCAAGGUUUGCGCUCUGCAACU	0	0	10	0	9	35	9.0	4
tcf-miR-n572	UUGGCUCGGAAAUGGGGAUUUUCU	0	12	15	7	9	7	8.3	3
tcf-miR-n573	UGAGUCGACAACGGGAACCGAACC	5	19	10	9	5	0	8.0	2
tcf-miR-n574	ACUGAGAACGGUUCGAGUCGACU	0	15	7	5	14	5	7.7	4

Table 3.4 (Continued)

miRNA name	miRNA sequence	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult	Average read	[a]Reliability level
tcf-miR-n575	AGCGCUCUGACGGAUUACUGAAA	0	6	0	0	0	40	7.7	4
tcf-miR-n576	CGAUCGACGACUCGAAGAUUGCCA	6	0	5	9	8	16	7.3	3
tcf-miR-n576-2	CGAUCGACGACUCGAAGAUUGCCA	6	0	5	9	8	16	7.3	4
tcf-miR-n577	UGGGUAGACUCUCCAUCAUUCGU	14	0	10	14	5	0	7.2	4
tcf-miR-n578	UCCGCCAUCAGCUGAGGAUUG	6	11	11	8	6	0	7.0	4
tcf-miR-n579	GUAUACAAACGAGGUACGGCUAA	0	0	14	12	8	5	6.5	4
tcf-miR-n580	UUUCGAAUUCGAACUUUGUUUU	0	0	5	12	16	6	6.5	4
tcf-miR-n581	AGGUCCCUGGUUCGAGUCC	36	0	0	0	0	0	6.0	4
tcf-miR-n581-2	AGGUCCCUGGUUCGAGUCC	36	0	0	0	0	0	6.0	4
tcf-miR-n582	AAAUAAACUGAUCGAGGUGCUU	0	0	0	0	0	35	5.8	4
tcf-miR-n582-2	AAAUAAACUGAUCGAGGUGCUU	0	0	0	0	0	35	5.8	4
tcf-miR-n583	GAACGGUCCGAGUCGACUCGCU	0	0	0	8	12	15	5.8	2
tcf-miR-n584	UAGGUCUUGGAAUGCUUGUUCCU	13	0	6	9	6	0	5.7	2
tcf-miR-n585	UUACGUAUUCGUUCGGCUGGCCCC	0	8	13	6	7	0	5.7	2
tcf-miR-n585-2	UUACGUAUUCGUUCGGCUGGCCCC	0	8	13	6	7	0	5.7	2
tcf-miR-n586	UCGAGAGUUUUCUCGCGUGCU	0	0	5	10	10	8	5.5	4
tcf-miR-n587	UGAGCACCAAUAAGCACACGGACU	0	8	13	6	5	0	5.3	4
tcf-miR-n588	UACGGCUAAGGGAACCUUUACC	0	0	5	5	10	11	5.2	4
tcf-miR-n589	ACAGAGUAUACGAGCGGAGAUCG	0	10	12	0	8	0	5.0	4
tcf-miR-n590	CUUCUUAUCCGUCUGUCCCGGU	0	11	9	10	0	0	5.0	4
tcf-miR-n591	AUUGUUUGAGAUAAGUGGACCC	0	0	0	0	0	28	4.7	3
tcf-miR-n592	UCCGAGAUCCGUCGAACUGGGAUC	0	8	9	0	5	5	4.5	4
tcf-miR-n593	UGCAAUGGGUCGCACUGAGACU	6	8	5	0	8	0	4.5	4

[a]Reliability level; see the main text in details.

Table 3.5 Nucleotide sequences of putative novel candidate miRNA precursors in *T. cancriformis*

miRNA name	precursor sequence
tcf-miR-n501	CUGUUGAUCAUUUGGCCAAUCCUGCUUACUCGUGAAAGCACUUGGUUAUAUGAUUGACAGCU
tcf-miR-n502	GGAUGGACAACACUUCGGUUGGCAGUUCCUGUUGAUUGUCACUGGGUGCGGUUCAUCCUG
tcf-miR-n503	GCUUCGAAUCUAUGGUUGCAAUGCUGACUUGCUUCAUUGCACUGUAUGGUUCGAGGCGA
tcf-miR-n504	CCCUAUUACUUGUUGGCUCAACAUUACGUCAUGUUGCACUAACCGGUAGUGGGCUU
tcf-miR-n505	AGCCUGAUAUCUUGCUCGACCUUAUGUAUGGCGGUGAGCAAAUCAUCAGGUGCGU
tcf-miR-n506	UUGGUUUGAGAUGAGUGGACCCGUAAAUGUAAUAUCCGGGUCCACUUAUCUCAAACAAU
tcf-miR-n507	AUCGGCACUGGGUAACAAUGAAUGUUCCACUAUGGAAAUAUUAAUAUCAUUGUUAUCCAGUGCCGUGUA
tcf-miR-n508	UGUGAUGUGUUAUUUCAAACUAGUGUGCUUUUAUGCAUUCACUAGUUAAAUAGCACAUCAUUAG
tcf-miR-n509	AAGGAACACUUGCUGUGGUAUGUGGUUGUGUUCUUGUGAGUCGUAUCACAGUGAGUGUUACCUAG
tcf-miR-n510	AGUUACAUAGGUUUAUUCCGCCGUUCGAUAUUUUGAUACCCGGUGAGAUCAACCUAUGUAGCAUU
tcf-miR-n511	GUUUUCACCUUGGCCAAUGUCGUAACGCUAGUAUCGUCUUGGUUAUGGUGAAAAUGG
tcf-miR-n512	CCUUGGUCAGCCAUGUGCUAUGCGAAUUUUAUAACAUUGCACUGGCCUGCCCGGGGGU
tcf-miR-n513	UGUCUUUUCCGCUUGACUGCCGAGCAAGGGCCGAUGAUCAAGUUGAAACCCCGGCAAUCUAGAGGAAGAGACAGA
tcf-miR-n514	AGGAGUGGCACUUUUUGAAAUACGGGAGCGUGAUUGUCGUGAUGGAUCUACAACAAUCAUUGCUCCCGUAGAAGUUUUCGCCACCCCUGAA
tcf-miR-n515	UUCUACACCUCGCAAUAUCGUCUUUGAUGUUGCGGUGAUAUAUUGCGAGGUGCAGAACU
tcf-miR-n516	UCGGUAUUUCUAUUUCUGUCCUUAAGAUAAGGACAGAAAUAGAAAUACCGGUAC
tcf-miR-n517	UUGGUCCCAGGACGGUAGAUGACGCUGCUCUUGCCGGUAUUGGUUACGGUACCACGUCAGGGC
tcf-miR-n518	UUGGUGGUUUCGCUUUCUGGAAGAUUUUUUGGUAUUUCCUAAAGCUAAGCCACCAGAGGG
tcf-miR-n519	UCCAUGAGCUCUCCAACGAACCUUGAAACUUUUGUUGUACUGGCAUAUGCAGCC
tcf-miR-n520	GGUAGACUAGGCUCAGUUUGACUGUUUUCCAAGUGCAGUCACUGGGUGUGGUUUAUCCG
tcf-miR-n521	GGGAGAACCCUACCUUAGUGGAUGGAGCCUAUUUAGUAACAUCACUGGGUAAGGUUCGCCCCCG
tcf-miR-n522	UGUGUAGGAUAGGUGGGAGGCUAUGAAGAUGUGACGCCAGUUACAUUGGAGCCAUCCUUGAAAUACCACCCUG
tcf-miR-n523	UAAGCAGCAUCGGGCUUGGUGACCUUGGAUGGGUGAGCACCAGGAAAGCUGUCGGCU
tcf-miR-n524	UGCAAGGUCGUUCUGCUUACGGUCGAAACGACCACGGUCGUUUUGACAAAGGUCGUUUCGACCCGCACC
tcf-miR-n525	UGGUACGCUGUUUUAGUCACGUGGUUGGUCUGUAACGCUUUUACCACC
tcf-miR-n526	UUCGUGAAUUUUAGCAUAAUGUCGUAACUGAGGAACUUAUCUAUUAUUCCUCUUACAAGGAUAUUAUGCUAAAUUUCAUGAGAUU
tcf-miR-n527	GGAUGGACUUGCCUUGGUUAUAUAAUGUCUAGGUAUCACCAGGUGAGAUUCAUCCAU
tcf-miR-n528	GUGGGGCCGCGAAUGGCGCGGCAGGAAGAUCCGCUUAUUCUGCGACAAUCUUGGCCUCGUCG
tcf-miR-n529	AAAGUUGGACAUUGUUUUGAAACUCUUCAUUACGGGUGAAUGAAGAGUUUCAAAACAAUGUCCAACUUUUAGAAACUU
tcf-miR-n530	AAGAAACGAAUCGGCGCUGAGACUAGUCACAGUUUCCAGAUGCACGGCUCGACUAG
tcf-miR-n531	UCGUAUACGAGACCGGGGUCUCAGCAUGUGUUGCAGAGGCCUCGUUUCGAGUGUACGAAG
tcf-miR-n531-2	UCGUAUACGAGACCGGGGUCUCAGCAUGUGUUGCUGAGGCCUCGUUUCAAGUGUACGAAG
tcf-miR-n532	CUACAUAGGUUGAUCUCACCGGGUAUCAAAAUAUCGAACGGCGGAAUAAACCUAUGUAACU
tcf-miR-n533	CUGGGCAACAAUGAAAAACGCUAUUAUAGCGUUUUUCAUUGUUGCCCAGUCCG
tcf-miR-n534	CAUUCUUGUAGCAGUAUCACCUUAGCAAGGUGAUACUACUACAAGAAUGCA
tcf-miR-n535	CCCUUAGUCCUUCUAGUGCAAUGUGGUUACUUCGAUUCAUUGCACUGGCCGGUCCAGGGUU
tcf-miR-n536	GCAAGAAACAAGUCGGAAAUCGGACGAAAAAAAGUUCUGUCUGAGCCCGGCUUUUACUGCCA
tcf-miR-n537	CCUUGUGCCAGCGGUUGCAUUGGCAAAUUUGUUCAUUGCACUGGCUAGCCCAGGGGA
tcf-miR-n538	UUUUUGAAACACAGCGGUACUGUGUGCUUGAAAAACUGAAGUUUUUGAAACACACUGUACCGUUGUGUUUCAAAAACU
tcf-miR-n539	UCUGAGAACGGUUCGACUCGACUCGCCUUAACCAUAGUGACUCGGCUCGACUCCACGUAACCUUAUGAGUCGACUCGAACCGUUCUCAGCCCU
tcf-miR-n540	UCAACAUUCUGUAACUCAGCCUUGCAAAGAGAACUAUAGGAGACCG
tcf-miR-n541	UGCGCACCGAGUUUUUGAAACACAGCGAUUACUUGUGUUUGAAAAACUCAGUGCGCACC
tcf-miR-n542	ACGGGUACACCUUAAGGGCAGCUUCCAUGGCAUCUGUCCCAGAAGCUUCCAUAGGUCGUACCUGAAU
tcf-miR-n543	UUUAAAAAAGAUCAGAUCAGAUCAGAUCAAAUGAUCUGAUCUGAUCUUUUUUACACC
tcf-miR-n544	UUGGAUAGAGGUAAUCUCGGCCUGGUCGACUUCUUCUGCUCGUAGAAGGGAGUCUUUCGGCUUCGGAUUACCUGCUUUGUCCGAAG
tcf-miR-n545	UGUGUCUUGUCGUGCUGCUGCUUAUACAUGUAAUGCAGACAGCAGAAAAAGGUACAGA
tcf-miR-n546	GUACGGUUUUCAGGGAGUCGGCUUUCGGGAAAAAAUUUCUCGCAAAUGAGCUUGACUGGCCUGAGCCGAGGACU
tcf-miR-n547	AGUAGUAAAGGGUGGGAUGCACUGUAAAGUAACAAAGUGCACUUACUGACUUUACUGCCA
tcf-miR-n548	UAUGAACUCGUCCCUUGGUCACGUGUGGGUGUCCCUCAGAUGGUUGAUCCACGCGUCCGGGUGGACAGAUGAUGCC
tcf-miR-n549	AUGGCGGACAAAGGGUUUGACAUUCAAAGAGUCUUGCCUGACGGUCUCGCUAUUA
tcf-miR-n550	UAAGAACGAGAGCAACUCAGAACCCCUCCGGGGUGAUGAUGAUAGCUCCGGAUUGUGCCAGUUCUGAGUUUCUCGUUCAUUGGCA
tcf-miR-n551	UCGACCAGUACUCGUCACGGCACCGUCUUGUUACUGAUCGGUACUGGUUCCU
tcf-miR-n552	UAUCUAGCAAUCCUGAUGAUGUCGCAAUCGUGGGAAAGAUUUGGUCCACUAAAUUUCCCCCACCCAUAUCAUCGUCUUCUAUGGAUGGCU
tcf-miR-n553	UGUAUAGUACUCGGCAUAAAAAAAAUGACAGUUCGCUUUUAUGCCGAGCACUAUACAGU
tcf-miR-n554	UUUUCCGGAUCUCUCGAAACCGUUUUUGGUGGAUUCGGAUGCAAGC
tcf-miR-n555	UUCUCUCCCGGACUCUUCUUGGUCCGGUGGACUAGAGACCUAAAAAGGUCUAUCCAUUGGAACCAUUCCAAGGCAGAACCGGAAUACGGCU
tcf-miR-n556	AUAACCUUUUCGUUGGAAAUCGCUGUCCCAGGAAUAGGAAAAGUCAGAGAUUUCCAAGGAAAAGGUUAUAAUAACC
tcf-miR-n557	UAGACCGCGCAAUUCAGAAACAUAGGAUUUGACUUCUGCUUGCUGCUUUUGAAUUGCGCGGUCUAGC
tcf-miR-n558	CUUGUUGUUGACGCGACUGCUGUUAGGAAGACCUUCCCAGUUGUAGUUCUUCCAAACUAGGUCUGUCGCAACGGCGAGCC
tcf-miR-n559	UGCGUGACAGAAUUCAUAAGAAAAAGUAUGAUAUAACAAUGUCUGAGAUUUCUGUCACGUACAG
tcf-miR-n560	GCUCUCAAAGUUGAGCUUCUCGAGUGGGAGAAGUCUUAGCGAAGGAGAGCUU
tcf-miR-n561	CCCCGUGGAAGUCUGGAUAUUUUAUAUAAAAAAUUCCAGGCUCUUCGGGAU
tcf-miR-n562	UAAAGAUUGACAAGCUGGCGUAGCAAAAGUUAUGUCAAUUGUCGAGUGGCUGCC
tcf-miR-n563	AUCGGACAAGUAGUUUUUACGAAAUCAUGACUAGAGUUUCCAAAUACUAGUAAAAUGGUUUCGUAAAAACCACAUGUCCGAUUUG
tcf-miR-n564	GAAAUGCUCCCGUCGUUUGGGUCUCCCUGGGCCCAAACGACUAGAGCAUUUUGC
tcf-miR-n565	UUGUCGGAGUAGAUACUCAAUGGCUUGCCGCGGGUCGCAGCGAAGCGACGCAGUACAAGUAGGAAGUCAUCCGUCGACAAGGU
tcf-miR-n566	AACCAGGCUGAGGUCUUGACGAGAAUUUCAAUGAGAUUCUUGUACUUUGGCCUGGAAACAAAU
tcf-miR-n567	UGUUUCGCUUGUUGUGCGGCGUUCUCCAUAGCUCGAGCUUUGUCUAGGAUCGUUUCAAGUGACACGUC
tcf-miR-n568	UAAGCUCUACGUCUGGAGGCAUCCGGCGAUACUGGCCACUUUUGCCAGUAUCUGGAUGCAUAGAGGCGAAGGAGUGUGGU
tcf-miR-n569	AUUGGACCUAGAUCUCGGAAACGCAGUGCUUAGACACUUCGCUUAGAGAUACUGGGAUAAAUGG
tcf-miR-n570	UAGGGGAUAGUGCAGAGGUCUAUUUAUUCACUAGACCUCUGCACUAUCCCCUGUA

Table 3.5 (Continued)

miRNA name	precursor sequence
tcf-miR-n571	CUGCAGAGCGCAAACCUUGAAGUACAUGUACUUCAAGGUUUGCGCUCUGCAACU
tcf-miR-n572	UUGGCUCGGAAAUGGGGAUUUUCUCCUCAGUUCCCUCUUCCGAACAAGCAAGAGCAGGGCACC
tcf-miR-n573	UUCUUCCCAAAAACCUUGUUGGCUAGUCACUAUAUAGCAGAAACUGUCGGUGUCGGUGAAUAUGAGUCGACAACGGGAACCGAACC
tcf-miR-n574	ACUGAGAACGGUUCGAGUCGACUCAUAAGGUUACGUGGAGUCGAGUCGAGUCACUAUGGUUAAAGCGAGUCGAGUCGAACCGUUCUCAGACC
tcf-miR-n575	AGCGCUCUGACGGAUUACUGAAAAAGCCUUCGAUUCCUUUUGAUACCAAAAGGAAUCGAAGGCUUUUUCAGUAAUCCGUCAGAGCGCU
tcf-miR-n576	CGAUCGACGACUCGAAGAUUGCCACGUUGCACCUGUUCGUGGCGUUCGGGUUACAGAUUGUC
tcf-miR-n576-2	CGAUCGACGACUCGAAGAUUGCCACGUUGCACCUGUUCGUGGCGUUCGGGUUACCGAUUGUC
tcf-miR-n577	UGGGUAGACUCUCCAUCAUUCGUUUGGUCGAUGGUAGUAGAAAUUGCUCGUCGGAUGUAAGUGGAUCUUUACCCACG
tcf-miR-n578	UCCGCCAUCAGCUGAGGAUUGAUGGUCACAUCAAUCCUCAGCUGAUGGCGGAUU
tcf-miR-n579	UGCCGUACCUCGUUUGUGUGAUGUGAUAAAACAAAUGUAUACAAACGAGGUACGGCUAA
tcf-miR-n580	AACAAAGUUCGAUUUCGAAACGUCGUGCGCACCGUUUCGAAUUCGAACUUUGUUUU
tcf-miR-n581	AGCUUGAACGAGUGGAUAGCGUACCUGCCUGUGGUGCGUGAGGUCCCUGGUUCGAGUCC
tcf-miR-n581-2	GCCUUGAACGAGUGGAUAGCGUACCUGCCUGUGGUGCGUGAGGUCCCUGGUUCGAGUCC
tcf-miR-n582	GCACCUCGAUCAGUUUAUUUUUUUUGCAAAAAAAUAAACUGAUCGAGGUGCUU
tcf-miR-n582-2	GCACCUCGAUCAGUUUAUUUUUUUGCAAAAAAUAAACUGAUCGAGGUGCUU
tcf-miR-n583	GAACGGUCCGAGUCGACUCGCUAAGGUUACGGCGAGUCGAGUCGAACCCUGACC
tcf-miR-n584	GUGCGCUGGUCAGGGCUUACUUUAGGUCUUGGAAUGCUUGUUCCU
tcf-miR-n585	UUACGUAUUCGUUCGGCUGGCCCCUCUUCCUCGCUUUCAGCUAGGGGGAUGGGUUCGGCCGAAGCGCCUAACGCCACG
tcf-miR-n585-2	UUACGUAUUCGUUCGGCUGGCCCCUCUUCCUCGCUUUCAGCUAGGGGAAUGGGUUCGGCCGAAGCGCCUAACGCCACG
tcf-miR-n586	UCGAGAGUUUUCUCGCGUGCUUGGUCGAUCAACUUAGAGCACGCGAGAAACUCAUUCUCGACA
tcf-miR-n587	UGAGCACCAAUAAGCACACGGACUCCAGAUUUCGCAUUGUAGCGCUUGCCGUGGCUUCGUUGGGACCAUCC
tcf-miR-n588	UACGGCUAAGGGAACCUUUACCGGCGCAGGCUGGUAAAGAGUCUUUAGCCGUCCC
tcf-miR-n589	ACUCCCUCUGUCUACUCUGUGCUCUGUAUACUCUGUGUUGAGCACAGAGUAUACGAGCGGAGAUCG
tcf-miR-n590	CCGGGGAAGAGGGAAAGGUGGCCCUGUUCAAGUAAGCAGCAAAUGUUCGCUUACGCGGGCUUCUUAUCCGUCUGUCCCGGU
tcf-miR-n591	AUUGUUUGAGAUAAGUGGACCCGGAUAUUACAUUUACGGGUCCACUCAUCUCAAACCAAUAC
tcf-miR-n592	UCCGUUGCGUGGAUCGCGAUUCGUGAGCCCGAAGAUCCGCGAACCGGGGAUCCAAAAUCCGCGAUCCGAGAUCCGUCGAACUGGGAUC
tcf-miR-n593	UGCAAUGGGUCGCACUGAGACUGAGGCAUGCCUCCCUCAGUGUUCGAGCUUUUGUCUU

3.3.2 Changes in the expression of miRNAs during *T. cancriformis* development

It has been reported that miRNAs are intimately related to development (Wienholds and Plasterk 2005; Kloosterman and Plasterk 2006), so I hypothesized that miRNA expression will change markedly with the dramatic morphological changes that occur of the early larvae of *T. cancriformis.* First, the expression of 87 conserved *T. cancriformis* miRNAs was analyzed during the six developmental stages based on the read counts for each miRNA. miRNA expression was normalized using spike reads (Tables 3.6–3.7), and these conserved miRNAs were then roughly clustered into seven groups based on their expression patterns (Groups 1-I to 1-VII in Figure 3.3A), and showed that the expression patterns of the conserved miRNAs varied throughout the six developmental stages of *T. cancriformis*. Some miRNAs were stage-specifically expressed in the egg, 4th instar larval, and adult stage (Groups 1-I to 1-III in Figure 3.3A). To validate the expression of the conserved miRNAs, northern blotting analyses were performed, and 13 miRNAs (tcf-let-7-5p, tcf-miR-1, tcf-miR-2b, tcf-miR-12, tcf-miR-34, tcf-miR-87, tcf-miR-125, tcf-miR-133, tcf-miR-184-3p, tcf-miR-276-3p, tcf-miR-279a, tcf-miR-375, and tcf-miR-750) were detected, at approximately 20–25 nt in length, in the adult stage (Figure 3.3B–C), suggesting that these 13 miRNAs are actually expressed in the adult stage of *T. cancriformis.* The expression patterns of six (tcf-let-7-5p, tcf-miR-2b, tcf-miR-12, tcf-miR-34, tcf-miR-87, and tcf-miR-125) of these 13 miRNAs were then investigated in the six developmental stages (Figure 3.3B). A strong correlation was observed between the northern blotting data and the read counts from the deep-sequencing analysis (tcf-let-7-5p: $r = 0.95$, $p = 0.00273$; tcf-miR-2b: $r = 0.96$,

Table 3.6 Normalization of small RNA reads in this study

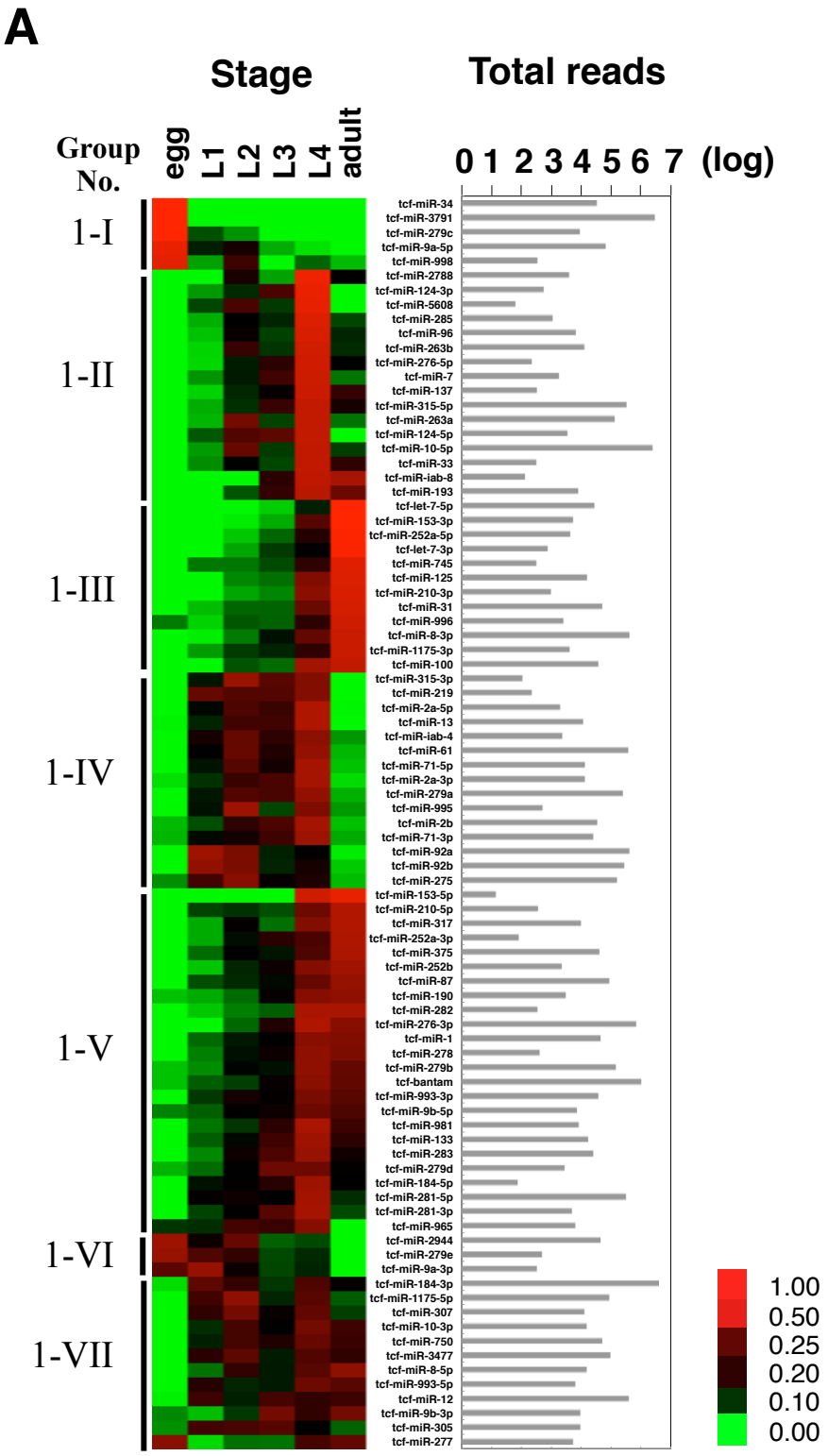
	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
Spike1	28,746	30,235	36,053	61,382	42,651	54,232
Spike2	2,150	3,143	3,545	6,407	4,063	4,264
Spike1/Spike2	13.4	9.6	10.2	9.6	10.5	12.7
Fold change to standard	1.432395941	1.155164531	0.996075363	0.567836288	0.855427095	0.740516285
Total small RNA reads	23,438,746	15,019,413	20,078,802	17,921,148	21,597,504	22,625,712
Normalized small RNA reads	33,573,565	17,349,893	20,000,000	10,176,278	18,475,090	16,754,708

Table 3.7 Normalized relative expression of conserved miRNAs in *T. cancriformis*

miRNA name	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
tcf-bantam	0.057121	0.104927	0.121672	0.167790	0.304315	0.244175
tcf-let-7-3p	0.000000	0.032283	0.068328	0.124791	0.164087	0.610511
tcf-let-7-5p	0.001179	0.007195	0.024495	0.053396	0.139872	0.773864
tcf-miR-1	0.007726	0.098377	0.143675	0.159217	0.302586	0.288420
tcf-miR-2a-5p	0.003569	0.152567	0.221907	0.203198	0.378162	0.040597
tcf-miR-2a-3p	0.046846	0.128939	0.201956	0.217808	0.355902	0.048549
tcf-miR-2b	0.060343	0.113037	0.193123	0.219975	0.355245	0.058278
tcf-miR-7	0.028306	0.076726	0.142161	0.208216	0.453131	0.091460
tcf-miR-8-5p	0.024163	0.092751	0.195611	0.143266	0.232017	0.312192
tcf-miR-8-3p	0.031796	0.042876	0.088634	0.147501	0.245602	0.443591
tcf-miR-9a-5p	0.548965	0.139738	0.176769	0.065802	0.046331	0.022395
tcf-miR-9a-3p	0.237579	0.326249	0.170550	0.118056	0.137836	0.009729
tcf-miR-9b-5p	0.085122	0.103034	0.161473	0.170818	0.257730	0.221823
tcf-miR-9b-3p	0.082183	0.059650	0.125729	0.268781	0.193866	0.269792
tcf-miR-10-5p	0.008680	0.073863	0.252942	0.123266	0.419078	0.122172
tcf-miR-10-3p	0.019225	0.132691	0.214815	0.156682	0.268976	0.207611
tcf-miR-12	0.042055	0.204916	0.140088	0.213217	0.194024	0.205700
tcf-miR-13	0.040978	0.136922	0.209634	0.208424	0.377097	0.026945
tcf-miR-31	0.004322	0.059801	0.099143	0.100518	0.256123	0.480092
tcf-miR-33	0.034592	0.080203	0.156356	0.118275	0.418329	0.192244
tcf-miR-34	0.983118	0.002664	0.001929	0.001117	0.002314	0.008857
tcf-miR-61	0.021304	0.162689	0.250492	0.183230	0.320953	0.061333
tcf-miR-71-5p	0.035781	0.145893	0.230012	0.175346	0.355787	0.057181
tcf-miR-71-3p	0.062016	0.151325	0.174871	0.207501	0.337722	0.066564
tcf-miR-87	0.029611	0.112611	0.134067	0.151208	0.248007	0.324496
tcf-miR-92a	0.039481	0.343180	0.277786	0.138372	0.158139	0.043043
tcf-miR-92b	0.036626	0.316887	0.277827	0.136978	0.177265	0.054417
tcf-miR-96	0.030354	0.057527	0.168112	0.121106	0.486599	0.136301
tcf-miR-100	0.000528	0.024334	0.110082	0.096487	0.347654	0.420914
tcf-miR-124-5p	0.000000	0.107623	0.225336	0.238700	0.422350	0.005990
tcf-miR-124-3p	0.000000	0.074859	0.134333	0.217805	0.548359	0.024643
tcf-miR-125	0.002450	0.032202	0.082040	0.094473	0.289410	0.499424
tcf-miR-133	0.001092	0.104171	0.152550	0.207491	0.343174	0.191522
tcf-miR-137	0.000000	0.051154	0.135269	0.167638	0.446997	0.198941
tcf-miR-153-5p	0.000000	0.000000	0.000000	0.000000	0.456560	0.543440
tcf-miR-153-3p	0.007579	0.018338	0.043438	0.066414	0.229446	0.634785
tcf-miR-184-5p	0.000000	0.147396	0.165226	0.188382	0.338366	0.160630
tcf-miR-184-3p	0.047858	0.240124	0.197854	0.126693	0.222208	0.165263
tcf-miR-190	0.057299	0.066663	0.097001	0.167196	0.296192	0.315649
tcf-miR-193	0.000000	0.032998	0.108779	0.197664	0.402223	0.258336
tcf-miR-210-5p	0.000000	0.118746	0.129337	0.112134	0.259174	0.380609
tcf-miR-210-3p	0.025512	0.022860	0.068992	0.083718	0.307255	0.491662
tcf-miR-219	0.000000	0.250103	0.228598	0.228671	0.292628	0.000000
tcf-miR-252a-5p	0.007572	0.027345	0.056087	0.099967	0.185788	0.623241
tcf-miR-252a-3p	0.000000	0.066700	0.149536	0.190166	0.217329	0.376269
tcf-miR-252b	0.000000	0.056270	0.136299	0.170237	0.293201	0.343994
tcf-miR-263a	0.032819	0.062422	0.269597	0.119439	0.422762	0.092960
tcf-miR-263b	0.021221	0.050973	0.201358	0.128493	0.464951	0.133004
tcf-miR-275	0.079399	0.212167	0.301354	0.167421	0.179530	0.060130
tcf-miR-276-5p	0.000000	0.049807	0.141728	0.190971	0.461043	0.156451
tcf-miR-276-3p	0.004690	0.034997	0.098616	0.183036	0.378952	0.299709

Table 3.7 (Continued)

miRNA name	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
tcf-miR-277	0.303924	0.049020	0.094420	0.093048	0.213090	0.246498
tcf-miR-278	0.000000	0.087223	0.148142	0.170203	0.313167	0.281265
tcf-miR-279a	0.031766	0.144352	0.227229	0.215360	0.317193	0.064099
tcf-miR-279b	0.057121	0.080401	0.154087	0.148205	0.314368	0.245817
tcf-miR-279c	0.756305	0.111744	0.077541	0.025047	0.017559	0.011804
tcf-miR-279d	0.062600	0.096864	0.157492	0.258854	0.258957	0.165234
tcf-miR-279e	0.330094	0.223342	0.196474	0.116441	0.133649	0.000000
tcf-miR-281-5p	0.018174	0.166631	0.171815	0.160727	0.349990	0.132662
tcf-miR-281-3p	0.022531	0.119272	0.156880	0.231541	0.353295	0.116481
tcf-miR-282	0.024267	0.055450	0.087189	0.104218	0.367142	0.361733
tcf-miR-283	0.025533	0.082587	0.170926	0.220219	0.327221	0.173515
tcf-miR-285	0.021006	0.064771	0.159819	0.136662	0.499567	0.118176
tcf-miR-305	0.078566	0.211565	0.214874	0.231563	0.161343	0.102087
tcf-miR-307	0.003299	0.194199	0.271075	0.168737	0.241493	0.121198
tcf-miR-315-5p	0.001444	0.065792	0.130752	0.200308	0.424513	0.177190
tcf-miR-315-3p	0.000000	0.144843	0.330081	0.233942	0.291134	0.000000
tcf-miR-317	0.002635	0.066577	0.166016	0.094003	0.290742	0.380027
tcf-miR-375	0.004102	0.096834	0.161315	0.145628	0.220961	0.371160
tcf-miR-745	0.000000	0.090985	0.090525	0.115253	0.191764	0.511473
tcf-miR-750	0.013542	0.140057	0.213584	0.182257	0.250405	0.200155
tcf-miR-965	0.127386	0.131482	0.212221	0.201636	0.297927	0.029348
tcf-miR-981	0.019146	0.091263	0.127908	0.190085	0.365067	0.206531
tcf-miR-993-5p	0.038487	0.186940	0.136431	0.143188	0.260883	0.234072
tcf-miR-993-3p	0.023603	0.125221	0.175490	0.167025	0.276109	0.232552
tcf-miR-995	0.035458	0.147377	0.339513	0.117859	0.285058	0.074735
tcf-miR-996	0.089417	0.051507	0.108138	0.101937	0.191044	0.457956
tcf-miR-998	0.523610	0.070915	0.205682	0.039613	0.100255	0.059925
tcf-miR-1175-5p	0.001489	0.208562	0.313804	0.136149	0.234346	0.105650
tcf-miR-1175-3p	0.033241	0.073111	0.123983	0.136834	0.192099	0.440731
tcf-miR-2788	0.000000	0.014121	0.177536	0.070388	0.571387	0.166567
tcf-miR-2944	0.343096	0.169956	0.252422	0.102789	0.116080	0.015657
tcf-miR-3477	0.009061	0.189782	0.238616	0.141475	0.226540	0.194526
tcf-miR-3791	0.921396	0.014440	0.024746	0.008305	0.014729	0.016385
tcf-miR-5608	0.000000	0.118343	0.218667	0.124656	0.538334	0.000000
tcf-miR-iab-4	0.000000	0.175875	0.248123	0.192016	0.308924	0.075062
tcf-miR-iab-8	0.000000	0.000000	0.035525	0.190369	0.414922	0.359185



(Legend on next page)

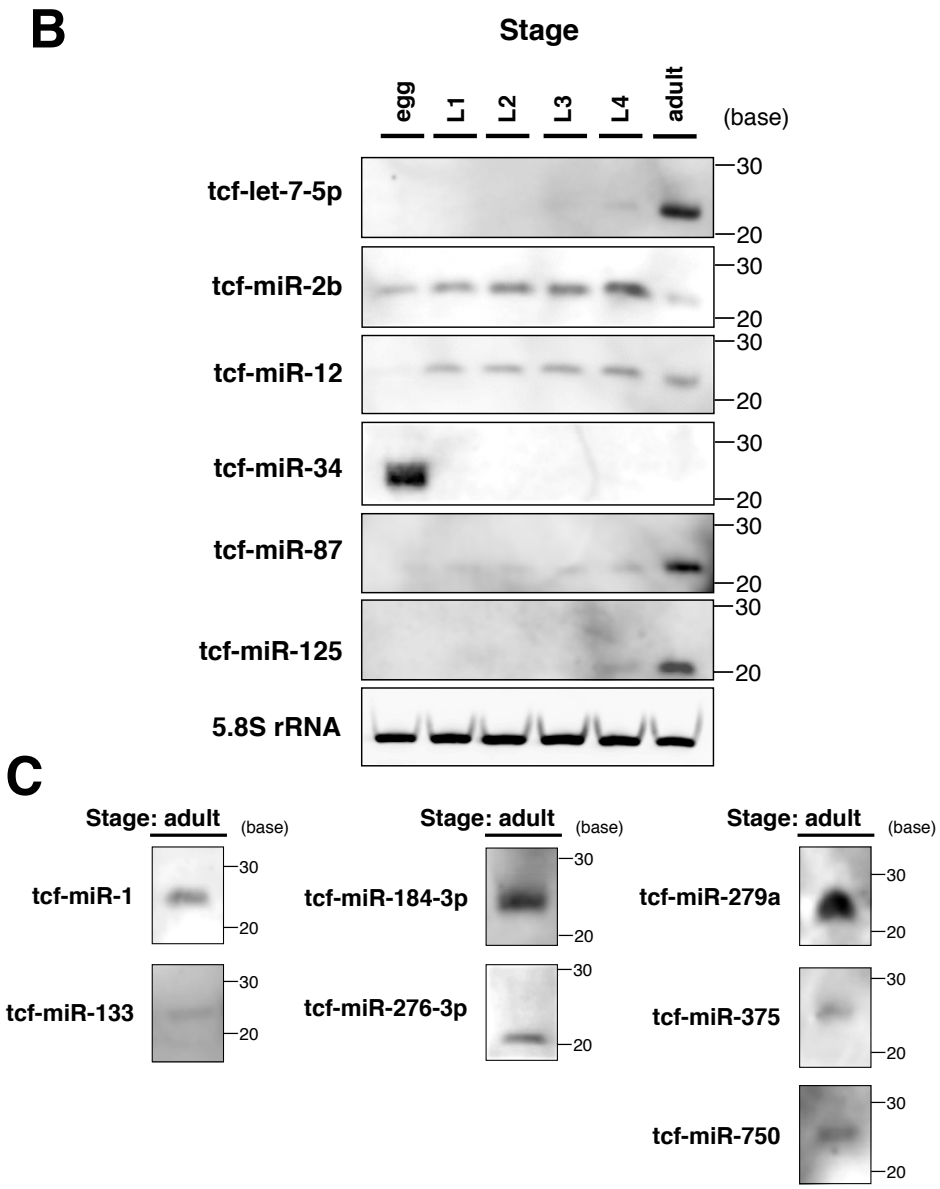


Figure 3.3 Expression of conserved *T. cancriformis* miRNAs

(A) Expression profiles of conserved miRNAs based on read numbers in the deep-sequencing analysis. miRNA reads were normalized to the spike reads in each stage. A more intense red color indicates a more strongly expressed miRNA, whereas a more intense green color indicates a more weakly expressed miRNA. Conserved miRNAs were categorized into seven groups based on their expression patterns (Groups 1-I to 1-VII). Total read numbers for the six stages are listed. (B) Northern blotting analysis of six conserved miRNAs during *T. cancriformis* development. A typical pattern of 5.8S rRNA expression is also shown as the loading control. (C) Northern blotting analysis of the expression of eight conserved miRNAs in the adult stage.

p = 0.00192; tcf-miR-12: r = 0.82, p = 0.04336; tcf-miR-34: r = 0.98, p = 0.00016; tcf-miR-87: r = 0.81, p = 0.04738; and tcf-miR-125: r = 0.98, p = 0.00027) (see Materials and Methods). These results show that this method efficiently extracted the conserved miRNAs and that their expression profiles based on read numbers are reliable.

I then investigated whether the expression patterns of the six conserved miRNAs were common to *T. cancriformis* and *D. melanogaster*, a representative model arthropod species. I used northern blotting data from a previous study of *D. melanogaster* (Sempere et al. 2003). The expression patterns of two (tcf-let-7-5p and tcf-miR-125) of the six conserved miRNAs were similar and increased toward the adult stages of *T. cancriformis* and *D. melanogaster* (Group 1-III in Figure 3.3A and B). In contrast, the expression patterns of four conserved miRNAs (tcf-miR-2b, tcf-miR-12, tcf-miR-87, and tcf-miR-34) were quite different in these species. *Triops cancriformis* tcf-miR-2b was predominantly expressed in the larval stages, especially increasing toward the 4th instar (Group 1-IV in Figure 3.3A and B), whereas in *D. melanogaster*, dme-miR-2 was expressed throughout the developmental stages, and especially in the egg stage. tcf-miR-12 was expressed from the 1st instar to the adult stage (Group 1-VII in Figure 3.3A and B), whereas the expression of dme-miR-12 decreased toward the adult stage. tcf-miR-34 was only expressed in the egg stage (Group 1-I in Figure 3.3A and B), whereas dme-miR-34 was strongly expressed in the adult stage. tcf-miR-87 was expressed throughout all six stages, but was especially strongly expressed in the adult stage (Group 1-V in Figure 3.3A and B), whereas dme-miR-87 was detected in the egg,

1st and 2nd instar larval, pupal, and adult stages. This inconsistency in the expression patterns of the conserved miRNAs suggests that although these miRNAs have very similar sequences (≧80%), they may play different roles in different species.

To examine whether the interactions between *D. melanogaster* miRNAs and their target genes are conserved in *T. cancriformis*, I searched for the target genes of the miRNAs of *T. cancriformis*. I focused on three miRNAs, tcf-miR-2b, tcf-miR-12, and tcf-miR-34, whose expression timing during development differs in *T. cancriformis* and *D. melanogaster*. It has been reported that dme-miR-2b targets *grim* (involved in apoptosis during the development), dme-miR-12 targets *cos* (involved in the maintenance of homeostasis in the hedgehog pathway), and dme-miR-34 targets *Eip74EF* (unknown function) (Burgler and Macdonald 2005; Leaman et al. 2005; Friggi-Grelin et al. 2008). I used a bioinformatic analysis to predict the orthologous genes of *grim*, *cos*, and *Eip74EF* in *T. cancriformis* (see Materials and Methods). The amino acid lengths of the deduced *T. cancriformis* proteins (grim, 194 amino acids; cos, 1,143 amino acids; and Eip74EF, 841 amino acids) were similar to those of *D. melanogaster* (grim, 138 amino acids; cos, 1,201 amino acids; and Eip74EF, 883 amino acids). The corresponding *T. cancriformis* proteins showed 53.6% amino acid sequence similarity to grim, 62.2% similarity to cos, and 61.4% similarity to Eip74EF of *D. melanogaster* (data not shown). The kinesin motor domain and ETS domain were predicted in cos and Eip74EF, respectively, in both species. Two software programs, RNAhybrid and miRanda (John et al. 2004; Rehmsmeier et al. 2004) were used for the target prediction, to extract reliable target sites in the 3′-untranslated region (UTR)

sequence of each gene. Initially, I confirmed that all three miRNA–target gene interactions were successfully predicted in *D. melanogaster*. *Triops cancriformis* miRNA target prediction identified miR-12/*cos* and miR-34/*Eip74EF*, whereas miR-2b/*grim* was not predicted. Although the expression patterns of miRNAs tcf-miR-34 and tcf-miR-12 were quite different during the developmental stages of *T. cancriformis* and *D. melanogaster*, the miRNAs targeted orthologous genes in the two species. This suggests that the timing of gene regulation by these miRNAs is important and that their roles may differ in different species. However, tcf-miR-2b did not target the *grim* gene in *T. cancriformis*, and the miRNA expression patterns of miR-2b during the developmental stages of these two species differed, suggesting that miR-2b target different genes, at least in *T. cancriformis* and *D. melanogaster*. Even when the miRNA–target gene interactions were conserved between species, I noted that the timing of the expression of some miRNAs differed, and the roles of these miRNAs may vary across species. I speculated that these different miRNA expression patterns may produce differences during species differentiation.

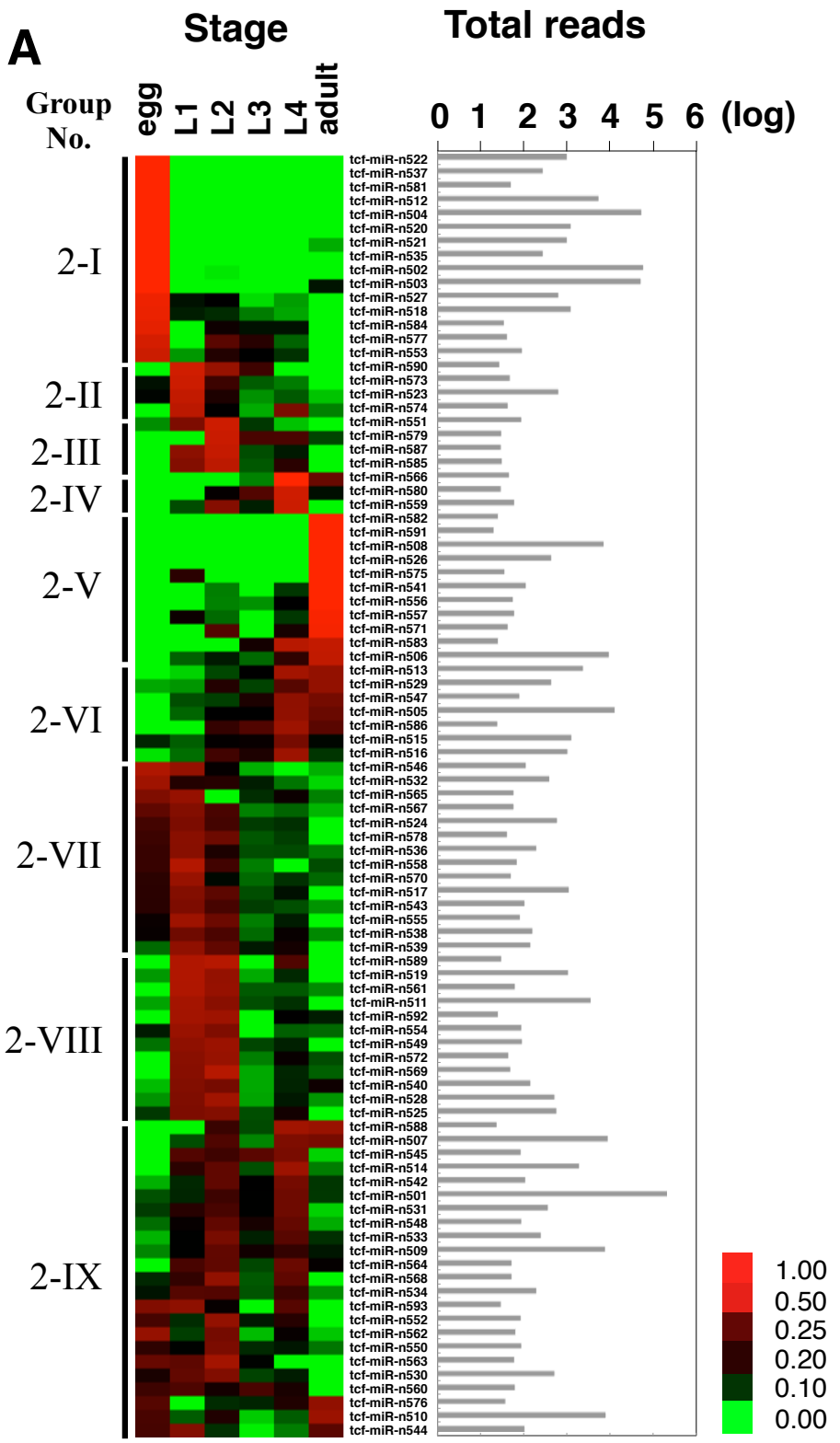
To characterize these novel candidate miRNAs, an expression analysis was conducted with the same method used for the conserved miRNAs (Tables 3.6 and 3.8). Based on their expression patterns during development, the candidate miRNAs were roughly clustered into nine groups (Groups 2-I to 2-IX in Figure 3.4A), showing that expression of these novel candidate miRNAs also varied throughout the six developmental stages. In particular, I found that some miRNAs were stage-specifically expressed in the egg, 1st instar, 2nd instar, 4th instar larval, or adult stage (Groups 2-I to

Table 3.8 Normalized relative expression of novel candidate miRNAs in *T. cancriformis*

miRNA name	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
tcf-miR-n501	0.111554	0.136144	0.206347	0.157172	0.262815	0.125967
tcf-miR-n502	0.861694	0.037547	0.044632	0.017245	0.021867	0.017015
tcf-miR-n503	0.831527	0.007357	0.007329	0.003515	0.004771	0.145500
tcf-miR-n504	0.950499	0.020240	0.015818	0.006514	0.005600	0.001329
tcf-miR-n505	0.010353	0.100467	0.159503	0.158920	0.315935	0.254822
tcf-miR-n506	0.039977	0.102603	0.142380	0.098340	0.194715	0.421986
tcf-miR-n507	0.019855	0.113685	0.224360	0.084021	0.284276	0.273803
tcf-miR-n508	0.001611	0.000000	0.000000	0.000000	0.000000	0.998389
tcf-miR-n509	0.083571	0.157717	0.244056	0.172861	0.197485	0.144309
tcf-miR-n510	0.216958	0.105602	0.183009	0.054018	0.105253	0.335160
tcf-miR-n511	0.070025	0.353763	0.303084	0.113113	0.130264	0.029752
tcf-miR-n512	0.986627	0.001093	0.004148	0.001075	0.004534	0.002523
tcf-miR-n513	0.000000	0.051609	0.116976	0.158739	0.351604	0.321072
tcf-miR-n514	0.026304	0.190920	0.243543	0.112619	0.338416	0.088198
tcf-miR-n515	0.135898	0.104023	0.165779	0.168924	0.272361	0.153015
tcf-miR-n516	0.045952	0.095454	0.209158	0.181062	0.340954	0.127420
tcf-miR-n517	0.189692	0.292519	0.238864	0.113306	0.147729	0.017890
tcf-miR-n518	0.552611	0.140489	0.133174	0.088725	0.068898	0.016103
tcf-miR-n519	0.074760	0.380048	0.323066	0.066683	0.134738	0.020705
tcf-miR-n520	0.887713	0.018282	0.037337	0.023177	0.033490	0.000000
tcf-miR-n521	0.886102	0.011715	0.012122	0.005759	0.018218	0.066086
tcf-miR-n522	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
tcf-miR-n523	0.154581	0.422768	0.180714	0.078153	0.107032	0.056751
tcf-miR-n524	0.212232	0.282018	0.212991	0.124289	0.131066	0.037404
tcf-miR-n525	0.125759	0.282384	0.293221	0.113394	0.173769	0.011473
tcf-miR-n526	0.000000	0.000000	0.011325	0.021951	0.038904	0.927820
tcf-miR-n527	0.561788	0.147336	0.158806	0.047981	0.072283	0.011806
tcf-miR-n528	0.077472	0.285611	0.348251	0.069101	0.143755	0.075811
tcf-miR-n529	0.068163	0.075911	0.185085	0.119666	0.230672	0.320503
tcf-miR-n530	0.178315	0.244915	0.288684	0.119287	0.141432	0.027367
tcf-miR-n531	0.124330	0.188000	0.216145	0.157104	0.262196	0.052224
tcf-miR-n532	0.344516	0.185225	0.187167	0.140843	0.092157	0.050093
tcf-miR-n533	0.063240	0.157638	0.275854	0.139024	0.233469	0.130775
tcf-miR-n534	0.146270	0.230023	0.228858	0.110171	0.205278	0.079399
tcf-miR-n535	0.878151	0.020952	0.032520	0.014419	0.021722	0.032236
tcf-miR-n536	0.200920	0.300919	0.179638	0.113785	0.115704	0.089033
tcf-miR-n537	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
tcf-miR-n538	0.167565	0.263155	0.220781	0.097892	0.168539	0.082068
tcf-miR-n539	0.097598	0.314835	0.244328	0.143154	0.174857	0.025228
tcf-miR-n540	0.060188	0.291234	0.272052	0.067603	0.137786	0.171137
tcf-miR-n541	0.000000	0.000000	0.087086	0.034752	0.127142	0.751019
tcf-miR-n542	0.064705	0.135674	0.242976	0.159035	0.270495	0.127115
tcf-miR-n543	0.188865	0.282863	0.215764	0.123002	0.112790	0.076716
tcf-miR-n544	0.213825	0.290993	0.130105	0.042383	0.087791	0.234903
tcf-miR-n545	0.000000	0.227278	0.207505	0.236587	0.277208	0.051422
tcf-miR-n546	0.378108	0.325256	0.166524	0.064953	0.000000	0.065158

Table 3.8 (Continued)

miRNA name	Egg	1st Instar	2nd Instar	3rd Instar	4th Instar	Adult
tcf-miR-n547	0.000000	0.112914	0.121704	0.180389	0.313557	0.271437
tcf-miR-n548	0.095930	0.167621	0.244600	0.177469	0.248255	0.066125
tcf-miR-n549	0.093445	0.313999	0.335736	0.117306	0.139514	0.000000
tcf-miR-n550	0.193449	0.156008	0.280255	0.134204	0.144410	0.091675
tcf-miR-n551	0.079663	0.282675	0.454252	0.126321	0.057089	0.000000
tcf-miR-n552	0.214509	0.133071	0.321283	0.143908	0.187230	0.000000
tcf-miR-n553	0.450813	0.075220	0.183771	0.160226	0.129971	0.000000
tcf-miR-n554	0.142406	0.331773	0.286081	0.037635	0.103944	0.098161
tcf-miR-n555	0.169565	0.341867	0.259411	0.087386	0.141770	0.000000
tcf-miR-n556	0.000000	0.000000	0.085603	0.078080	0.161734	0.674584
tcf-miR-n557	0.000000	0.170162	0.097818	0.000000	0.126009	0.606011
tcf-miR-n558	0.200669	0.388394	0.209315	0.087505	0.000000	0.114116
tcf-miR-n559	0.000000	0.114087	0.295126	0.140203	0.450584	0.000000
tcf-miR-n560	0.206400	0.221937	0.175424	0.218192	0.178046	0.000000
tcf-miR-n561	0.000000	0.378937	0.326750	0.106441	0.106900	0.080972
tcf-miR-n562	0.324074	0.121964	0.270430	0.059953	0.167732	0.055846
tcf-miR-n563	0.252747	0.240889	0.351516	0.154847	0.000000	0.000000
tcf-miR-n564	0.000000	0.216648	0.242855	0.117146	0.256693	0.166658
tcf-miR-n565	0.285409	0.326074	0.000000	0.132000	0.170446	0.086071
tcf-miR-n566	0.000000	0.000000	0.000000	0.085263	0.660583	0.254154
tcf-miR-n567	0.241138	0.291701	0.217991	0.086034	0.100805	0.062331
tcf-miR-n568	0.135128	0.196154	0.319487	0.107136	0.242095	0.000000
tcf-miR-n569	0.000000	0.298121	0.395482	0.067636	0.135856	0.102905
tcf-miR-n570	0.191130	0.330296	0.151897	0.097417	0.130449	0.098810
tcf-miR-n571	0.000000	0.000000	0.228575	0.000000	0.176669	0.594756
tcf-miR-n572	0.000000	0.303588	0.327223	0.087053	0.168611	0.113525
tcf-miR-n573	0.147796	0.452926	0.205552	0.105462	0.088264	0.000000
tcf-miR-n574	0.000000	0.404680	0.162842	0.066309	0.279697	0.086473
tcf-miR-n575	0.000000	0.189622	0.000000	0.000000	0.000000	0.810378
tcf-miR-n576	0.229938	0.000000	0.133247	0.136729	0.183092	0.316994
tcf-miR-n577	0.474740	0.000000	0.235807	0.188198	0.101255	0.000000
tcf-miR-n578	0.204954	0.303025	0.261292	0.108331	0.122398	0.000000
tcf-miR-n579	0.000000	0.000000	0.445456	0.217665	0.218604	0.118274
tcf-miR-n580	0.000000	0.000000	0.166432	0.227709	0.457381	0.148478
tcf-miR-n581	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
tcf-miR-n582	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
tcf-miR-n583	0.000000	0.000000	0.000000	0.175288	0.396099	0.428613
tcf-miR-n584	0.534466	0.000000	0.171537	0.146683	0.147315	0.000000
tcf-miR-n585	0.000000	0.292583	0.409969	0.107867	0.189582	0.000000
tcf-miR-n586	0.000000	0.000000	0.198128	0.225895	0.340304	0.235672
tcf-miR-n587	0.000000	0.309338	0.433447	0.114045	0.143170	0.000000
tcf-miR-n588	0.000000	0.000000	0.203119	0.115793	0.348876	0.332212
tcf-miR-n589	0.000000	0.380640	0.393862	0.000000	0.225498	0.000000
tcf-miR-n590	0.000000	0.464603	0.327778	0.207620	0.000000	0.000000
tcf-miR-n591	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
tcf-miR-n592	0.000000	0.352914	0.342350	0.000000	0.163339	0.141397
tcf-miR-n593	0.289768	0.311580	0.167919	0.000000	0.230733	0.000000



(Legend on next page)

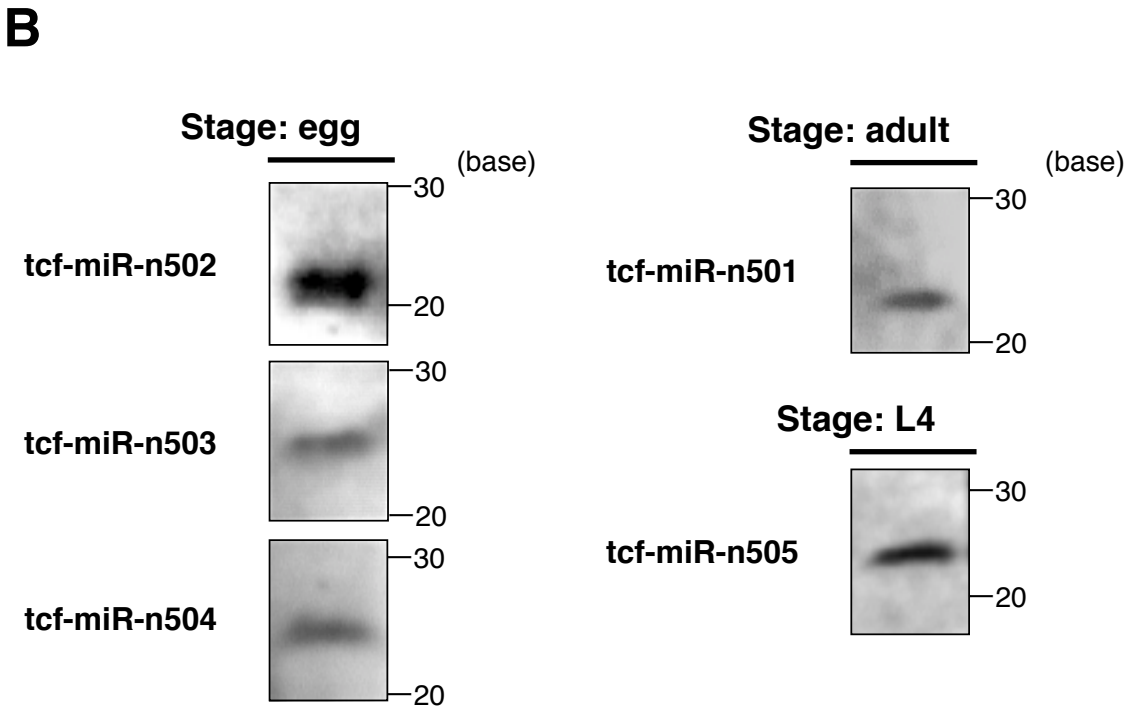


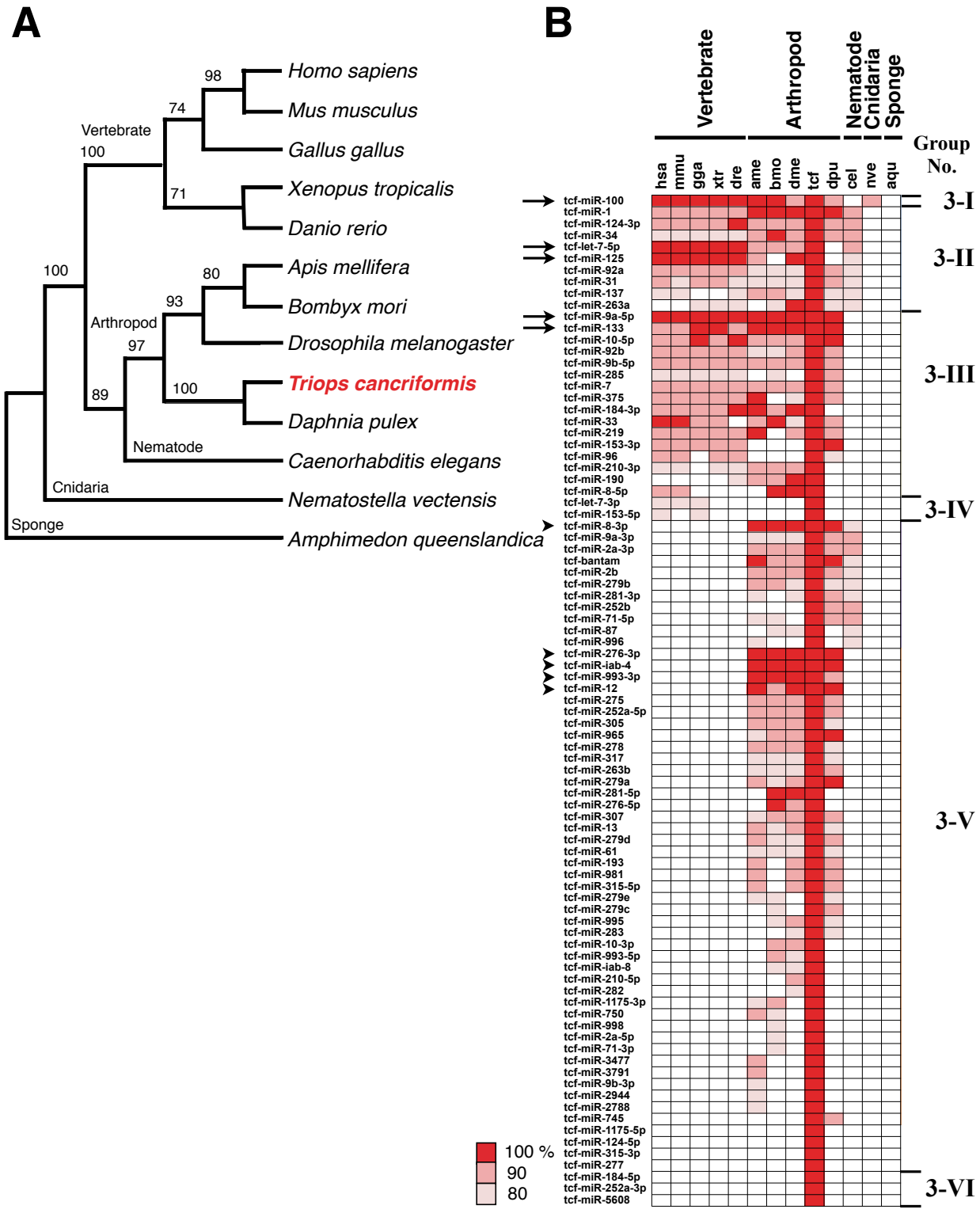
Figure 3.4 Expression of novel candidate *T. cancriformis* miRNAs

(A) Expression profiles of novel candidate miRNAs in the deep-sequencing analysis. miRNA reads were normalized to the spike reads in each stage. The more intense red color indicates a strongly expressed miRNA, whereas the more intense green color indicates a more weakly expressed miRNA. Novel candidate miRNAs were categorized into nine groups based on their expression patterns (Groups 2-I to 2-IX). Total read numbers for the six stages are listed. (B) Northern blotting analysis of the expression of five novel miRNAs.

2-V in Figure 3.4A), although each developmental stage of *T. cancriformis* is only short (0–0.5 h for 1st instar, 3–8 h for 2nd instar, 13–22 h for 3rd instar, and 26–37 h for 4th instar larvae), in accordance with the dramatic morphological changes in *T. cancriformis* larvae. To confirm the expression of these novel candidate miRNAs, a northern blotting analysis of eight of them (from tcf-miR-n501 to tcf-miR-n508) was performed. As a consequence, the expression of five of the eight miRNAs was detected in one stage that showed higher read counts for each candidate miRNA (egg stage for tcf-miR-n502, tcf-miR-n503, and tcf-miR-n504; 4th instar larval stage for tcf-miR-n505; and adult stage for tcf-miR-n501) (Figure 3.4B). These results show that at least these five novel miRNAs are actually expressed in the cells. Because the expression of these five miRNAs changed dynamically throughout the six developmental stages, they may be involved in stage-specific gene regulation. For instance, tcf-miR-n505 was increasingly expressed toward the 4th instar (Group 2-VI in Figure 3.4A), with high read counts (4,599 read counts) in 4th instar larvae (Table 3.4). Of the 93 novel candidate miRNAs, 10 were strongly expressed in 4th instar larvae (three in Group 2-IV, seven in Group 2-VI in Figure 3.4A). Several morphological characteristics of *T. cancriformis* change in the 4th instar larvae, when the compound eyes expand and the number of body segments increases. I speculate that some miRNAs that are strongly expressed in the 4th instar larvae, including tcf-miR-n505, are involved in these differentiation processes. To understand the exact functions of these novel miRNAs, an expression analysis at the tissue level must be performed in future work.

3.3.3 Evolution conservation analysis of *T. cancriformis* miRNA sequences and miRNA clusters

To clarify the evolutionary position of *T. cancriformis*, a phylogenetic tree was constructed based on the 18S rRNA sequences of *T. cancriformis* and 12 model species from a wide range of Metazoa. The phylogenetic analysis showed that *T. cancriformis* is closely related to *Daphnia pulex*, in the crustaceans (Figure 3.5A). The branching order of *T. cancriformis* was very similar to that on a phylogenetic tree constructed in a previous study (Bourlat et al. 2008). To investigate the evolution of the *T. cancriformis* miRNA sequences, I compared the 87 conserved *T. cancriformis* miRNAs found in this study with 7,634 miRNAs previously reported in 12 model species and registered in miRBase release 20.0. These conserved miRNAs were roughly classified into six groups based on their sequence conservation (Groups 3-I to 3-VI in Figure 3.5B), and 81 of the 87 *T. cancriformis* miRNAs shared sequence similarity (≧80%), based on seed matching, with those of arthropod species (Groups 3-I to 3-III, and 3-V in Figure 3.5B, and pink circle in Figure 3.5C), although the organism is called “living fossil”. Among these 81 miRNAs, 26 were also conserved in vertebrates (Groups 3-I to 3-III in Figure 3.5B, and the overlapping region between the pink and light blue circles in Figure 3.5C). Five of these 26 miRNAs (tcf-let-7-5p, tcf-miR-9a-5p, tcf-miR-125, tcf-miR-100, and tcf-miR-133) shared identical sequences with those of a wide range of bilateria (indicated with arrows in Figure 3.5B). In particular, tcf-miR-9a-5p had an identical sequence from organisms ranging in vertebrates to arthropods. Five miRNAs (tcf-miR-8-3p, tcf-miR-12, tcf-miR-276-3p, tcf-miR-993-3p, and tcf-miR-iab-4) also



(Legend on next page)

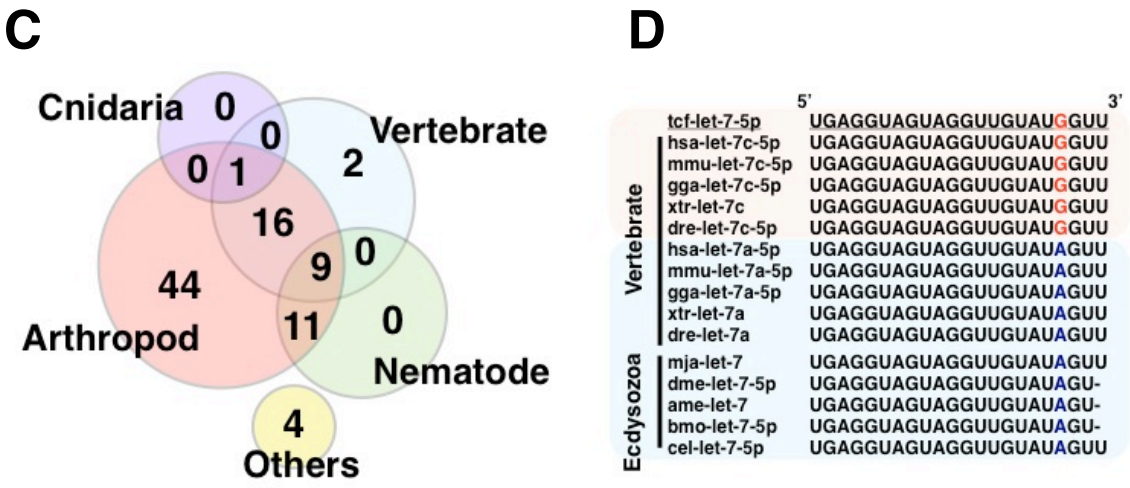


Figure 3.5 Evolution of conserved miRNAs found in *T. cancriformis*

(A) Phylogenetic tree of *T. cancriformis* and 12 metazoan animals constructed from 18S rRNA sequences. The evolutionary position of *T. cancriformis* is shown in red. Bootstrap support values are indicated near the branches. (B) Distributions of conserved miRNAs across metazoan animals. The presence of metazoan miRNAs that share sequence similarity (≧80%) and complete seed matches with *T. cancriformis* miRNAs is shown with colors. Based on miRNA conservation, 87 conserved miRNAs were classified into six groups (Groups 3-I to 3-VI). Arrows and arrowheads indicate the miRNAs highly conserved among bilaterians and arthropods, respectively. hsa *H. sapiens*, mmu *M. musculus*, gga *G. gallus*, xtr *X. tropicalis*, dre *D. rerio*, ame *A. mellifera*, bmo *B. mori*, dme *D. melanogaster*, tcf *T. cancriformis*, dpu *Daphnia pulex*, cel *C. elegans*, nve *N. vectensis*, aqu *A. queenslandica.* (C) Conserved miRNAs occurring in *T. cancriformis* and other metazoan animals. The number indicates the conserved *T. cancriformis* miRNA that share sequence similarity with miRNAs of vertebrate, arthropod, nematode, cnidaria, or other species. (D) Nucleotide sequence comparison of let-7 in bilaterian animals. A sequence alignment of bilaterian let-7 is shown. The colored 19th nucleotide, "G" or "A", shows the difference in the let-7 sequences of vertebrates and ecdysozoans. Underlining indicates the let-7 sequence of *T. cancriformis*.

shared identical sequences with various arthropods (indicated with arrowheads in Figure 3.5B).

According to the literature, 95 miRNAs of *Marsupenaeus japonicus*, which belongs to the Crustacea, have been reported (Ruan et al. 2011; Huang et al. 2012). Therefore, I comparatively analyzed the 180 *T. cancriformis* miRNAs (87 conserved miRNAs and 93 novel candidate miRNAs) and the 95 *M. japonicas* miRNAs. Based on seed matching, 55 of the 95 *M. japonicus* miRNAs shared sequence similarities (≧ 80%) with the *T. cancriformis* conserved miRNAs, whereas no *M. japonicus* miRNAs shared sequence similarity with the 93 novel *T. cancriformis* candidate miRNAs. Of these 55 miRNAs, 13 shared identical sequences and 44 shared ≧90% sequence similarity. The number of *M. japonicus* miRNAs that shared ≧90% sequence similarity with those of *T. cancriformis* was almost equivalent to the number it shares with other arthropods (from 36 to 42) (Figure 3.5B), suggesting that *M. japonicus* and *T. cancriformis* are phylogenetically related. Sequence alignments of these well-conserved miRNAs in *M. japonicus* and other model species identified interesting characteristics of tcf-let-7. Generally, vertebrates have multiple copies of let-7, whereas ecdysozoans have only a single copy, and ecdysozoans let-7 is identical in vertebrates let-7a, except at a single nucleotide, one base from the 3′ end. In *T. cancriformis*, a single copy of let-7 was detected that is identical to vertebrate let-7c, which differs from arthropod let-7 at the 19th nucleotide, which is altered from “A” to “G” (for dme-let-7-5p: UGAGGUAGUAGGUUGUAUAGU, whereas for hsa-let-7c-5p and tcf-let-7-5p: UGAGGUAGUAGGUUGUAUGGUU) (Figure 3.5D), suggesting that *T. cancriformis*

let-7 evolved in a unique way. Further analysis is required to clarify why the *T. cancriformis* let-7 sequence is more similar to those of the vertebrates than to those of the arthropods. How was this vertebrate-type let-7 obtained by *T. cancriformis*? I believe that both horizontal gene transfer and mutation were involved. Unfortunately, I can find no evidence to support the hypothesis that let-7 was obtained by *T. cancriformis* through horizontal gene transfer from another species. However, the mutation site is at the 19th nucleotide from the 5′ end of the miRNAs, so it is unlikely to affect the target mRNAs because this site is not in the seed region, which is important for miRNA target recognition. This inference must be confirmed by comparing the target mRNAs of miRNAs in which the nucleotide at this position is “A” or “G”. Therefore, this vertebrate-type sequence is considered to have been retained throughout the evolution of *T. cancriformis*. It should be possible to clarify the acquisition of the vertebrate-type let-7 by *T. cancriformis* by examining the small RNAs and genomic DNAs of closely related species, such as *Triops longicaudatus* and *Triops granarius*.

Next, I investigated whether the 93 novel *T. cancriformis* candidate miRNAs share sequence similarity with the genomic DNA sequences of related model species, such as *C. elegans*, *D. melanogaster*, and *Daphnia pulex*. Forty-eight of the 93 novel *T. cancriformis* candidate miRNAs share ≧80% sequence similarity, based on complete seed matching, with the genomic DNA sequences of these organisms (Figure 3.6A). Of these 48 candidates, two in *C. elegans*, one in *D. melanogaster*, and six in *Daphnia pulex* potentially form secondary structures typical of precursor miRNAs (Figure 3.6B) and meet reliability criteria (2), (4), and (5). In the *C. elegans* genomic DNA sequence,

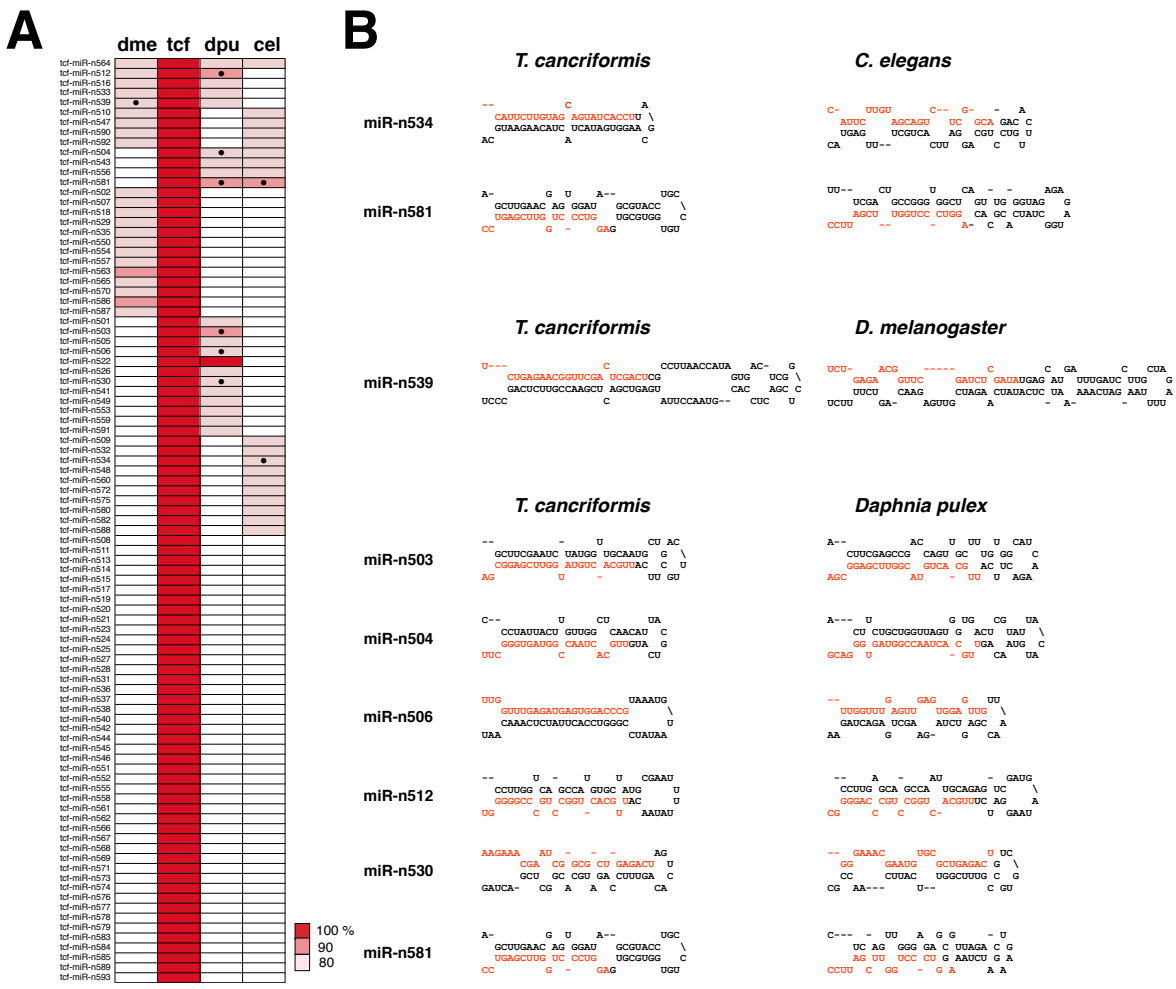


Figure 3.6 Conservation of the nucleotide sequences of novel *T. cancriformis* candidate miRNAs in ecdysozoan genomic DNA sequences

(A) The presence of *C. elegans, D. melanogaster*, and *Daphnia pulex* genomic sequences that share sequence similarity (≧80%) and a complete seed match with novel *T. cancriformis* candidate miRNAs is shown in red (80%–100%). (B) Possible secondary structures of miRNA precursors that share sequence similarity in *T. cancriformis* and *C. elegans*, *D. melanogaster*, or *Daphnia pulex*. Red indicates mature miRNA sequences.

two predicted miRNAs (cel-miR-n534 and cel-miR-n581), which are located in the exon region of *clec-41* and a tRNA pseudogene, respectively (data not shown), share sequence similarity with tcf-miR-n534 and tcf-miR-n581, respectively. It has been reported that most miRNAs are located in intergenic and intronic regions. However, exonic miRNAs located within the exons of host protein-coding genes are rarely reported (Wang 2010), suggesting that cel-miR-n534 and cel-miR-n581 are unlikely to be expressed as miRNAs in *C. elegans*. In contrast, dme-miR-n539, which shares sequence similarity with tcf-miR-n539, is located in the intronic region of *Toll-9* in the *D. melanogaster* genomic sequence (data not shown), suggesting that dme-miR-n539 is expressed as an miRNA in the fly. *Daphnia pulex*, which belongs to the crustaceans like *T. cancriformis*, has six candidate miRNAs that share sequence similarity with novel *T. cancriformis* miRNAs (Figure 3.6B). From a sequence alignment of these six miRNA sequences, three miRNAs (miR-n503, miR-n504, and miR-n512) are seen to have the same seed sequences (5′-UUGCACU-3′) in both *T. cancriformis* and *Daphnia pulex* (Figure 3.7A). Furthermore, these three miRNAs are closely located on genomic contigs (i.e., form an miRNA cluster) and are possibly encoded by the same primary miRNA (pri-miRNA) in both species (Figure 3.7B–C). tcf-miR-n503 and tcf-miR-n504 were also detected in this northern blotting analysis (Figure 3.4B) and most importantly, the expression of the miRNAs forming this miRNA cluster correlated significantly ($r > 0.98$, $p \leq 0.001$; Table 3.9). These data support the existence of an miRNA cluster in *T. cancriformis*.

I next checked whether another miRNA cluster exists in the *T. cancriformis*

Table 3.9 miRNA clusters in *T. cancriformis*

Cluster	Strand	Contig#	Start position	End position	Length (nt)	Expression correlation
miR-96/263b/263a	minus	contig_68-70816	44256	45925	1670	r > 0.93 **
miR-277/34	plus	contig_643-10482	320	1625	1306	r = 0.66
miR-2a-2/2a/13/2b/71	minus	contig_751-13972	11521	12303	783	r > 0.96 **
miR-12/3477/283	minus	contig_869-45389	28247	29019	773	r > 0.60
miR-n504/n512/n537/n503	plus	contig_1320-28899	25634	26273	640	r > 0.98 ***
miR-2788/193	minus	contig_1401-52320	26513	26775	263	r = 0.87 *
miR-305/275	minus	contig_1756-59422	31314	31547	234	r = 0.81 *
miR-61/279b	plus	contig_1838-20014	11507	11737	231	r = 0.53
miR-252a/252b	minus	contig_2308-38970	25406	27071	1666	r > 0.83 *
miR-87/n505	plus	contig_2981-25385	3101	3296	196	r = 0.90 *
miR-n509/995	plus	contig_3153-8869	5995	6309	315	r = 0.92 **
miR-92a/92b	plus	contig_4248-31752	30650	30887	238	r = 0.99 ***
let-7/miR-100	minus	contig_5716-15662	2819	3063	245	r > 0.81 *
miR-n527/n518/n502/n520/n535	minus	contig_8886-8311	1674	2310	637	r > 0.95 **
miR-3791/279c/n521/n511/998/2944/9b/279e	minus	contig_9912-3069	1103	3031	1929	r > -0.19
miR-1175/750	minus	contig_10104-5963	445	745	301	r > -0.02

Statistically significant differences: * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$.

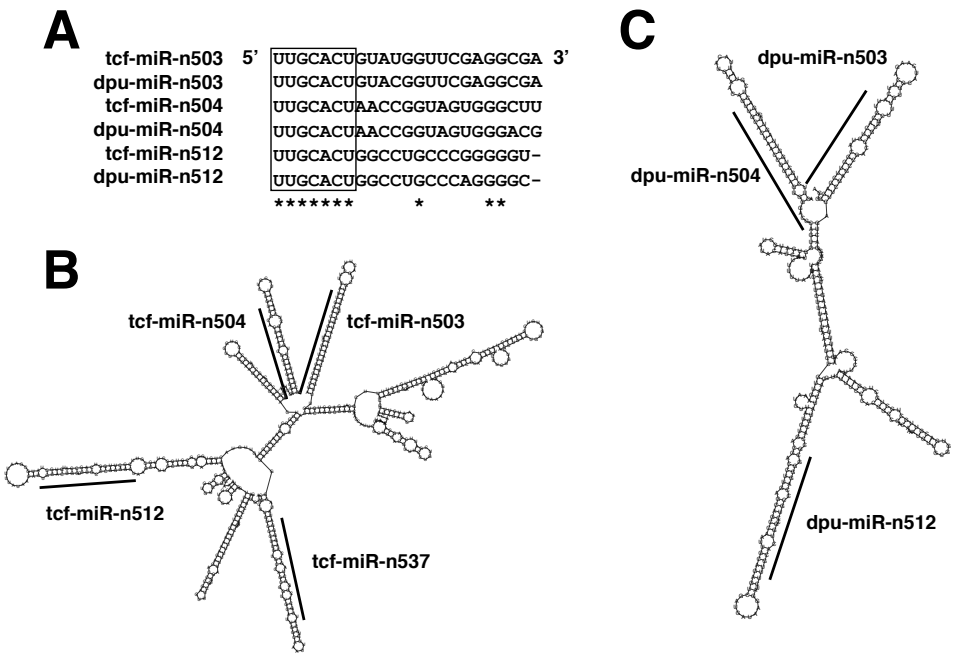


Figure 3.7 Novel *T. cancriformis* candidate miRNAs and their sequence conservation in *Daphnia pulex*

(A) Sequence alignment of miR-n503, miR-n504, and miR-n512 from *T. cancriformis* and *Daphnia pulex*. Asterisks indicate 100% conserved residue among the sequences. The common seed sequence regions (5′-UUGCACU-3′) are boxed. (B–C) Predicted pri-miRNAs consisting of at least miR-n503, miR-n504, and miR-n512 in the genomes of (B) *T. cancriformis* and (C) *Daphnia pulex*. Each line indicates an individual mature miRNA sequence.

draft genome, which was determined in this study. In *D. melanogaster*, several miRNAs encoded within 10 kilobase (kb) on the same genomic strand are defined as an miRNA cluster (Marco et al. 2013). Because there is currently no information about the transcriptional unit in *T. cancriformis*, I set a more stringent criterion to define an miRNA cluster in this organism: several miRNAs encoded within 2 kb on the same genome strand. This analysis showed that *T. cancriformis* had 16 miRNA clusters, consisting of 2–8 miRNAs (Table 3.9). It has been reported that 74 (31%) of all annotated miRNAs in the genome of *D. melanogaster* are clustered (Marco et al. 2013), whereas 48 (28%) of the annotated miRNAs in the draft genome of *T. cancriformis* are clustered, suggesting that the proportions of miRNAs forming clusters are almost the same in the two species. I also noted that the expression of miRNAs forming 11 of these clusters was significantly correlated ($p \leq 0.05$; Table 3.9). These results suggest that the miRNAs in each cluster are transcribed from the same transcriptional unit. Next, I compared the compositions of the *T. cancriformis* miRNA clusters with those of the arthropod *D. melanogaster* and *Tribolium castaneum* clusters (Marco et al. 2010; Marco et al. 2013). Of the 11 miRNAs clusters of *T. cancriformis*, the compositions of four (tcf-miR-2a-2/2a/13/2b/71, tcf-miR-305/275, tcf-miR-92a/92b, tcf-let-7/miR-100) were (partly) conserved relative to those of *D. melanogaster*, whereas the compositions of five clusters (tcf-miR-2a-2/2a/13/2b/71, tcf-miR-2788/193, tcf-miR-305/275, tcf-miR-92a/92b, and tcf-let-7/miR-100) were (partly) conserved relative to those of *Tribolium castaneum*.

3.3.4 Phylogenetic evolutionary analysis of *T. cancriformis* DICER and AGO family proteins

A bioinformatics analysis showed that *T. cancriformis* has only one DICER (*dicer1*), three AGO (*ago1–3*), one PIWI (*piwi*), and one AUB (*aubergine*) (Figures 3.8–3.9) (see Materials and Methods). These were predicted using DICER and AGO family proteins from *H. sapiens*, *D. melanogaster*, *M. japonicus*, and *Daphnia pulex* as the query sequences (see Materials and Methods). I refer to each protein as "tcf XXX" (e.g., tcf DICER1) hereafter. *Triops cancriformis dicer1* encodes the tcf DICER1 protein of 2,048 amino acids (a.a.), and I predicted five domains: DEAD (E-value = $1.51e^{-21}$), helicase C terminal (E-value = $3.00e^{-16}$), PAZ (E-value = $1.53e^{-24}$), RNase III 1 (E-value = $2.47e^{-23}$), and RNase III 2 (E-value = $1.45e^{-47}$) in tcf DICER1. The dsrm domain, which is the RNA-binding domain in the C-terminal region, was not predicted in my first prediction because the E-value was high. However, by taking into consideration the conservation of amino acid sequences, dsrm can be regarded a domain of tcf DICER1 (Figures 3.8 and 3.10). Tcf DICER1 shows 44.9% amino acid identity and 76.4% similarity with *Daphnia pulex* DICER1, 20.5% identity and 56.8% similarity with *D. melanogaster* DCR1, and 9.3% identity and 47.1% similarity with *D. melanogaster* DCR2. The five functional domains, excluding the DEAD domain, are particularly well conserved in DICER1 rather than in DICER2 among these related species (Figure 3.10).

To investigate the evolution of the tcf DICER1 protein, a phylogenetic tree was constructed using the amino acid sequences of DICER proteins from

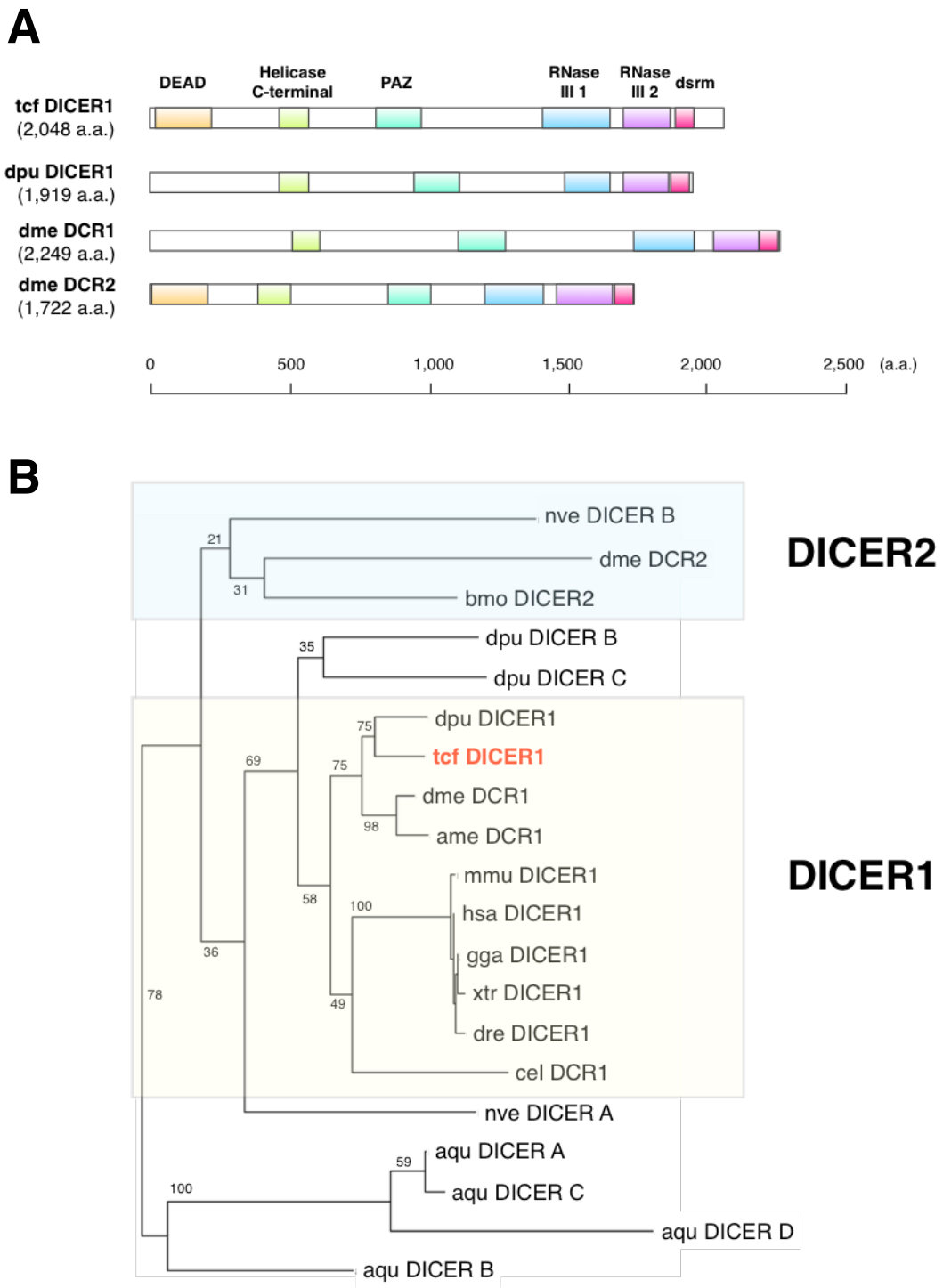


Figure 3.8 Evolution of *T. cancriformis* DICER protein

(A) Comparison of the domain structures of the DICER proteins of *T. cancriformis*, *Daphnia pulex*, and *D. melanogaster*. Each domain structure is based on UniProt and SMART predictions. (B) Phylogenetic tree of DICER proteins among several metazoan species. The evolutionary position of tcf DICER1 is shown in red. Bootstrap support values are indicated near the branches.

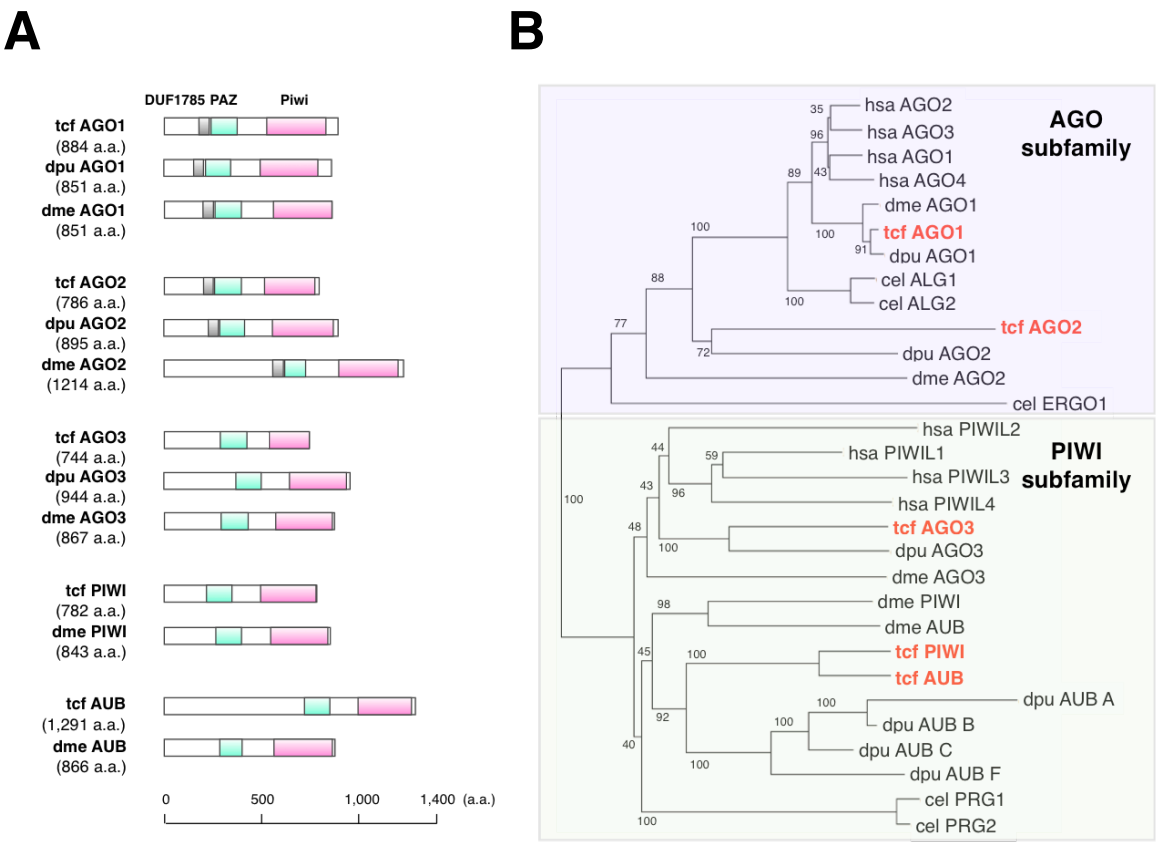


Figure 3.9 Evolution of *T. cancriformis* AGO family proteins

(A) Comparison of the domain structures of the AGO family protein in *T. cancriformis*, *Daphnia pulex*, and *D. melanogaster*. Each domain structure is based on UniProt and SMART predictions. (B) Phylogenetic tree of the AGO family proteins of several metazoan species. The evolutionary positions of the tcf AGO family proteins are shown in red. Bootstrap support values are indicated near the branches.

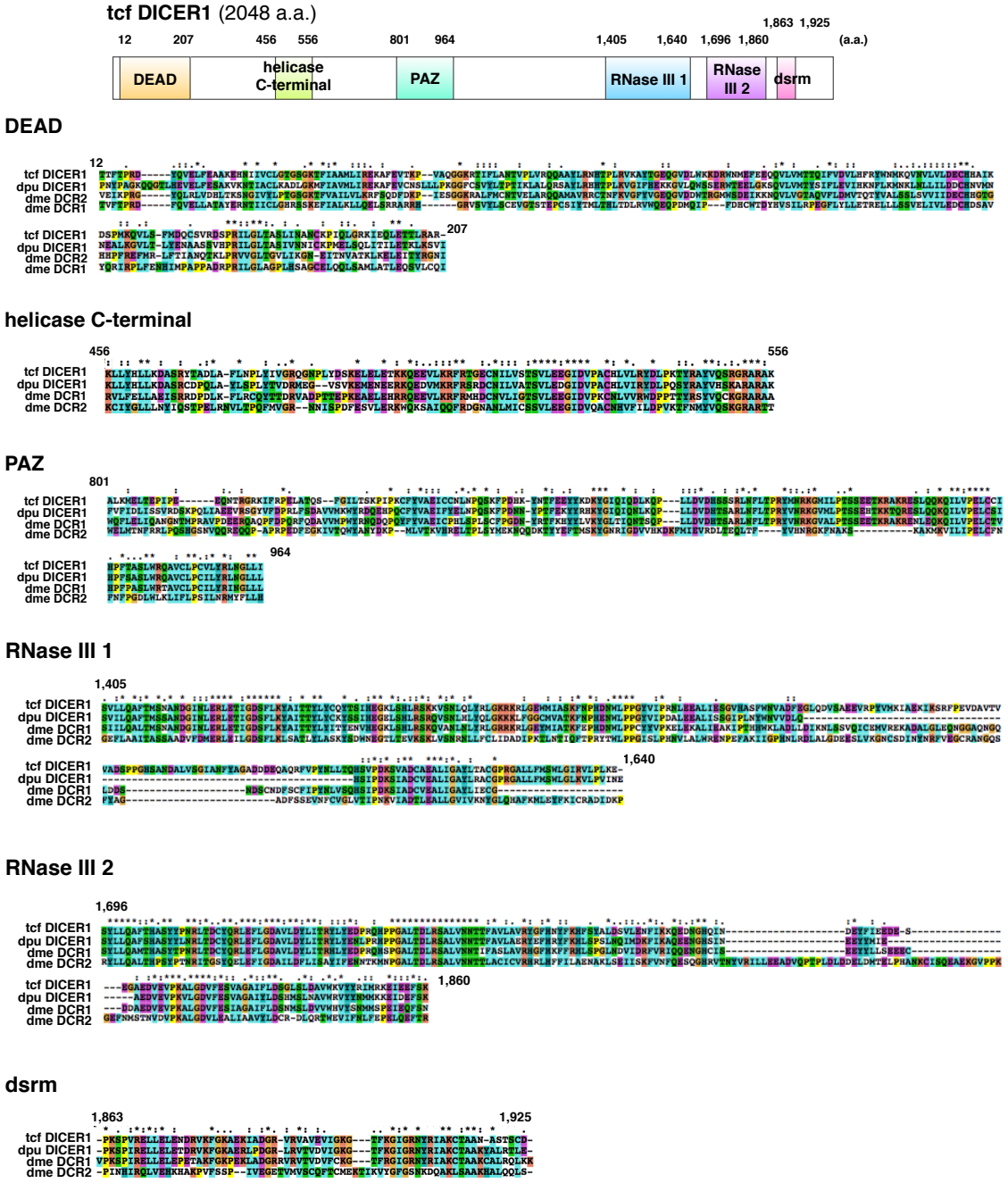


Figure 3.10 Amino acid sequence alignments of DICER domains

Amino acid sequence alignments of the DEAD, helicase C-terminal, PAZ, RNase III 1, RNase III 2, and dsrm domains of the DICER proteins in *T. cancriformis*, *Daphnia pulex*, and *D. melanogaster*. "*" identical sequences; ":" and "." similar amino acid properties. DEAD domain of dme DCR1 is not predicted by the SMART but added here for the amino acid sequence comparison.

T. cancriformis and several model species. On the phylogenetic tree, tcf DICER1 was positioned close to those of the crustacean *Daphnia pulex* (Figure 3.8B). It has been reported that most insects, including *D. melanogaster*, and some crustaceans express both DICER1 and DICER2 proteins (Su et al. 2008; Yao et al. 2010; Chen et al. 2011; Li et al. 2013; Mukherjee et al. 2013), and that most of those DICER1 proteins have lost the DEAD domain. In contrast, vertebrates and *C. elegans* have single DICER proteins (DICER1), containing the DEAD domain. This analysis showed that *T. cancriformis* has a single DICER1 protein containing the DEAD domain with the DECH motif (Figure 3.10), suggesting that this domain is functional. DEAD is the RNA helicase domain in the N-terminal region, and may be involved in the autoinhibition of the DICER activity (Ma et al. 2008). I found that tcf DICER1 domain is also more similar to the DICER1 in the vertebrates than to those in the arthropods like the let-7 miRNA sequence, suggesting that miRNA system of *T. cancriformis* evolved in a unique fashion. Because only the DICER1 protein was found in *T. cancriformis*, I looked for the DICER2 protein among the DICER proteins of other species, such as *B. mori* and *Litopenaeus vannamei* (see Materials and Methods). However, I found no DICER2 protein containing the PAZ, RNase III 1, and RNase III 2 domains encoded in the *T. cancriformis* genomic DNA contigs. However, there is still a slight possibility that *T. cancriformi*s has a DICER2 protein in which the domains differ considerably from those of previously identified DICER proteins, or that *T. cancriformi*s DICER2 will be found when more precise genomic sequence information becomes available, because the current data are draft genome sequences that still include many gaps. To confirm the

identity of DICER1, I checked the conservation of the 5′ pocket motif, which recognizes the 5′ phosphorylated end of RNAs for cleavage in humans (Park et al. 2011). The amino acid sequence of the 5′ pocket motif in tcf DICER1 is more similar to those of hsa DICER1, dme DCR1, and dpu DICER1 than to that of dme DCR2 (Figure 3.11), and four of the five 5′-interacting residues are conserved in tcf DICER1. Although it is still unclear whether all five interacting residues are required for the activity of the protein in *T. cancriformis*, it is likely that the DICER protein deduced from the *T. cancriformis* draft genome is DICER1. However, it is also possible that this domain has been swapped between *T. cancriformis* DICER1 and another as-yet-unidentified DICER protein, because the quality of the current draft genome is not high.

Most animals express DICER1 and/or DICER2. Therefore, I considered the evolution of DICER. Because most animals express DICER1, there may be two patterns of DICER2 acquisition, gain or loss. Most of the vertebrates and *C. elegans* express only one DICER1, suggesting that the animals originally had only a single DICER1, which then duplicated in the insects and several crustaceans. However, it has been reported that *Trichoplax adhaerens* and *Nematostella vectensis* express DICER2, suggesting that DICER duplicated early in animal evolution (Mukherjee et al. 2013). Because some species, including *Daphnia pulex* and *A. queenslandicas*, have more than three DICERs, it is possible that the DICER proteins are likely to generate diversity. Furthermore, because DICER2 is involved in the siRNA pathway in *D. melanogaster*, the evolution of the siRNA pathway must be clarified if we are to understand DICER

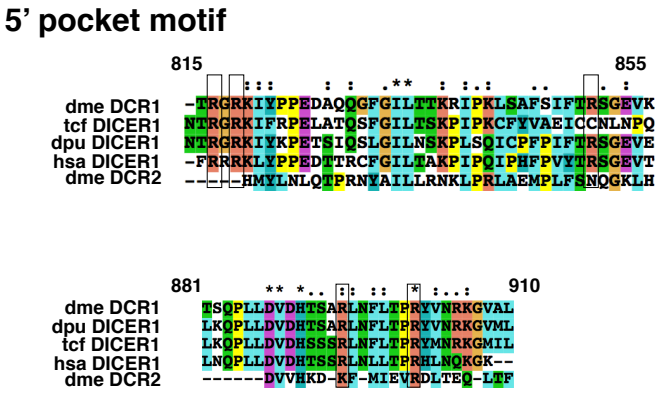
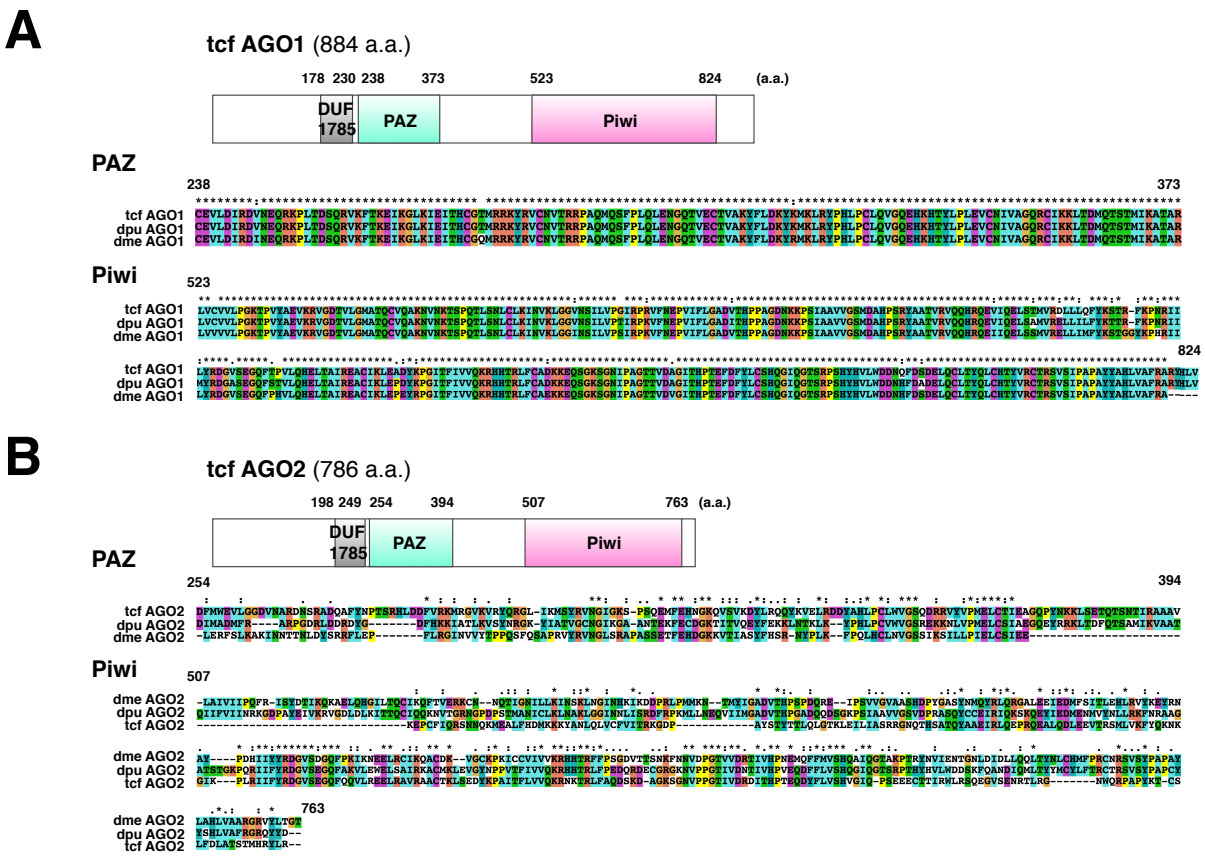


Figure 3.11 Amino acid sequence alignments of the 5′ pocket motif of DICER

Amino acid sequence alignments of the 5′ pocket region of DICER in *T. cancriformis*, *H. sapiens*, *Daphnia pulex*, and *D. melanogaster*. “*” identical sequences; “:” and “.” similar amino acid properties.

evolution.

The AGO family proteins play important roles in small-RNA-guided gene-silencing processes (Meister 2013), and are basically subdivided into the AGO subfamily and PIWI subfamily. In this study, I identified tcf AGO1 (884 a.a.) and tcf AGO2 (786 a.a.) of the AGO subfamily, and tcf AGO3 (744 a.a.), tcf PIWI (782 a.a.), and tcf AUB (1,291 a.a.) of the PIWI subfamily of *T. cancriformi*s. I predicted that all these tcf AGO family proteins contain PAZ and Piwi domains, known from the AGO family proteins of other species (Figures 3.9 and 3.12). Most of the PAZ and Piwi domains were predicted with E-values of $<1.00e^{-19}$, with the exception of the PAZ domains in AGO1 (E-value = $5.41e^{-3}$) and AGO2 (E-value = 0.46), indicating that the prediction of these AGO family proteins is reliable. It has been reported that the AGO3, PIWI, and AUB proteins are involved in piRNA biogenesis (Brennecke et al. 2007; Gunawardane et al. 2007). Although the presence of piRNAs in *T. cancriformis* has yet to be demonstrated, it is highly likely that *T. cancriformis* has a piRNA regulatory system. When I examined the sequence similarities in the AGO family proteins, the tcf AGO1 protein shared a high degree of similarity with those of *Daphnia pulex* and *D. melanogaster* (96.6% and 89.9% amino acid identity, respectively), and most of the amino acid sequences of the PAZ and Piwi domains are identical to the corresponding domains in these species (Figure 3.12). However, the tcf AGO2 protein showed 35.8% and 19.2% amino acid identity to the *Daphnia pulex* and *D. melanogaster* protein, respectively, and other tcf AGO family proteins showed 15.6%–46.3% identity with the proteins of these species. The *D. melanogaster* AGO proteins reportedly play distinct



(Legend on next page)

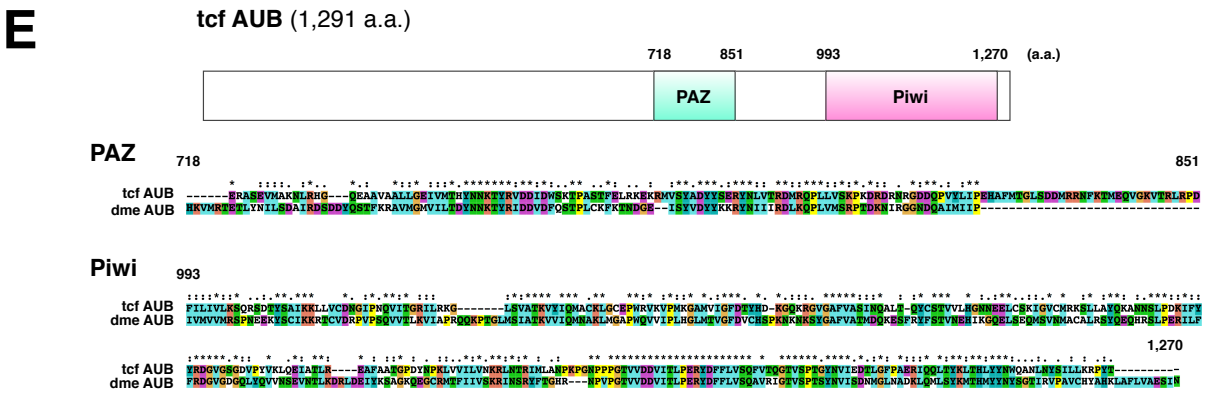


Figure 3.12 Amino acid sequence alignments of AGO domains

Amino acid sequence alignments of the PAZ and Piwi domains of (A) AGO1, (B) AGO2, and (C) AGO3 in *T. cancriformis*, *Daphnia pulex*, and *D. melanogaster*; and (D) PIWI and (E) AUB in *T. cancriformis* and *D. melanogaster*. "*" identical sequences; ":" and "." similar amino acid properties.

roles. For instance, AGO1 is involved in miRNA-directed RNA cleavage in RNAi, and AGO2 is involved in siRNA-directed RNAi (Okamura et al. 2004). In contrast, *H. sapiens* AGO1–4 are involved in miRNA biogenesis (Peters and Meister 2007). Phylogenetic analysis positioned *T. cancriformis* AGO1 close to AGO1 from *Daphnia pulex*, *M. japonicus*, *D. melanogaster* and AGO1–4 from *H. sapiens*, whereas the other AGO family proteins were diverged from the corresponding proteins in other species (Figure 3.9B). Because the miRNA regulatory system is largely conserved and plays important roles in eukaryotes, it is conceivable that the AGO1 proteins are under selective pressure to maintain their functional domains and their functions. In conclusion, throughout this chapter, I have shown the importance of non-model species, as well as model species, in the evolutionary analysis of miRNAs. I have also provided detailed sequence information previously unknown to the scientific community involved in RNA research. This study provides a foundation for further discussion of the evolution of miRNAs and the components of the RNAi machinery.

Chapter 4

Conclusions

4.1 miRNA evolution in bilaterian animals

MicroRNAs (miRNAs) are present in a wide range of species and evidence suggests that they have contributed to the evolution of animals (Niwa and Slack 2007; Prochnik et al. 2007; Grimson et al. 2008). To better understand miRNA evolution, I analyzed the evolution of miRNA/target-gene pairs in the bilaterian animals (Chapter 2), and specifically those of the living fossil *Triops cancriformis* (Chapter 3). Here, I describe the conclusions I reached in each chapter.

In Chapter 2, five highly conserved miRNAs were extracted from five bilaterian animal species (*Homo sapiens, Mus musculus, Gallus gallus, Drosophila melanogaster*, and *Caenorhabditis elegans*) (Takane et al. 2010). A procedure to extract potential evolutionarily conserved miRNA/target-gene pairs was developed based on orthologous gene information. This analysis yielded 31 evolutionarily conserved miRNA/target-gene pairs from 357,430 pairs. The downregulation of six candidate pairs (of the six tested pairs) was experimentally validated using HeLa cells, indicating the efficiency of this method. These findings demonstrate that miRNA target sites are conserved among various species and suggest that gene regulation mediated by let-7, miR-1, and miR-124, in particular, played important roles throughout evolution in processes such as development, differentiation, and muscle movement. These results also suggest that miRNA-mediated gene regulation and the tissue-specific expression of miRNA/target-gene pairs already existed in the common ancestor of the bilaterians. Therefore, this study provides new insight into miRNA functions in the early stages of

animal evolution.

In Chapter 3, I describe the identification of 180 miRNAs (87 conserved miRNAs and 93 novel candidate miRNAs) and the subsequent deduction of the components of the RNAi machinery: the DICER1, AGO1–3, PIWI, and AUB proteins (Ikeda et al. 2015). A comparative miRNA analysis of *T. cancriformis* and *D. melanogaster* showed inconsistencies in the expression patterns of four conserved miRNAs (e.g., *T. cancriformis* miR-34 is highly expressed in the egg, whereas *D. melanogaster* miR-34 is highly expressed in the adult). This suggests that although the miRNA sequences of the two species are very similar, their roles differ between the species. An analysis of the conservation of miRNA sequences revealed that most conserved *T. cancriformis* miRNAs share sequence similarities with those of the arthropods, although *T. cancriformis* is called a "living fossil". However, I found that let-7 sequence and domains of DICER1 in *T. cancriformis* are more similar to those of the vertebrates than to those of the arthropods, suggesting that the miRNA system of *T. cancriformis* evolved in a unique manner.

In Chapter 2, although it was possible to predict miRNAs and their target genes from worms to humans, the data suggest that the expression of miRNAs during development differs among species, even when the miRNAs and their target genes are conserved, as shown in Chapter 3. Together, these findings suggest that miRNA functions can change depending on the temporal expression patterns of the conserved miRNAs or mRNAs. These different expression patterns may also contribute to the differences among species. I also found that let-7 and DICER of *T. cancriformis* are

more similar to those of vertebrates than to those of arthropods, suggesting that the miRNA system of *T. cancriformis* has evolved uniquely. From this analysis, the importance of analyzing the miRNAs of both non-model species and model species is clear.

4.2 Future perspectives

To clarify miRNA evolution in the bilaterian animals, two analyses were performed using bioinformatics and experimental biology. The perspectives derived from these analyses are described below. In Chapter 2, miRNA targets were predicted using the entire set of known miRNA and 3′-untranslated region (UTR) data using an informatics approach, and the tissue-specific expression of miRNA/target-gene pairs was determined with a database search. However, when considering miRNA functions, stage specificity as well as tissue specificity is important. Understanding the stage-specific expression of miRNAs will contribute to a better understanding of their functions. The definition of an evolutionarily conserved miRNA was that it displayed >75% sequence similarity and complete seed matching in all five model species (*H. sapiens*, *M. musculus*, *G. gallus*, *D. melanogaster*, and *C. elegans*). However, my analysis in Chapter 3 revealed that nematode miRNAs have tended to be lost during evolution (Figure 3.5B). miRNA data are currently available for several species. Analysis of these data, particularly those of non-model species, may extend our understanding of the coevolution of miRNAs and their target mRNAs.

Chapter 3 provides detailed information on the miRNAs and the components of the RNAi machinery in *T. cancriformis* and its draft genomic DNA sequence. The data presented here indicate that the miRNA systems of *T. cancriformis* evolved in a unique manner. Therefore, it will be important to analyze the miRNAs and the components of the RNAi machinery in other non-model species to determine whether

other molecules are also conserved beyond species. At present, the quality of the draft *T. cancriformis* genome is not high because I sequenced it only once, and its size is unclear. Therefore, the unique evolution of the miRNA system in *T. cancriformis* that I propose here must be validated in future work. Sequencing the let-7 and *DICER* genes from the genomic DNA of *T. cancriformis* will establish both a reliable let-7 sequence and the organization of the DICER domain. Examining the genomic DNAs and small RNAs of closely related species, such as *Triops longicaudatus* and *Triops granaries*, will also clarify the evolution of let-7 and DICER. In my analysis, specific miRNAs were determined in six different developmental stages. However, the sizes of the individual organisms differed in each stage. To allow more reliable comparisons of miRNAs to be made, their levels in individual tissues during each stage of development must be determined. Understanding the functions of the stage-specific miRNAs of *T. cancriformis* will be interesting, because the morphology of *T. cancriformis* larvae changes dramatically in concert with changes in the expression of its miRNAs. Deciphering the roles of miRNAs in each developmental stage of *T. cancriformis* will require an analysis of the effects of miRNA overexpression and inhibition. However, it is currently difficult to knockdown gene expression in *T. cancriformis* because of technical problems. Therefore, to understand the miRNA functions in *T. cancriformis*, it will be easier to determine the miRNA targets *in silico* and *in vitro* especially with transfection assays in the cells of a related model species, such as *D. melanogaster*. Finally, in annotating the *T. cancriformis* miRNAs, I chose small RNA sequences with high read numbers as miRNAs. However, it is still possible that some miRNA

paralogues and novel miRNAs will be found in the future when the quality of the draft genome and small RNA sequences is improved.

The findings presented here suggest that the timing of the expression of miRNAs and their mRNA targets influences the differences that exist between species. Therefore, how differences in miRNA expression between species affect miRNA functions must be addressed in the future. Morphogenesis differs among species and a definition of the common developmental stages among species is yet to be established, but this definition will be essential. In conclusion, the findings described here contribute to our knowledge of the evolution of the miRNA systems and the roles of small RNAs in biology and evolution.

Acknowledgments

First, I would like to thank Professor Akio Kanai for teaching me the essence of science. He has always been scientifically critical and encouraging throughout my 10 years of campus life at Keio University. His jokes and the snacks that he gives me always make me smile. Professor Masaru Tomita has provided me with an excellent research environment at the Institute for Advanced Biosciences (IAB), where all the ideas for this work were conceived. Associate Professor Kazuharu Arakawa has given me valuable advice in pursuing my research and insightful advice in my role as an editor of IAB web publicity, with which I have been associated for six years. I extend my special thanks to my dissertation committee members, Professor Tomoyoshi Soga, Professor Mitsuhiro Itaya, Professor Mitsuhiro Watanabe, Associate Professor Yasuhiro Naito, and Associate Professor Hiroki Kuroda, for their valuable comments and suggestions regarding this dissertation.

I am deeply grateful to my research advisor, Dr. Kosuke Fujishima, who provided me with brilliant advice and was very enthusiastic about my research. I thank Kenji Igarashi for valuable advice on *Triops cancrifromis*; Kaoru Sugahara, Megumi Uetaki, Kalyn Kawamoto, Midori Kawasaki, and Emmy Umeda for their support with IAB web publicity; Yuka Hirose, Yuki Usui, and Yohei Chiwata for joining me in the long hours of study and laboratory work; Yuka Watanabe, Atsuko Shinhara, Hiromi Toyoshima, Nobuto Saito, Haruna Imamura, Kentaro Hayashi, Kiyofumi Hamashima, Shinnosuke Murakami, and Cornelia Amariei for their friendship, encouragement, and

helpful research discussions; Asako Soga, Emiko Noro, and Kiriko Hiraoka for their support with experiments; and Satomi Yokoi, Ayumi Mikami, Ayami Mizukami, Miwa Hiramoto, Kazune Sugahara, Aya Saito, and Maki Oike for their academic and administrative support.

Finally, I especially thank my dearly beloved family. My husband Yoshiki Ikeda supported me unconditionally throughout my PhD studies and offered suggestions of inestimable value. I thank him very much for being my partner for life. My wonderful and loving son Haruki Ikeda enriches my life with his endearing smiles, and I thank him for his unwavering moral support. I express my deep appreciation to my dear parents, Makoto Takane and Hiroko Takane, for their patience, faith, and love. I owe them so much for their overwhelming kindness and I thank them from the bottom of my heart.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF et al. 2000. The genome sequence of Drosophila melanogaster. *Science* **287**: 2185-2195.

Adlakha YK, Saini N. 2014. Brain microRNAs and insights into biological functions and therapeutic potential of brain enriched miRNA-128. *Mol Cancer* **13**: 33.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M et al. 2003. A uniform system for microRNA annotation. *RNA* **9**: 277-279.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215-233.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366.

Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC, Tassani S, Piva F et al. 2013. An estimation of the number of cells in the human body. *Ann Hum Biol* **40**: 463-471.

Bourlat SJ, Nielsen C, Economou AD, Telford MJ. 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol* **49**: 23-31.

Bousquet-Antonelli C, Deragon JM. 2009. A comprehensive analysis of the La-motif protein superfamily. *RNA* **15**: 750-764.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089-1103.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78-94.

Burgler C, Macdonald PM. 2005. Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. *BMC Genomics* **6**: 88.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309**: 1559-1563.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642-655.

Caygill EE, Johnston LA. 2008. Temporal regulation of metamorphic processes in Drosophila by the let-7 and miR-125 heterochronic microRNAs. *Curr Biol* **18**: 943-950.

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**: 584-589.

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL,

Wang DZ. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* **38**: 228-233.

Chen X, Li Q, Wang J, Guo X, Jiang X, Ren Z, Weng C, Sun G, Wang X, Liu Y et al. 2009. Identification and characterization of novel amphioxus microRNAs by Solexa sequencing. *Genome Biol* **10**: R78.

Chen YH, Jia XT, Zhao L, Li CZ, Zhang S, Chen YG, Weng SP, He JG. 2011. Identification and functional characterization of Dicer2 and five single VWC domain proteins of Litopenaeus vannamei. *Dev Comp Immunol* **35**: 661-671.

Choy EY, Siu KL, Kok KH, Lung RW, Tsang CM, To KF, Kwong DL, Tsao SW, Jin DY. 2008. An Epstein-Barr virus-encoded microRNA targets PUMA to promote host cell survival. *J Exp Med* **205**: 2551-2560.

Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES. 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* **104**: 19428-19433.

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F et al. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* **38**: 813-818.

Crick FH. 1958. On protein synthesis. *Symp Soc Exp Biol* **12**: 138-163.

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-235.

Dieci G, Preti M, Montanini B. 2009. Eukaryotic snoRNAs: a paradigm for gene

expression flexibility. *Genomics* **94**: 83-88.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**: 14863-14868.

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* **23**: 5866-5878.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806-811.

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

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92-105.

Friggi-Grelin F, Lavenant-Staccini L, Therond P. 2008. Control of antagonistic components of the hedgehog signaling pathway by microRNAs in Drosophila. *Genetics* **179**: 429-439.

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC*

Bioinformatics **8**: 69.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221-224.

Griffiths-Jones S. 2004. The microRNA Registry. *Nucleic Acids Res* **32**: D109-111.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154-158.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91-105.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193-1197.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**: 23-34.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5'

end formation in Drosophila. *Science* **315**: 1587-1590.

Haag J, Aigner T. 2007. Identification of calponin 3 as a novel Smad-binding modulator of BMP signaling expressed in cartilage. *Exp Cell Res* **313**: 3386-3394.

He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D et al. 2007. A microRNA component of the p53 tumour suppressor network. *Nature* **447**: 1130-1134.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* **435**: 828-833.

Hillier LW, Coulson A, Murray JI, Bao Z, Sulston JE, Waterston RH. 2005. Genomics in C. elegans: so many genes, such a little worm. *Genome Res* **15**: 1651-1660.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**: 3429-3431.

Huang T, Xu D, Zhang X. 2012. Characterization of host microRNAs that respond to DNA virus infection in a crustacean. *BMC Genomics* **13**: 159.

Huang Y, Gu X. 2007. A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs. *BMC Genomics* **8**: 66.

Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834-838.

Igarashi K. 1971. Ecological Studies on Triops longicaudatus (LE CONTE) Inhabiting Shonai District, Japan. (III) Some Features of the Young Individuals. *Journal of*

the Yamagata Agriculture and Forestry Society **28**: 4-5.

Ikeda KT, Hirose Y, Hiraoka K, Noro E, Fujishima K, Tomita M, Kanai A. 2015. Identification, expression, and molecular evolution of microRNAs in the "living fossil" Triops cancriformis (tadpole shrimp). *RNA* **21**: 230-242.

International Wheat Genome Sequencing C. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human MicroRNA targets. *PLoS Biol* **2**: e363.

Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D, Wilson M, Wang X, Shelton J, Shingara J et al. 2007. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res* **67**: 7713-7722.

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ. 2005. RAS is regulated by the let-7 microRNA family. *Cell* **120**: 635-647.

Karijolich J, Yu YT. 2010. Spliceosomal snRNA modifications and their function. *RNA Biol* **7**: 192-204.

Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, Hammond M, Rocca-Serra P, Cox T, Birney E. 2004. EnsMart: a generic system for fast and flexible access to biological data. *Genome Res* **14**: 160-169.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**: 1278-1284.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**: 126-139.

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A. 2004. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* **18**: 1165-1178.

Kloosterman WP, Plasterk RH. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**: 441-450.

Korn M, Rabet N, Ghate HV, Marrone F, Hundsdoerfer AK. 2013. Molecular phylogeny of the Notostraca. *Mol Phylogenet Evol* **69**: 1159-1171.

Kozomara A, Griffiths-Jones S. 2013. miRBase: annotating high confidence microRNAs using data. *Nucleic Acids Res* **42**: D68-D73.

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**: 495-500.

Kurihara Y, Takashi Y, Watanabe Y. 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**: 206-212.

Ladewig E, Okamura K, Flynt AS, Westholm JO, Lai EC. 2012. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res* **22**: 1634-1645.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-858.

Lai EC. 2005. miRNAs: whys and wherefores of miRNA-mediated regulation. *Curr*

Biol **15**: R458-460.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858-862.

Leaman D, Chen PY, Fak J, Yalcin A, Pearce M, Unnerstall U, Marks DS, Sander C, Tuschl T, Gaul U. 2005. Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. *Cell* **121**: 1097-1108.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-854.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415-419.

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**: 4051-4060.

Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40 (D1)**: D302-305.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787-798.

Li X, Yang L, Jiang S, Fu M, Huang J, Jiang S. 2013. Identification and expression analysis of Dicer2 in black tiger shrimp (Penaeus monodon) responses to immune challenges. *Fish Shellfish Immunol* **35**: 1-8.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769-773.

Liu J, Bang AG, Kintner C, Orth AP, Chanda SK, Ding S, Schultz PG. 2005. Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis. *Proc Natl Acad Sci U S A* **102**: 1927-1932.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* **303**: 95-98.

Ma E, MacRae IJ, Kirsch JF, Doudna JA. 2008. Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* **380**: 237-243.

Marco A, Hui JH, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. *Genome Biol Evol* **2**: 686-696.

Marco A, Ninova M, Ronshaugen M, Griffiths-Jones S. 2013. Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic Acids Res* **41**: 7745-7752.

Mathers TC, Hammond RL, Jenner RA, Hanfling B, Gomez A. 2013. Multiple global radiations in tadpole shrimps challenge the concept of 'living fossils'. *PeerJ* **1**:

e62.

Mattick JS, Makunin IV. 2006. Non-coding RNA. *Hum Mol Genet* **15 (suppl 1)**: R17-R29.

Meister G. 2013. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**: 447-459.

Moss EG, Lee RC, Ambros V. 1997. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**: 637-646.

Mukherjee K, Campos H, Kolaczkowski B. 2013. Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol Biol Evol* **30**: 627-641.

Nahvi A, Shoemaker CJ, Green R. 2009. An expanded seed sequence definition accounts for full regulation of the hid 3' UTR by bantam miRNA. *RNA* **15**: 814-822.

Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. *Curr Opin Genet Dev* **17**: 145-150.

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**: 1655-1666.

Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. 2009. The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res* **37**: D155-158.

Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. 2011. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**: 201-205.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86-89.

Peters L, Meister G. 2007. Argonaute proteins: mediators of RNA silencing. *Molecular Cell* **26**: 611-623.

Pfeffer S, Lagos-Quintana M, Tuschl T. 2005. Cloning of small RNA molecules. *Curr Protoc Mol Biol* **Chapter 26**: Unit 26.24.

Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C et al. 2004. Identification of virus-encoded microRNAs. *Science* **304**: 734-736.

Prochnik SE, Rokhsar DS, Aboobaker AA. 2007. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev Genes Evol* **217**: 73-77.

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**: 1507-1517.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-906.

Rick M, Ramos Garrido SI, Herr C, Thal DR, Noegel AA, Clemen CS. 2005. Nuclear localization of Annexin A7 during murine brain development. *BMC Neurosci* **6**:

25.

Rother S, Meister G. 2011. Small RNAs derived from longer non-coding RNAs. *Biochimie* **93**: 1905-1915.

Ruan L, Bian X, Ji Y, Li M, Li F, Yan X. 2011. Isolation and identification of novel microRNAs from Marsupenaeus japonicus. *Fish Shellfish Immunol* **31**: 334-340.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**: 83-86.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

Schurko AM, Logsdon JM, Jr., Eads BD. 2009. Meiosis genes in Daphnia pulex and the role of parthenogenesis in genome evolution. *BMC Evol Biol* **9**: 78.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58-63.

Sempere LF, Sokol NS, Dubrovsky EB, Berger EM, Ambros V. 2003. Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. *Dev Biol* **259**: 9-18.

Shields JM, Rogers-Graham K, Der CJ. 2002. Loss of transgelin in breast and colon tumors and in RIE-1 cells by Ras deregulation of gene expression through Raf-independent pathways. *J Biol Chem* **277**: 9790-9799.

Shivdasani RA. 2006. MicroRNAs: regulators of gene expression and cell differentiation. *Blood* **108**: 3646-3653.

Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246-258.

Smyth N, Vatansever HS, Murray P, Meyer M, Frie C, Paulsson M, Edgar D. 1999. Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation. *J Cell Biol* **144**: 151-160.

Stanke M, Tzvetkova A, Morgenstern B. 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* **7 (Suppl 1)**: S11.

Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M. 2007. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. *Genome Res* **17**: 1865-1879.

Su J, Oanh DT, Lyons RE, Leeton L, van Hulten MC, Tan SH, Song L, Rajendran KV, Walker PJ. 2008. A key gene of the RNA interference pathway in the black tiger shrimp, Penaeus monodon: identification and functional characterisation of Dicer-1. *Fish Shellfish Immunol* **24**: 223-233.

Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* **29**: 288-299.

Takahashi F. 1975. Effect of Light on the Hatching of Eggs in Triops granarius (Notostraca : Triopsidae). *Environ control biol* **13**: 29-33.

Takane K, Fujishima K, Watanabe Y, Sato A, Saito N, Tomita M, Kanai A. 2010. Computational prediction and experimental validation of evolutionarily

conserved microRNA target genes in bilaterian animals. *BMC Genomics* **11**: 101.

Takane K, Kanai A. 2011. Vertebrate virus-encoded microRNAs and their sequence conservation. *Jpn J Infect Dis* **64**: 357-366.

Thadani R, Tammi MT. 2006. MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics* **7 (Suppl 5)**: S20.

Toyofuku T, Zhang H, Kumanogoh A, Takegahara N, Suto F, Kamei J, Aoki K, Yabuki M, Hori M, Fujisawa H et al. 2004. Dual roles of Sema6D in cardiac morphogenesis through region-specific association of its receptor, Plexin-A1, with off-track and vascular endothelial growth factor receptor type 2. *Genes Dev* **18**: 435-447.

Vaucheret H, Vazquez F, Crete P, Bartel DP. 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev* **18**: 1187-1197.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**: 669-687.

Waidner LA, Burnside J, Anderson AS, Bernberg EL, German MA, Meyers BC, Green PJ, Morgan RW. 2011. A microRNA of infectious laryngotracheitis virus can downregulate and direct cleavage of ICP4 mRNA. *Virology* **411**: 25-31.

Wang J, Czech B, Crunk A, Wallace A, Mitreva M, Hannon GJ, Davis RE. 2011. Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. *Genome Res* **21**:

1462-1477.

Wang Z. 2010. MicroRNA: A matter of life or death. *World J Biol Chem* **1**: 41-54.

Watanabe Y, Tomita M, Kanai A. 2007. Computational methods for microRNA target prediction. *Methods Enzymol* **427**: 65-86.

Watanabe Y, Tomita M, Kanai A. 2008. Perspective in the evolution of human microRNAs: copy number expansion and acquisition of target gene specialization. *Prog theor phys Suppl* **173**: 219-228.

Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**: 737-738.

Wei C, Salichos L, Wittgrove CM, Rokas A, Patton JG. 2012. Transcriptome-wide analysis of small RNA expression in early zebrafish development. *RNA* **18**: 915-929.

Wienholds E, Plasterk RH. 2005. MicroRNA function in animal development. *FEBS Lett* **579**: 5911-5922.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-862.

Wu L, Belasco JG. 2005. Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol Cell Biol* **25**: 9198-9208.

Yao X, Wang L, Song L, Zhang H, Dong C, Zhang Y, Qiu L, Shi Y, Zhao J, Bi Y. 2010. A Dicer-1 gene from white shrimp Litopenaeus vannamei: expression pattern in

the processes of immune response and larval development. *Fish Shellfish Immunol* **29**: 565-570.

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**: 932-935.

Yu JY, Chung KH, Deo M, Thompson RC, Turner DL. 2008. MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation. *Exp Cell Res* **314**: 2618-2633.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821-829.