Summer 6-10-2022

CHARACTERIZING ENDOGENOUS DICER PRODUCTS TO UNRAVEL NOVEL RNAI BIOGENESIS PATHWAYS

Jacob Oche Peter

CHARACTERIZING ENDOGENOUS DICER PRODUCTS TO UNRAVEL NOVEL RNAI BIOGENESIS PATHWAYS

by

Jacob Oche Peter

A Dissertation
Submitted to the Graduate School,
the College of Arts and Sciences
and the School of Biological, Environmental, and Earth Sciences
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved by:

Dr. Alex Sutton Flynt, Committee Chair
Dr. Glenmore Shearer
Dr. Chaoyang Zhang
Dr. Dmitri Mavrodi
Dr. Shahid Karim

August 2022

Published by the Graduate School

ABSTRACT

RNA interference (RNAi) is a pervasive gene regulatory mechanism in eukaryotes based on the action of multiple classes of small RNA (sRNA). Exploiting RNAi pathways in non-model systems have great potential for creating potent RNAi technologies. Here, we accessed RNAi-mediated control of gene expression in the two-spotted spider mite, *Tetranychus urticae* (*T. urticae*) using engineered dsRNA designed to modulate the host RNAi pathway and increase RNAi efficacy. Analysis of Dicer (Dcr) generated fragments revealed how exogenous RNAs access the host RNAi pathway in this animal, opening avenues for designing RNAi technology for their control. Further, some organisms incorporate RNA-dependent RNA polymerases (RdRPs) activity into RNAi pathways either through precursor amplification or by directly generating sRNAs from Dcr-generated fragments. While this enzymatic activity has prominent roles in plants and fungi, the involvement of RdRP in RNAi pathways of many animals is controversial. In this work, we investigated the contribution of RdRP to sRNA biogenesis in *Branchiostoma floridae* (*B. floridae*), as well as their potential role in the RNAi pathways of *Crassostrea gigas* (*C. gigas*) challenged with oyster herpes virus type 1 (OsHV-1). With adequate RdRP transcripts expressed in a variety of organs and tissues and at different stages of development in these animals, their translation is highly likely. Through inhibiting RNA Pol II's activity and analyzing potentially RdRP synthesized nascent RNAs, we detected RdRP activity restricted to a few loci. Using this approach, we implicated RdRPs in playing a critical role in somatic genome protection through transposon repression. Although we failed to recover a wide coverage of RdRP activity in Amphioxus genome. Our observation suggests a scenario where RdRP function has become marginal, consistent with their loss in several animal lineages such as vertebrates. These explain the challenges of detecting a role in the creation of dsRNAs. This observation shows a framework for how an essential gene regulation mechanism slowly decays on its way to being lost.

ACKNOWLEDGMENTS

I would like to acknowledge my Lord and Savior Jesus Christ, who has given me this rare privilege to attain this height in my academic career. Further, I would like to thank my mother, Mrs. Rosemary Oche, who has not only been a pillar for me all these years but also a constant source of strength and unquantifiable love and compassion. Also, my heart is filled with unmeasurable gratitude for my advisor, Dr. Alex Flynt, who gave me the opportunity to become a scientist and has over the years become a great role model whose style and patience are worthy of emulation. Lastly, I would like to acknowledge my siblings, Dr. Isaac Akogwu, Emmanuel Oche, Maryann Peter James, Samuel Akogwu, and most especially, Daniel Oche Peter.

DEDICATION

I dedicate my dissertation work to the most important people in my life, my family who have been my biggest supporter during my academic journey. Also, in memory of my late father, Peter Obiabo Akogwu, I say a special feeling of thank you for bringing me into this beautiful planet we call home. You are gone but not forgotten.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF SCHEMES

LIST OF ABBREVIATIONS

AGO	Argonaute
ATP	Adenosine triphosphate
Aub	Aubergine
CHS	Chimeric Chalcone Synthase
DGCR8	DiGeorge syndrome critical region 8
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra acetic acid
ENA	European nucleotide archive
ERD	Eukaryotic RdRP domain
FPKM	Fragment per kilobase million
FSW	Filtered salt water
FTCD	Formimidoyl-transferase-cyclodeaminase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GTR	General time reversible
GWAS	Genome Wide Association Study
HE	High expressed
HEN1	Hua Enhancer factor 1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid

HIV	Human immune deficiency virus
HLB	Huanglobing
IGV	Integrative genome viewer
IPTG	Isopropyl β-d-1-thiogalactopyranoside
KOA	Potassium acetate
KOH	Potassium hydroxide
LINE	Long interspaced nuclear elements
LTR	Long tandem repeat
MEGA	Molecular Evolutionary Genetics Analysis
miRNA	MicroRNA
mRNA	messenger RNA
NAT	Natural antisense
NCBI	National center for biotechnology information
NGS	Next generation sequencing
PAZ	PIWI AGO Zwille
piRNA	PIWI interacting RNA
PIWI	P-element Induced Wimpy testis
PMSF	Phenylmethylsulfonyl fluoride
PTGS	Post transcriptional gene silencing
QDE-1	Quelling dependent enzyme
RanGTPase	Ran-related nuclear protein guanosine triphosphatase

RdRP	RNA dependent RNA polymerase
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RPM	Read per million
RT-PCR	Reverse transcription PCR
SDE	Silencing defective
SGS2	Suppressor of gene silencing
SINE	Short interspaced nuclear elements
siRNA	Short interfering RNA
SRA	Sequence read archive
sRNA	small RNA
ssRNA	Single stranded RNA
TE	Transposable elements
TM	Trademark
USA	United States of America
UTR	Untranslated region
WAGO	Worm AGO
WB	Whole body
WCR	Western corn rootworm

CHAPTER I - INTRODUCTION

1.1 RNA interference and discovery

RNA interference (RNAi) is a pervasive gene regulation mechanism brought about by complementary base pairing of transcripts and small ribonucleic acid (RNA) effector molecules leading to post-transcriptional gene silencing (PTGS). The discovery of this phenomenon has indeed revolutionized our understanding of gene function through reverse genetics.

Through an accidental discovery, Napoli, and Lemieux (1990) found that introduction of chalcone synthase (CHS) gene in pigmented petunia petals resulted in a 50-fold reduction in CHS messenger RNA (mRNA) levels when compared with wild type through what the authors described as an "unclear mechanism of reversible trans homologous gene suppression" (Napoli et al., 1990). Two years later, a similar study was reported for *Neurospora crassa* (*N. crassa*) by Romano and Macino where they termed the phenomenon, "quelling". The duo had observed a progressive and reversible transcriptional inhibition of the endogenous genes, albino-3 (al-3) and albino-1 (al-1) when the fungus is transformed with an exogenous homologous sequence (Romano & Macino, 1992). These findings preceded Ambros and Ruvkun's 1993 discovery of the first microRNA (miRNA) Lin-4, an ~22 nucleotide (nt) single stranded RNA (ssRNA), required for embryonic patterning in *Caenorhabditis elegans* (*C. elegans*). The very first evidence of an RNAi-like phenomena would be reported 3 years later by Guo and Kemphues through intentional use of sense/antisense RNA (around 4kb) to induce gene knockdown of the par-1 gene in *C. elegans* (Guo & Kemphues, 1995). Interestingly, this same study revealed that sense RNA designed against par-1 was also able to induce knockdown of the transcripts

increasing researchers' curiosity about the behavior of the Watson crick base paring and what we understood about the central dogma.

These earlier works created the foundation upon which the discovery of RNAi was built. After Guo and Kempheus's discovery, Fire and Mello took to the lab to attempt to understand and possible provide explanations about Guo's result. In 1998, Fire and Mello published results of their findings. Not only did the results provide excellent answer for the strange behavior of sense/antisense RNAs ability to induce gene knock down as reported by Guo and Kempheus, but also provided new insight into the role of double stranded RNA (dsRNA) as potent molecules for eliciting RNAi. Through extensive purification of sense and antisense ssRNA against unc-22, Fire and Mello discovered that individual ssRNA elicit gene knockdown 10 – 100-fold less compared to when sense/antisense ssRNAs are injected into the same worm (Fire et al., 1998). Through this finding, it was proposed that the sense RNAi response observed by Guo and his fellow researchers was possibly the result of dsRNA contamination in the preparation resulting from phage RNA polymerase which elicited the RNAi response. This work provided the framework for understanding the concept of RNAi and the role of "dsRNA" in this mechanism, earning the duo a Nobel prize in physiology and medicine in 2006.

Following miRNA discovery in 1993 and the discovery of dsRNA five years later, various small non-coding RNA (ncRNA) molecules capable of entering host RNAi pathways were revealed in different eukaryotic systems. Three major classes of small RNAs (sRNAs) resulting from non-coding regions of the genome are discussed in this chapter including miRNA, short interfering RNA (siRNA), and P-element Induced WIMPI testis (PIWI) interacting RNA (piRNA).

1.2 Major Classes of Small RNA

PTGS through RNAi (RNAi) is mediated by sRNAs. These ncRNAs act through complementary base pairing with target RNAs, triggering mRNA destruction or translational inhibition as well as epigenetic regulation (Shirayama et al., 2012). Studies involving diverse sRNA (sRNA) species have increased tremendously since the emergence and continued advancement of next generation sequencing (NGS) technologies which offers a cheaper and more efficient alternative for high throughput biomolecule assays. As a result, corresponding efforts in computational approaches to distinguish sRNA classes have become well-developed. Although, hundreds of sRNAs have been identified in animals, different high-throughput sequencing and bioinformatics tools repeatedly reveal the existence of new ones.

1.2.1 miRNA biogenesis and function

miRNAs are conserved sRNA molecules first discovered 1993 by Ambros and Ruvkun groups through the groundbreaking discovery of *C. elegans* lin-4. They found that lin-4 sequence originated from and controls the lin-14 gene through a complementary antisense RNA-RNA based interaction with lin-14's 3' untranslated region (UTR) (R. C. Lee et al., 1993a). This discovery happened around the same time quelling and earlier RNAi mechanisms were discovered. Following this discovery, lin-4 was identified to play a significant role in temporal pattern formation in *C. elegans,* directly implicating the 22-nt RNA molecule in embryonic development (R. C. Lee et al., 1993b). To date, miRNAs have been studied actively and their importance in normal cellular maintenance or animal development cannot be over emphasized. As a result, aberrant expression of these short oligos have been linked with a variety of human diseases including cancers like chronic

lymphocytic leukemia associated with miR15 and miR16 gene deletion, lung cancer and Parkinson's disease also linked to let-7, immune related disease like type II diabetes and viral diseases like influenza virus and Human immune deficiency virus (HIV-1) have all been linked to several miRNA dysregulation (Calin et al., 2002; J. Huang et al., 2007; Y. Li & Kowdley, 2012; Löfgren et al., 2012; Randall et al., 2007; Yanaihara et al., 2006).

miRNAs are non-coding RNAs (~22 nt) derived from conserved short hairpin RNAs (shRNAs). They are cleaved from heterogenous transcripts by the microprocessor complex that contain the RNase III enzyme Drosha at its center. Several sequence elements such as a mismatch (bulge) near the base of the hairpin recruits the microprocessor to produce a pre-miRNA hairpin that exhibits a 3' 2nt overhang characteristic of RNase III processing. After this initial pre-processing, the pre-miRNA is exported into the cytoplasm by exportin 5 through the support of Ras-related nuclear protein guanosine triphosphatase (RanGTPase) (Dahlberg & Lund, 1998), where another RNase III enzyme, Dicer (Dcr) cleaves the loop from the pre-miRNA (György et al., 2001; Y. Lee et al., 2003). The resulting ~22nt asymmetric duplex RNA typically exhibits a central mismatch. The duplex RNA is separated based on thermodynamics of the 5' base identity so that the strand with the least stable 5' end is favorably selected and loaded into an AGO protein to form the RNA silencing complex (RISC). The passenger strand is degraded, leading to highly asymmetric accumulation of guide strand relative to passenger strand (Schwarz et al., 2003). Upon RISC formation, the guide strand leads AGO to the miRNA target through complementary base paring. This binding result in either target degradation or translation inhibition (Fig 1.1).

1.2.2 siRNA Biogenesis

siRNAs are similar in size and between 21-23nt long in animals. Unlike miRNAs, their biogenesis takes place in the cytoplasm where they originate from a variety of dual strand reads of various origin including viral dsRNA, endogenous dsRNA generated from primary transcripts, dsRNA of exogenous sources, transposable elements (TEs) etc. Endogenous-siRNAs (endo-siRNAs) are derived from substrates like Cis-natural antisense transcript (Cis-NAT) RNAs, and long hairpin RNAs among other sources. On the other hand, exogenous siRNAs arise from sources such as viral dsRNA, and long dsRNA introduced into the cell. When siRNA precursors and the right RNAi factors are available, either an siRNA dedicated Dcr like protein found in fruit flies or a Dcr enzyme with an overlapping function (as seen in humans and worms) can initiate the siRNA maturation process. Following cleavage, the siRNAs are loaded into AGO proteins in a similar manner to miRNA loading for RISC formation and target repression through degradation or translation inhibition (Fig 1.1).

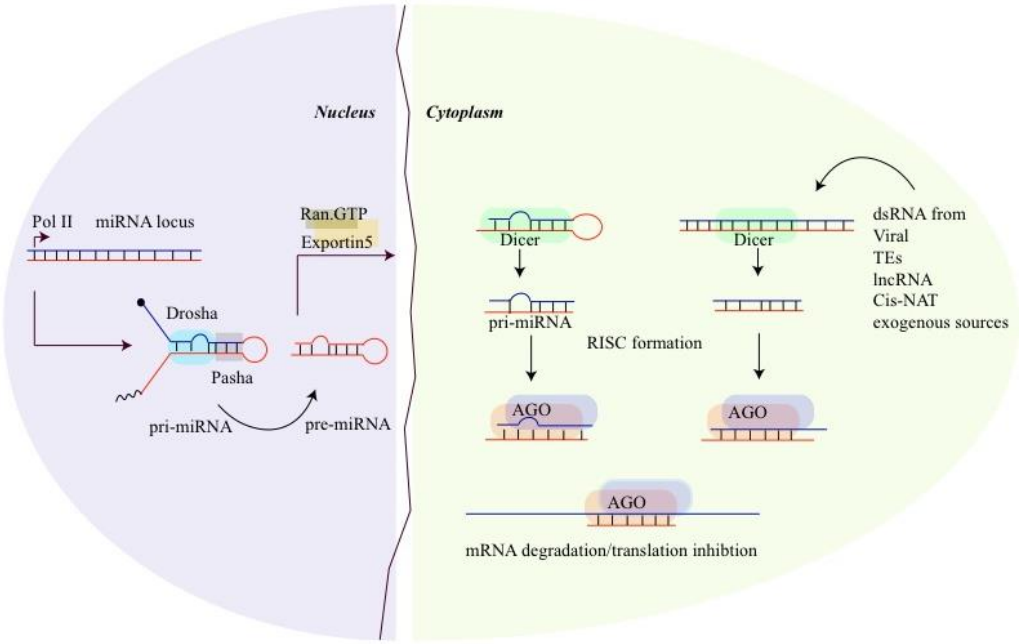


Figure 1.1 *Biogenesis of canonical miRNA and siRNAs.*

Diagram showing the biogenesis of canonical miRNA and siRNAs. The left (purple) section represents miRNA pre-processing in the nucleus while the right (green) shows maturation of pre-miRNA and long dsRNA in the cytoplasm

1.2.3 piRNA Biogenesis

Insect and vertebrate piRNAs are longer (25-30 nt) than miRNAs and siRNAs. These sRNA species were originally thought to be gonad restricted where they suppress TEs. It was however discovered that most insects have somatic piRNAs that appear to regulate gene expression in addition to suppressing TEs (C. Li et al., 2009). piRNAs are Dcr independent, 2'O-methylated, exhibit a bias for 5' uracil, and associate with PIWI proteins rather than AGO subfamily of the Argonautes.

Biogenesis of piRNAs may begin either in the cytoplasm or other cellular components. Due to their association with TEs, pinpointing their origin can be a difficult task. In fruit flies, where the biogenesis of piRNAs have been well characterized, attention have been given to unique mapping piRNAs originating from certain extended genomic

loci in other to understand their behavior based on their origin and subsequently, unravel the mechanism behind their biogenesis. These loci are termed piRNA clusters in flies. Two kinds of clusters have been identified in flies based on whether they map to one (uni-strand clusters) or both (dual-strand clusters) genomic strands. Transcriptionally, majority of the uni-strand clusters have an upstream Pol II promoter and associated transcripts that resemble canonical mRNAs possessing a 5'-cap and a poly-A tail with occasional splicing patterns. This is not the case for dual-strand clusters, where clusters lack clear Pol II associated promoters and transcripts lack Poly-A tail in most cases (Brennecke et al., 2007; Mohn et al., 2014). Two biogenesis pathways have been identified in flies: "Ping-Pong" and "Phasing". Both pathways involve recruitment of RNA fragments created by PIWI-mediated slicing.

The Ping-Pong cycle involves the activity of Aub and AGO3 proteins collaborating to generate piRNAs through an alternating slicing mechanism using both piRNA cluster transcript and active TE transcripts as substrate for the amplification process (Fig 1.2). In this pathway primary piRNAs generated through primary processing from piRNA cluster transcripts exported into the nuage of the cell are recruited by AGO3. Upon binding of the 5' uridylated (5'U) piRNAs by AGO3 to a cluster transcript, AGO3 slices the transcript exactly 10nt from the 5' end generating a substrate (piRNA with a U at the 5'end) either for Aub or phased piRNA pathway. In the ping-pong pathway, the AGO3 generated antisense 5' U bearing piRNA is then processed into appropriate length by the exonuclease, trimmer, and its cofactor papi. The processed piRNA is then loaded unto Aub and methylated by the Hua Enhancer factor 1 (HEN1). The Aub bound methylated piRNA then locates and bind to the active TE transcript. Upon binding, Aub slices the TE transcript in

a similar manner, generating an AGO3 associated piRNA with an adenine (A) at the 10th position. This piRNA is again trimmed to the appropriate length by trimmer/papi. This alternating slicing mechanism is known as the ping-pong piRNA biogenesis pathway. The signature 5’ U at position 1 and A at position 10 of piRNAs is therefore essential for identifying piRNAs generated through this pathway (Fig 1.2A).

The phasing pathway spins out of the ping-pong cycle whereby AGO3 generated antisense transcript from piRNA cluster transcript become substrate for cleavage by zucchini/mino from the unbound 3’ end before trimming by trimmer/papi. In fruit flies, the product of zucchini/mino cleavage is a 5’U antisense piRNA that is 3’-O methylated by HEN1 that is transported across the nuclear membrane where they then transcriptionally repress piRNA clusters and contribute to generating ping-pong pathway substrate (X. Huang et al., 2017; Luo & Lu, 2017).

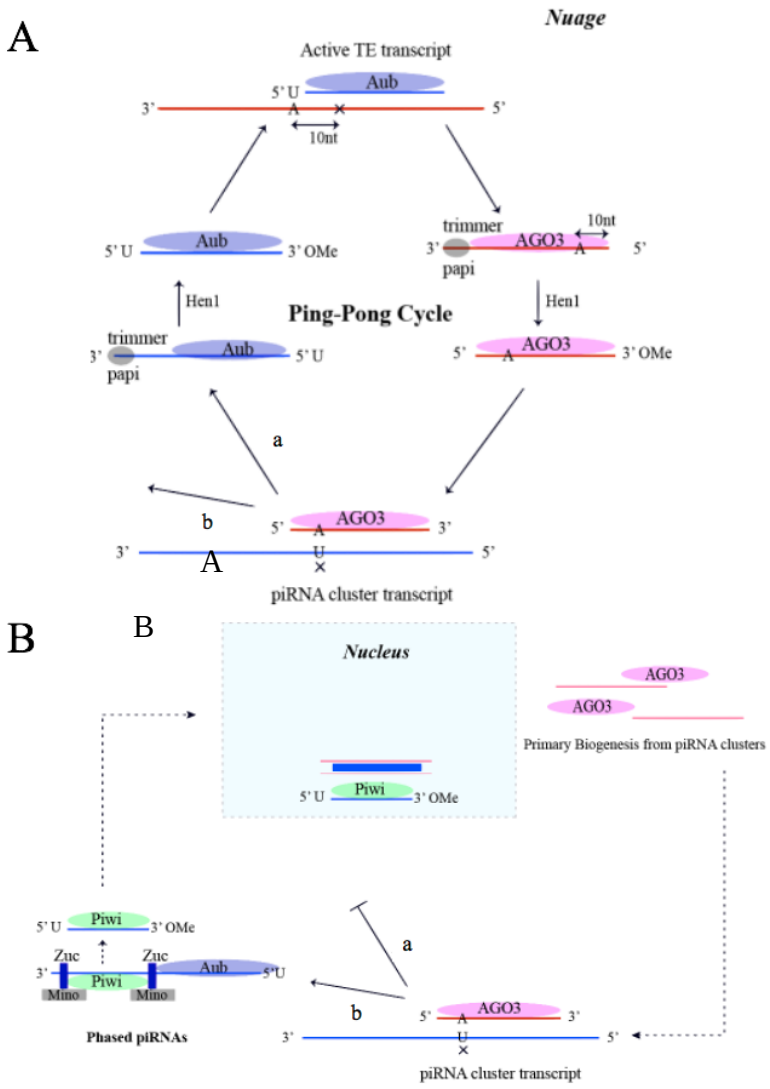


Figure 1.2 *Biogenesis of piRNA; Ping-pong and phasing pathways.*

A depiction of piRNA biogenesis through ping-ping amplification pathway (A) or phasing pathway (B) showing how PIWI proteins generate piRNAs from piRNA cluster transcripts to control TEs (Mohn et al., 2015).

1.3 RNAi factor biochemistry

A discussion about sRNA biogenesis is certainly incomplete without adequate understanding of the key proteins that are involved in their maturation and function. Key players of the RNAi process termed RNAi factors are discussed below. For a successful RNAi process to occur, RNAi factors including the complex, Drosha / DiGeorge syndrome

critical region 8 (Drosha/DGCR8 also known as the microprocessor), Dcr, Argonaute family of proteins, helicases, HEN1, and in some cases RNA dependent RNA polymerase (RdRP) are required. While some are dispensable for RNAi, others like Argonaute (AGO) and Dcr are essential. For example, RNAi involving miRNAs and siRNAs in human embryonic stem cells and *Drosophila melanogaster* (*D.me*) occur without the need for RdRPs (Gurung et al., 2021; Su et al., 2009a).

1.3.1 Dicer

Dcr protein alongside PIWI perhaps constitute one of the most critical proteins in the RNAi pathway as it is required for maturation of numerous sRNA effectors including miRNAs and siRNAs. Essential for miRNA biogenesis is the Drosha/DGCR8 complex that acts in the nucleus to initiate the maturation process of miRNA through cropping long primary transcripts pri-miRNA. This processing produces a shRNA (pre-miRNA) that is exported to the cytoplasm by exportin 5 for further processing by Dcr into ~22 nt mature miRNA duplexes. This short duplex RNA then proceeds into the miRNA pathway (S. Li & Patel, 2016).

Dcrs, whose activity was first characterized in 2000 at the Hannon group at the Cold Spring Harbor Laboratory through high-speed centrifugation and assaying of enzymatic activity of *D. melanogaster* lysate (Hammond et al., 2000). Armed with key enzymatic domains relevant for dicing dsRNA, eukaryotic Dcr are structurally unique with conservation across all eukaryotic supergroups. Through cryo-electron microscopy and crystallography, the structure of human Dcrs have been revealed as an L-shaped protein comprising three main sections: (1) the head which bears the PAZ domain, a dsRNA binding module with a looped enriched in basic amino acid that necessitate flexibility of

binding affinities of the protein, (2) a body which harbors the critical RNase III domain, and (3) a base where the helicase domain is found (J et al., 2006; P.-W. Lau et al., 2012). The helicase domain of Dcrs typically have an adenosine triphosphate (ATP) hydrolyzing residue (DExD, where x is any amino acid). However, this residue, though present in mammalian Dcr is not required for ATP hydrolysis.

Like AGOs, the number of Dcr protein homologs encoded in eukaryotic genomes varies. In *Drosophila,* different Dcrs have been shown to have dedicated roles in sRNA biogenesis. Mutagenesis studies involving *D. melanogaster* Dcr1 & 2 respectively implicated these proteins in miRNA and siRNA maturation (Y. S. Lee et al., 2004). In other animal genomes like humans, worms, lophotrochozoans and basal chordates like amphioxus, only one homolog of the protein is encoded. In humans and some invertebrates, this single Dcr tend to have overlapping function, engaging both miRNA and siRNA precursors (Svobodova et al., 2016).

1.3.2 Argonaute Family

These family of highly conserved proteins play essential role in the RNAi process by serving a core function in the RNA-induced silencing complex (RISC) during RNAi. Functionally, Argonautes are separated into three major sub-groups; AGO, PIWI and the worm specific group III Argonautes (WAGO), with all subfamilies playing essential function in RNAi. From a structural perspective, AGOs are characterized by four major domains; an N-terminal domain that is usually variable, a conserved PAZ (PIWI AGO Zwille) domain, middle and PIWI domain (Tolia & Joshua-Tor, 2007). Biochemical studies performed in-vivo revealed that the N-terminal domain is essential for unwinding dsRNA as well as possessing slicing activity in AGO (Hauptmann et al., 2013; Kwak &

Tomari, 2012). The PAZ domain necessitates ssRNA binding of the protein whereas, the MID-PIWI domain harbors the nucleotide binding pocket for anchoring the 5'end of the sRNA strand (Medley et al., 2021). The critical RNase H-like fold of the PAZ domain in some AGO proteins give the protein the critical target cleavage activity (Ji-Joon et al., 2004). In addition to high conservation in the AGO subfamily, different animal genomes encode different number of AGO homologs. For example, amphioxus, *D.me*, humans, and worms (*C. elegans*) encode 1, 4, 2, and 27 AGO homologs respectively (Yigit et al., 2006). Variation in homolog number is equally seen in the less conserved PIWI subfamily where most animals where the protein is encoded have at least two homologs.

In flies, three PIWI proteins have been identified including AGO3, PIWI, and Aubergine (Aub) where they are implicated in both germline and somatic piRNA pathways (Saito et al., 2009; Saito & Siomi, 2010). Also, AGOs are functionally distinct in flies with AGO1 and AGO2 respectively dedicated to miRNAs and siRNAs associated PTGS. In mammals, AGOs have been shown to have overlapping roles so that in human embryonic stem cells, AGO1 and AGO2 are capable of loading both bulge miRNA and perfectly matched siRNA duplexes (Su et al., 2009b). This phenomenon is also found in some plants. Classical Argonaute proteins tend to have DDH metal binding residues required for slicer activity. These residues are however absent in the WAGO subfamily. As a result, a slicer-independent RNAi mechanism involving translation inhibition is predicted for this subgroup.

In the siRNA biogenesis pathway, RNAi factors known as RdRP have been implicated in potentiating siRNA process in some eukaryotes through direct secondary amplification of siRNAs or antisense dsRNA synthesis from primary Pol II transcripts.

1.3.3 RNA dependent RNA polymerase

RdRP or RNA replicase is an ancient enzyme that catalyzes the replication of RNA from an RNA template. In this chapter, emphasis will be placed on eukaryotic RdRPs and highlight its critical roles in developmental regulation, maintenance of genome integrity, and defense against foreign nucleic acids. RdRP genes in eukaryotic ancestor have had three copies including RdRPα (alpha), RdRPβ (beta) and RdRPγ (gamma). While RdRPα genes have been sequenced from all three kingdoms, RdRPβ has only been found in animals and fungi and RdRPγ only in plants and fungi. Thus, RdRPβ was likely lost in the plant lineage, while RdRPγ was lost in animals (Zong et al., 2009a). We explicated RdRPs in *C. elegans* and plants to illuminate the critical function of this protein since these two systems have served as models where the protein have been vastly exploited.

The discovery of dsRNA as potent effector of gene knockdown through RNAi by Fire and Mello attracted attention of the scientific community, driving further research to understand the molecular components of the newly discovered gene regulatory mechanism. In 1998 when Fire and Mello published their findings, they observed that even few molecules of dsRNA when injected into *C. elegans* have the tendency to elicit significant gene repression arguing against stoichiometric interaction with cellular mRNAs. As a result, they suspected a potential amplification mechanism (Fire et al., 1998). This amplification mechanism eventually became known to be secondary siRNA synthesis by RdRP in worms.

RdRPs were initially identified in RNA viruses as enzymes that replicate the viral genomes (Shu & Gong, 2016). Later on, scientists revealed that some eukaryotes also contain RdRPs, such as responsible for the role in *C. elegans* described above, which are

involved in RNAi and differ structurally from viral RdRPs (Zong et al., 2009a). Among eukaryotes, the first activated-RdRP was isolated in tomato (RdRP1) and then in the *Neurospora crassa*, quelling dependent enzyme 1 (QDE-1) (Makeyev & Bamford, 2002; Schiebel et al., 1998) both demonstrated a role for RdRP in eukaryotic RNA silencing. Afterwards, RdRP-encoding genes were found in a wide range of eukaryotic species, including *C. elegans* (nematode), *Arabidopsis thaliana* (*A. thaliana*) (flowering plant), *Dictyostelium discoideum* (slime mold), and *Schizosaccharomyces pombe* (fission yeast) (Tijsterman et al., 2002). In another study, RdRP appeared to be involved in germ line development in nematodes (Smardon et al., 2000) and centromere function in fission yeast (Volpe et al., 2002). Unlike the wide distribution of RdRP genes in plants and fungi, only a few animals have putative RdRPs, and the RdRP homologs could not be detected in most animals, including human and fruit fly, even though RNAi occurs in these organisms (STEIN et al., 2003). One phylogenetic analysis of RdRP genes from a few species supports the monophyletic origin of eukaryotic RdRP genes (Cerutti & Casas-Mollano, 2006), however, the evolutionary relationships among eukaryotic RdRP genes remain unclear, especially between major groups of eukaryotic species.

According to a computational analysis on RdRP-like genes from available databases, it was shown that eukaryotic RdRP domain shared a conserved motif “DXDGD” (X represents any amino acid), as active site (Iyer et al., 2003). This was determined from the species whose genomes have been extensively annotated including RdRP domains of suppressor of gene silencing (SGS), silencing defective (SDE), RdRP1, RdRP2, RdRP6 in *A. thaliana* (to be discussed later), EGO-1 from *C. elegans*, QDE-1 from *N. crassa*, and RDP-1 from fission yeast, (50)In the past, 161 RdRP-like genes were collected from 56

species of a wide range of eukaryotes, from diverse unicellular fungi to multicellular organisms, and among them, 39 RdRP genes were observed in 12 plants, including 6 RdRP genes in *A. thaliana* and 3 in tomato (both eudicots), 5 in rice (monocot), 3 in *Selaginella moellendorffii* (basal vascular plant) and 3 in *Physcomitrella patens* (moss) (Zong et al., 2009a). Similarly, it has been demonstrated that RdRP genes are also widely distributed in three major groups of fungi: Ascomycetes, Basidiomycetes and Zygomycetes with a total of 76 RdRP genes from 24 fungal genomes (Pinzón et al., 2019a).

In another study, a single RdRP gene was found in *S. pombe* (Sigova et al., 2004), 3 and 4 genes in the nematodes *Caenorhabditis briggsae* and *C. elegans*, respectively, as well as one gene from the partial genomic sequences of another nematode, the soybean root parasite *Heterodera glycines*. Accordingly, the Chordate *Branchiostoma floridae* (*B. floridae*) was found with 6 RdRP homologs, and *Nematostella vectensis* (*N. vectensis,* a Cnidarian) with 4 genes. Recently, searches of the assembled genomic sequences of the Lyme disease tick, *Ixodes scapularis*, has been revealed to have 4 genes containing the RdRP homolog (Pinzón et al., 2019a).

1.3.3.1 Origin of RdRP

The origin of the RdRP was traced back to tRNA ancestors (de Farias et al., 2017). The most ancient part of the protein at the origin of RdRP is the cofactor-binding site that had the capacity of binding to magnesium, calcium, and ribonucleotides. RdRP originated from junctions of proto-tRNAs that worked as the first genes at the emergence of the primitive translation system, where the RNA was the informational molecule (de Farias et al., 2017). The proximity between the RdRP from *N. crassa* and the ancestor of RdRP suggests two scenarios. In the first scenario, the ancestor virus lineage was incorporated in

the eukaryotic genome and was lost during the evolutionary process. The second scenario suggests a recent acquisition from a modern virus by lateral gene transfer, and the structural similarity between the eukaryotic RdRP and the ancestral RdRP is a product of evolutionary convergence. In this scenario, the selective pressure that promoted the convergence was the replication of sRNA molecules, both in eukaryotes and in primitive RNA-based genomes (de Farias et al., 2017).

Based on genetic divergence and molecular evolutionary analysis, RdRP genes likely experienced early duplications that produced several copies before and after the divergence of major eukaryotic lineages. As mentioned earlier, it has been said that several duplicated copies have been stably maintained during the evolution of major groups of plants and fungi, while they seem to have been lost in vertebrates, insects and some other animals, hence the information for stably maintained distinct groups of plant and fungal genes and their distinct function suggests that they represent ancient functional differentiation, and relatively recent gains and losses in plants, fungi, and animals indicate that the function of RdRP genes continues to evolve (de Farias et al., 2017). Opposite of previously vision, now we know RdRPs are much more common in animals than previously thought, but their genes were independently lost in many lineages, and all hypotheses regarding the evolution of RdRP genes and possible functions of the majority of RdRP genes, requires more investigation to be functionally characterized.

1.3.3.2 Eukaryotic RdRPs encoded in the nematode, *C. elegans*

C. elegans which have served as a model for understanding RdRP function in animals has a complex endogenous sRNA landscape. These distinct sRNA populations can be distinguished by size, 5′ nucleotide bias and 5′-end phosphorylation some of which

include miRNAs, 21U, 26G and 22G RNAs. Further, 21U-RNAs are 21 nucleotides long and have a 5′ bias for uridine monophosphate; 26G-RNAs are 26 nucleotides long and have a 5′ bias for guanosine monophosphate; and 22G-RNAs are 22 nucleotides in length and have a 5′ bias for guanosine triphosphate (Ruby et al., 2006). In fact, 26G and 22G RNAs are recognized as products of RdRP activity via synthesizing sRNAs from template RNAs in a non-primed and non-processive manner, hence, it has been hypothesized that all RdRP products have an antisense orientation.

The genome of *C. elegans* encodes four RdRP genes including ego-1 and *rrf*-1 which are necessary for 22G-RNA biogenesis; *rrf*-3, required for 26G-RNA biogenesis; and *rrf*-2 with an unknown function. Accordingly, *C. elegans* produces 22G-RNA in response to exogenous dsRNA (Almeida et al., 2019). Since all sRNA pathways converge on 22G-RNA, many sRNA pathways compete for shared factors required for 22G-RNA amplification and silencing. The RdRPs *rrf*-1 and EGO-1 are required for 21U-RNA-driven 22G-RNA biogenesis. However, only *rrf*-1 was implicated in 26G-RNA-driven 22G-RNA biogenesis. 26G-RNAs are produced by ERI complex of RdRP, *rrf*-3 and Dcr-1 (Almeida et al., 2018; Thivierge et al., 2012; Vasale et al., 2010). The conserved CHHC zinc finger-containing GTSF-1 and Tudor domain-containing ERI-5 proteins form a pre-complex with *rrf*-3, which is required to bring *rrf*-3 to the remaining ERI complex members (Thivierge et al., 2012). It is not known how the ERI complex is recruited to specific target transcripts, but a current model of 26G-RNA biogenesis implies RNA synthesis by *rrf*-3 antisense to a template target RNA, creating a dsRNA intermediate with a blunt end which creates a substrate for Dcr-1 cleavage and selectively stimulates the production of a 26-nucleotide long cleavage product (Blumenfeld & Jose, 2016). Based on this current model,

rrf-3 would then synthesize another 26G-RNA from the next cytosine available 3′ of the Dcr-1 cleavage site, thereby initiating another cycle of biogenesis. Regarding *rrf*-3-like RdRP orthologues which has been distributed throughout the nematode phylum, it has been argued that ancestral nematode sRNA biogenesis is perhaps like *rrf*-3-mediated 26G-RNA biogenesis in *C. elegans* (Sarkies et al., 2015a). This have been widely supported by the nature of RdRP homologs distributed throughout eukaryotes.

1.3.3.3 RdRPs in plant

The knowledge of regulation of RdRPs and its interacting protein partners in plants might help in developing resistant plants to biotic stresses. Early genetic screens in *Arabidopsis* were instrumental in uncovering numerous proteins required for RNA silencing pathways.

Further, RdRPα genes have duplicated in plants to yield separate RdRP1, RdRP2, and RdRP6 subgroups, with apparent functional diversification, and *A. thaliana* has one of each type. *A. thaliana* RdRP1, RdRP2, and RdRP6 all share the C-terminal canonical catalytic DLDGD motif of eukaryotic RdRPs and have orthologs in many plant species. Initially, these enzymes were studied because of their role in plant antiviral and transgene silencing, but it is becoming increasingly apparent that they have additional molecular functions, including the control of chromatin structure (RdRP2) and the regulation of cellular gene expression (RdRP6) (Willmann et al., 2011b). RdRP2 is involved in the biogenesis of repeat-associated siRNAs (rasiRNAs) which are about 24 nucleotides in length and are associated with heterochromatic and repetitive regions of the *A. thaliana* genome, including pericentromeric regions and telomeres (Vrbsky et al., 2010).

A. thaliana also has three RdRPγ genes (RdRP3, RdRP4, and RdRP5; also called RdRP3a–RdRP3c), which share an atypical DFDGD amino acid motif in the catalytic domain. To date these three proteins have not been assigned functions, but the presence of at least one RdRPγ gene in several other sequenced plant genomes, including rice (*Oryza sativa*), poplar (*Populus trichocarpa*), moss (*Physcomitrella*), and a lycophyte (*Selaginella* sp.), as well as in many fungi, suggests that one or more RdRPγ proteins in Arabidopsis may have functional significance, likely in RNAi (Willmann et al., 2011b).

Apart from defense mechanism, RdRPs have also been associated with several aspects of *A. thaliana* development, and most of these roles are attributed to RdRP6 (Elmayan et al., 1998). Mutations in RdRP1 are reported to neither have morphological effect. The only characterized developmental role for RdRP2 in *A. thaliana* is in the development of the female gametophyte. By contrast, loss-of-function mutations of the RdRP2 ortholog in maize are reproducibly late flowering and have an abnormally shaped vegetative meristem (Willmann et al., 2011b). In summary, plant RdRPs play significant roles in *A. thaliana* development and environmental responses through their functions in multiple sRNA-mediated RNA silencing pathways. However, our understanding of the full impact of these proteins on plant biology and endogenous RNA silencing pathways is still quite lacking.

CHAPTER II - GUIDING RNAI DESIGN THROUGH CHARACTERIZATION OF ENDOGENOUS SMALL RNA PATHWAYS

2.1 Abstract

RNAi is a common eukaryotic gene regulation process driven by sRNA effectors. Mechanisms that govern sRNAs behavior have only been extensively described in a handful of organisms, which suggests that the most effective RNAi approach in many organisms, such as insect pests, remains to be determined. Taking advantage of advances in high throughput sequencing, characterization of sRNA molecules can be achieved through bioinformatic approaches without the need for genetic experiments. This chapter describes pipelines for the characterization of the three main classes of sRNAs (microRNAs, small-interfering RNAs, and PIWI-associated RNAs) using computationally determined sRNA biogenesis signatures. Understanding the abundance of different sRNA classes will lead to better-informed RNAi strategies.

2.2 Introduction

Post-transcriptional gene silencing through RNAi is mediated by sRNAs. These ncRNAs act through complementary base pairing with target RNAs, triggering mRNA destruction or translational inhibition as well as epigenetic regulation (Rinn & Chang, 2012). Diverse biogenesis pathways and activities of sRNAs have been identified over the last decade, mostly due to advances in sequencing technologies. In addition, the simplicity involved in library preparation methods for sRNAs where fragmentation steps used for mRNA prior to sequencing are avoided, the availability of essential molecular signatures, amenability, and cost effectiveness in addition to how relevant these molecules have contributed to increased interest from researchers in pursuing bioinformatics-based

approaches rather than biochemical means for studying sRNA species. Due to the ability of NGS library preparation methods to capture unmodified reads or reads with terminal modifications without the need for fragmentation, it is possible to recover sRNA populations in their original states and attribute them to the different sRNA classes through in-silico means based on known biogenesis signatures and available computationally defined methods. Taking advantage of the enormous amount of NGS datasets available on the national center for biotechnology information (https://www.ncbi.nlm.nih.gov) database especially sRNA libraries, the journey to performing or completing an RNAi study is already halfway.

In animals that lack available genetic tools, identifying potent sRNA effectors can be difficult and researchers must resort to rigorous biochemical characterization methods that are time consuming and requiring high level molecular assay expertise and technical knowledge of the experiments to select a candidate sRNA locus whose efficacy might not be strong enough to elicit the appropriate response.

Several uses for RNAi based pest control methods have been elucidated and are currently being used. For example, the commercial corn product “Stax Pro” uses a host induced genetic approach to elicit RNAi in the Western corn rootworm (WCR) *Dabrotica virgifera virgifera*. The transgenic host crop was engineered to expresses a hairpin dsRNA that target an integral member of the endosomal sorting complex, *snf7*, an essential gene of WCR whose knockdown have been shown to result in larval lethality (Bolognesi et al., 2012; Head et al., 2017; Ramaseshadri et al., 2013).

Although, hundreds of sRNAs have been identified in animals, high-throughput sequencing and bioinformatics tools repeatedly reveal the existence of new miRNAs,

siRNAs and piRNAs. Insects are one of the most diverse and successful groups, so it is not surprising that they also have diverse and elaborate RNAi pathway (Vogel et al., 2019). As a result, RNAi approach for each insect should be tailored to the prevalent sRNA pathway. To achieve this, an in-depth knowledge and understanding of the endogenous RNAi biology of the animal in question would therefore be valuable to identifying an effective RNAi strategy. Moreover, these insights could lead to novel strategies in pest management. There is therefore a need to design an all in one in-silico approach or methodology for characterizing and identifying various sRNA classes as well as their abundance that could be used for generating an efficacious RNAi approach. This chapter provides the design, pipeline, and advice to demonstrate tool usage using an example sRNA sequencing data from the WCR. Due to the availability of several open source and online tools for identifying miRNAs, this chapter will focus more of siRNA and piRNA identification.

2.3 Relevance

Efficacy of RNAi process is very important in RNAi experiments. Many challenges exist for designing RNAi mediated gene silencing experiments for animal models where genetic tools are absent or improperly defined. Example of such animals include the many insect pests of economic importance that affect a wide variety of crops creating huge economic loss throughout the world.

Citruses, an excellent source of oranges, lemons, and processed fruit juices, serve as the go to source of vitamins, sugars, and other organic compounds (Rampersaud & Valim, 2017). The major global growers of citrus including China, United States of America, and Brazil have reported losses due to the emergence of Huanglobing (HLB) bacterial disease caused by gram-negative bacteria members in the Candidatus

Liberibacter genus and transmitted by the insect vector *Diaphorina citri* (Bassanezi et al., 2020; Graham et al., 2020; Zhou, 2020). RNA mediated control strategies have been proposed for use against this pest. Another example where an RNAi based control method was used is the piRNA mediated RNAi approach proposed for the control of the white fly (*Bemisia tabeci*), an invasive crop damaging pest that transmit plant viruses throughout the world. Using computational methods, piRNA candidates have been identified in this animal and fused with the aquaporin1 and alpha glucosidase 1 to serve as silencing triggers against the respective targets following feeding (Mondal et al., 2020). This is one of many ways in which endogenous sRNAs can be exploited for design of RNAi triggers in non-model animals through the protocol designed in this work.

2.4 Materials

2.4.1 Required Installations

- Numpy, python (versions 2 and 3) and pysam are required for the pipeline.
- fastqc: https://anaconda.org/bioconda/fastqc
- fastx_toolkit: https://anaconda.org/bioconda/fastx_toolkit
- bowtie (version 1): https://anaconda.org/bioconda/bowtie
- samtools: https://anaconda.org/bioconda/samtools
- mirdeep2:https://www.mdc-berlin.de/content/mirdeep2-documentation
- sra-tools: https://anaconda.org/bioconda/sra-tools
- R-programming language: https://anaconda.org/r/r
- Trimmomatic: http://www.usadellab.org/cms/?page=trimmomatic

2.5 Method

2.5.1 Preliminary sRNA quality assessment

Initial Quality determination: following total RNA extraction with TRIzol, purity and amount should be estimated with Nano drop or qubit, degradation and RNA quality should be assessed with bioanalyzer, gel electrophoresis should be used as an alternative to assess RNA integrity. The first option is highly recommended.

2.5.2 Library preparation and sequencing

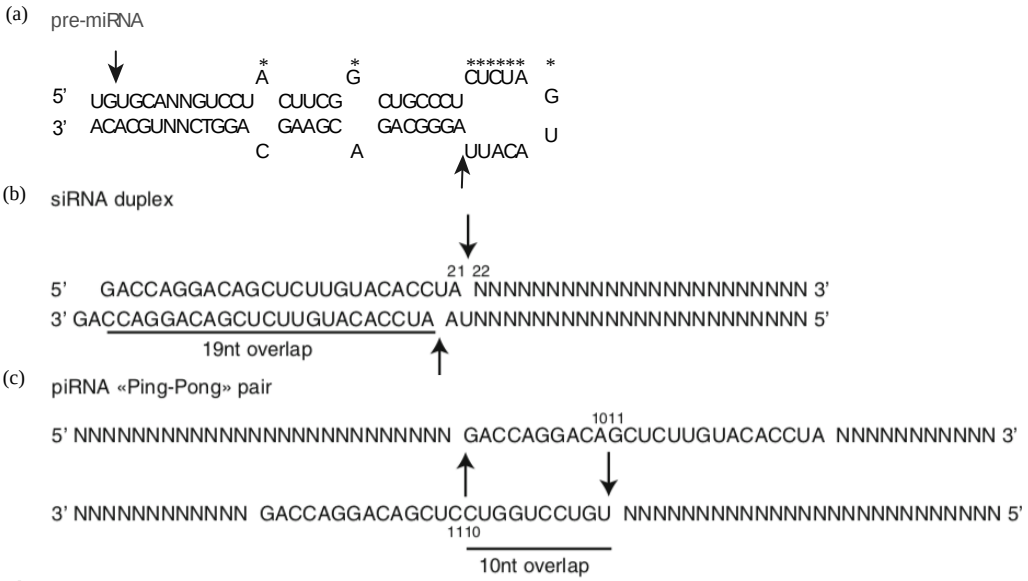
- Select a sRNA library prep kit: NEBNext sRNA prep kit, NextFlex, SMARTer, smRNA kit, CATS and TruSeq sRNA library prep kit are all commercially available.
- Follow kit-specific protocol for library preparation
- Size selection: Use bead-based selection or PAGE purification to select desired sRNA reads.
- Sequence reads on an Illumina sequencing platform or any other appropriate platform.

2.5.3 Obtaining and processing the sequencing data

In this Chapter, a dataset from WCR, SRA database accession number: SRR10430839 was used here to demonstrate the pipeline presented in this protocol using “prefetch”. If sRNA sequencing was done, proceed to the next step. Following download of 3 biological replicate data, the single end sequence read archive (sra) files were decompressed with the fastq-dump tool from the sra-toolkit and combined into one file.

2.5.4 Preparing the data

Adapter identification: using fastqc tool of the sra-tookit, quality of the library was determined to identify adapter sequences contamination (Fig 2.1C) and to remove them. sRNA sequencing libraries with adapter only contamination greater than 60% should not be used for analysis. Clip adapters from the 3' end of reads using either Trimmomatic or the fastx_clipper in the fastx-toolkit suit. Only retain reads for subsequent analysis that are clipped and are longer than 15nt.



(d)

Overrepresented sequences

Sequence	Count	Percentage
TAGCTTATCAGACTGATGTTGACAGATCGGAAGAGCACACGTCTGAACTC	4955599	38.54100535612445
TGAGAACTGAATTCCATAGGCTGAGATCGGAAGAGCACACGTCTGAACTC	809006	6.291853836264156
TGAGAACTGAATTCCATAGGCTGTAGATCGGAAGAGCACACGTCTGAACT	590670	4.593796962527039
TGAGGTAGTAGTTTGTGCTGTTAGATCGGAAGAGCACACGTCTGAACTCC	443029	3.445553819410824
TAGCTTATCAGACTGATGTTGACTAGATCGGAAGAGCACACGTCTGAACT	268444	2.087760055206136
TGAGGTAGTAGATTGTATAGTTAGATCGGAAGAGCACACGTCTGAACTCC	231425	1.799853491886874

Figure 2.1 *Major sRNA classes and biogenesis.*

(A) A canonical miRNA precursor with downward-facing arrow showing region of Drosha cleavage and upward-facing arrow showing region of Dcr cleavage. N represent n number of

Figure 2.1 (continued)

nucleotide and asterisk (*) represents mismatched regions. (B) shows a typical siRNA duplex with Dcr activity on 21 nt region and a signature staggered cleavage pattern (arrow) with 2nt 3' overhang. (C) Shows ping-pong generated piRNA pair with a characteristic 10nt overlap. PIWI slicing positions are represented with arrows (Antoniewski, 2014). D is an example fastqc result showing over-represented reads that contains the adapter used for library preparation.

2.5.5 In-silico identification of miRNA

Many tools including miRDeep2 and miRanalyzer (Friedländer et al., 2012; Hackenberg et al., 2011) have been developed for miRNA identification that incorporates sequencing data. As a result, this study only reported the result of this analysis without going in depth. These tools are freely available through the links below and are provided with straight forward instruction for installation and use.

miRDeep2: https://www.mdc-berlin.de/content/mirdeep2-documentation

miRanalyzer (sRNAbench): https://arn.ugr.es/srnatoolbox/srnabench/

To demonstrate miRNA discovery with these tools, miRdeep2 was used to identify miRNAs in the downloaded WCR sRNA data. Following a successful run of miRdeep2, the generated files will include a html link (shown in Fig 2.1). The example novel miRNA in Fig 2.2, A and B comprise three major segments, the matured (red = 5' arm), the Star (purple = 3' arm) and the loop (yellow) regions. The characteristic loop and bulges facilitate Drosha cleavage and essentially differentiating miRNA precursors from siRNA precursors. Hence, one can successfully identify miRNAs from seq-data leaving behind sRNAs (such as siRNAs) that did not result from short hairpin structures.

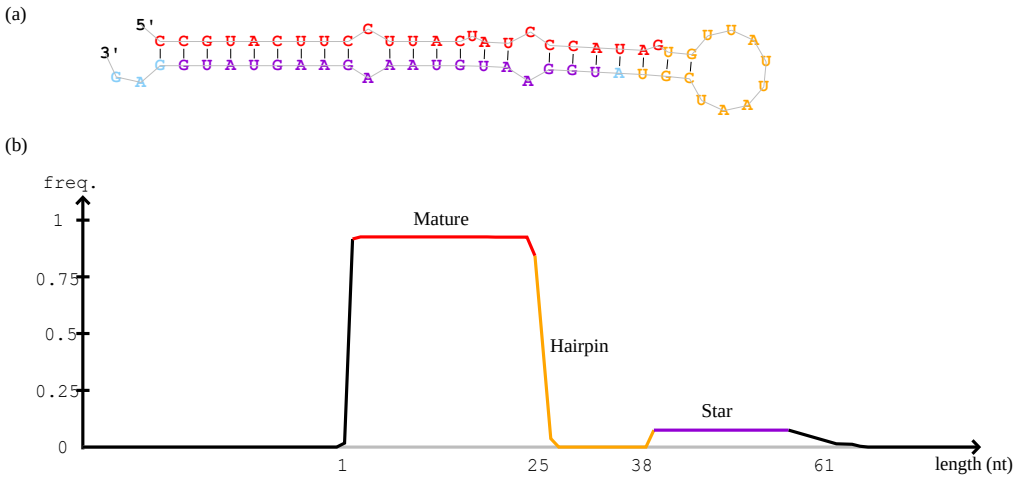


Figure 2.2 *WCR miRNA NW_021040030.1_748375.*

A portion of miRDeep2 output with A showing the hairpin structure possessing GA at the 3' end suggesting Drosha cleavage. B is a graphical representation of read frequency (y-axis) against read length (x-axis). In A and B, the red portion (mature) is the 5' arm, purple portion (star) is the 3' arm and the yellow portion is the hairpin.

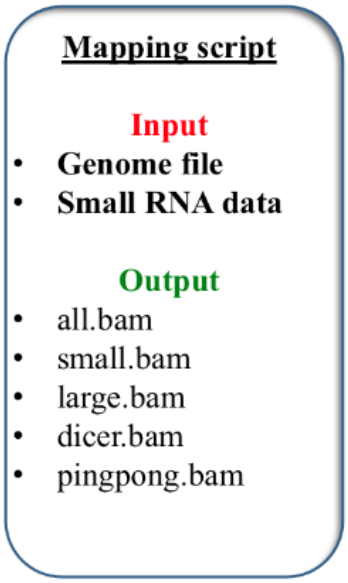
2.5.6 In-silico identification of siRNAs and ping-pong generated piRNAs

Executing the steps bellow by obtaining relevant scripts from the Flynt lab or the published chapter (Peter et al., 2022) in the order presented will identify all centers or loci with high Dcr activity and ping-pong amplification evident by finding accumulation of dual strand reads with 2nt 3' overhang Dcr signatures for siRNAs and 5' U-bias > 25nt reads with preference for an adenine at the 10th position for ping-pong generated piRNAs or trailing 5'U of Aub cleavage.

The first process involves mapping a sRNA sequencing data to the relevant genome sequence. The shell script can be obtained and saved as "mapping_smallRNAs1.sh". Before running the script on the command line, permission should be giving for it to execute using" chmod +x mapping_smallRNAs1.sh". In the script, the first line is the shebang line that specifies the type of script being run and is required for the shell script to

work every file required as specified in the comment (#) lines must be provided in the current working directory (cwd) where necessary.

The mapping script takes two inputs: a genome sequence file in multi-fasta format corresponding to the genome of the model organism and a sequencing data file in fastq format. Users should create the required python script (overlapping_reads.py) by downloading it from the published article referenced and saving it with the same name (Antoniewski, 2014). This script is needed to identify reads that exhibit 2nt 3' overhangs typical of Dcr processed reads or the 10nt overlap for the ping-pong piRNA signatures as will be seen when we attempt to recover this piRNA reads later.



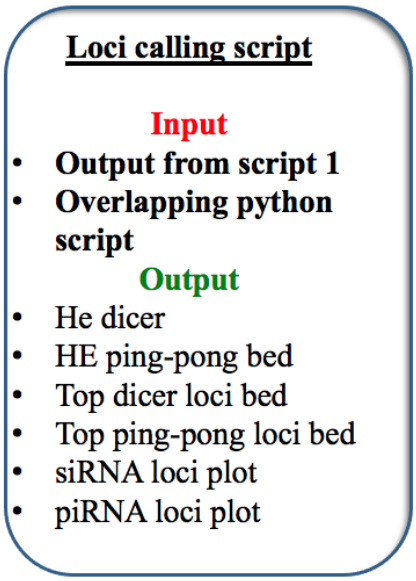
Scheme 2.1 *Mapping script showing major mapping steps for sRNA identification*

This scheme shows summary of the mapping script with major input and output files required for the first step in the small RNA identification pipeline,

To identify loci in which the most siRNAs and ping-pong piRNAs are being generated in the model animal, the following steps should be followed or the smallRNA1

or a custom script may be requested from the Flynt's lab at the University of Southern Mississippi or from the book chapter wherein the steps are published

The script's operations are as follows; The mapped reads from the steps above are used to identify regions of higher Dcr-associated reads and piRNAs by searching for loci were read depth exceed a threshold. This threshold can be adjusted as needed to increase or decrease stringency.



Scheme 2.2 *Loci calling script showing required input and output of HE sRNA producing loci including siRNA and ping-pong generated piRNAs.*

This scheme shows summary of the siRNA/piRNA loci identification script with major input and output files required for the second step in the small RNA identification pipeline.

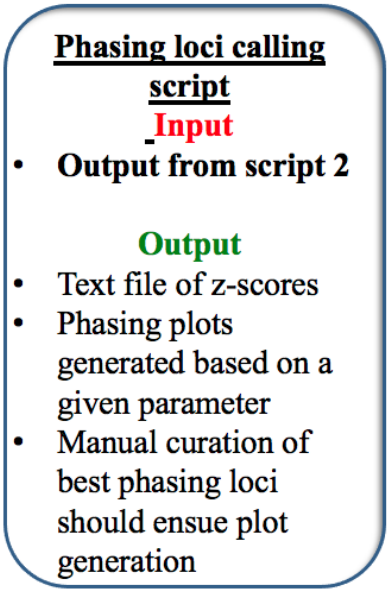
The results are then sorted in a defined order so that the best candidate loci are selected. Usually, these top loci would be most efficient for designing RNAi reagents due to the high siRNA/piRNA expression levels and the increased potential for recruitment by Argonaute proteins. The selected top loci (100 loci in this example) are individually queried for the presence of dual strand reads that meets either Dcr or ping-pong criteria. These are then counted and collected as read depth. Depth of the reads for all loci are then

plotted as a bar graph using ggplot2. The plotting command is included in the above step and can be found in the script upon request (Fig 2.3A-B).

siRNA producing loci comprise regions of high dual strand Dcr signatures where reads overlap perfectly leaving 2nt overhang on either 3' ends. These loci are listed in the script output named: siRNALoci.txt (top 100 only in this example). Also, ping-pong piRNA producing loci represent regions of high ping pong activity where piRNA sized reads overlap each other by 10 nt. They are listed in the script output named: piRNALoci.txt (only top 100 in this example). This sort of overlap signature is often observed in the ping-pong activity seen when Aub and AGO3 produce piRNAs in flies as discussed in chapter I.

2.5.7 In-silico identification of phased piRNAs

A file with the command below should be obtained from the Flynt lab or the published chapter and saved as "smallRNA2phasing.sh" (Peter et al., 2022). Also, a second file, "phasingPlots.R" with the command following the smallRNA2phasing.sh should be created and saved to the cwd. Executing the smallRNA2phasing script will generate phasing piRNAs. This script will use some of the files generated earlier by the earlier steps. In addition, the python script "piPipes_nuc_percentage.py" should be downloaded and saved from Han et. al. (Han et al., 2015) .



Scheme 2.3 *Loci calling script showing required input and output of piRNAs generated in phase through phasing biogenesis pathway*

This scheme shows summary of the phased piRNA loci generation script with major input and output files required for the third step of the small RNA identification pipeline.

Once the script is generated or obtained as mentioned above, it can be executed as follows; Usage on the command line is: ./smallRNA2phasing.sh genome.fasta

By using the longer reads (25 to 30nt) that mapped to the reference genome, the reads are converted into a bed format that is then intersected with all regions where read depth exceeds a certain threshold. This threshold can be adjusted as needed to increase stringency in locus selection. The result of this intersection is then sorted in descending order so that the top 100 loci are separated and queried for the presence of phased piRNAs. Each line in the resulting file (top 100) representing each locus is then passed into the python script generated earlier, "piPipes_nuc_percentage.py" (this should be present in the directory at the time of script execution) (Han et al., 2015).

piPipes_nuc_percentage.py searches the reads in each location and count each nucleotide by base position across all reads, calculates and assign z-scores to each

nucleotide (“A”, “C”, “G” and “T”) before creating 100-line graphs using Z-score information of only U at all nt positions (Fig 2.3, C-D). The line graphs are then scanned for the presence of Zucchini (WCR Zuc protein: XP_028134317.1) like signature characterized by a 1’ U-bias usually between 1-4 position and a downstream trailing U originating from HEN1 methylation of Aub processing (absent in Fig 2.3C but present in Fig 2.3D) (Mohn et al., 2015). The locus in Fig 2.2D is therefore an evidence of phasing biogenesis and should be considered a potential phasing piRNA locus.

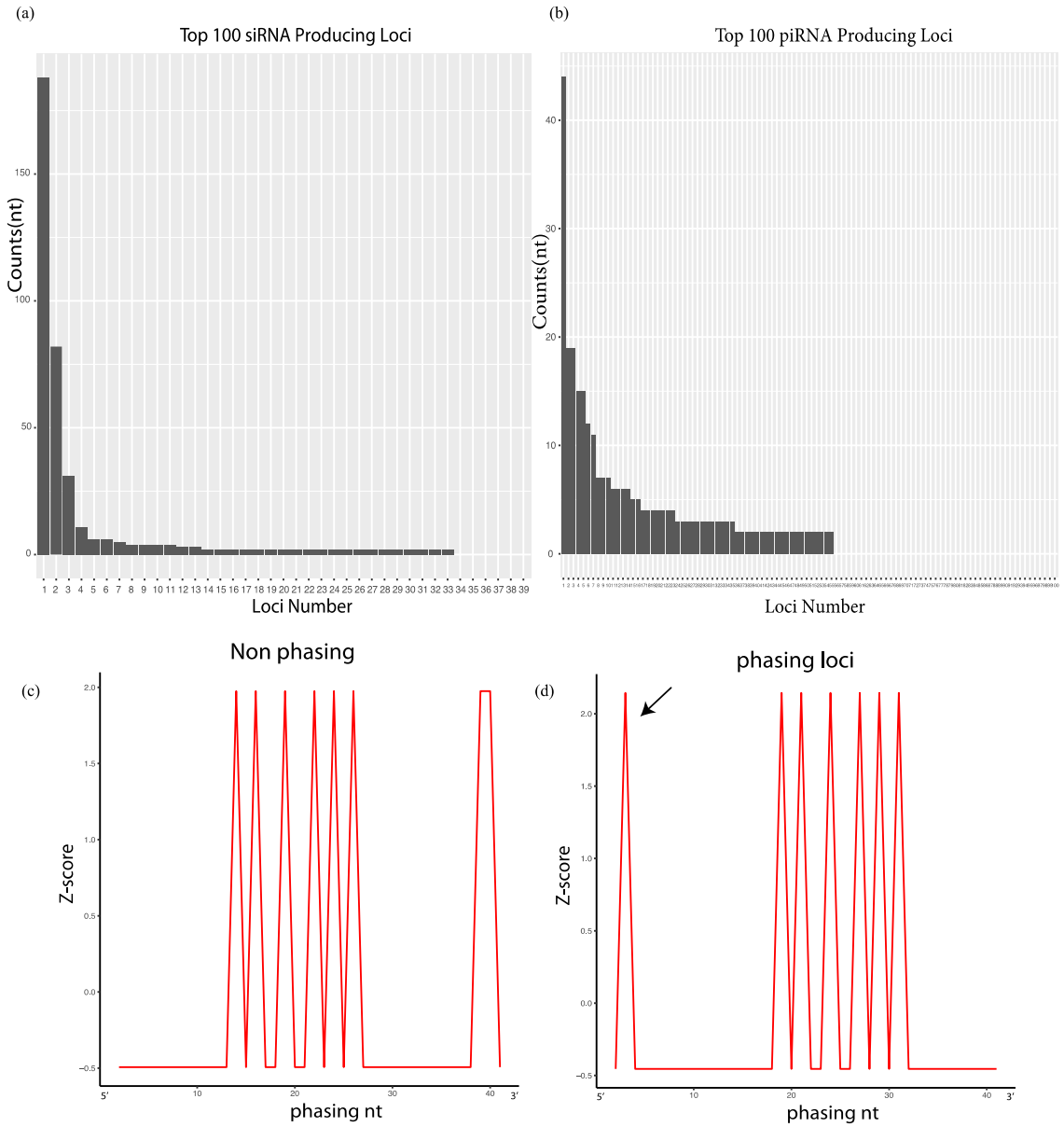


Figure 2.3 *Top loci containing high read pairs activity.*

A and B respectively represents the top 100 high-expressed siRNA and ping-pong piRNA producing loci with y-axis showing read count (nt) and x-axis showing each locus. C and D show distribution of each nt in 2 of 100 locus. C is an example of a non-phasing locus and D is a phasing piRNA locus. In D, arrow points to 5'U generated by Zucchini during phasing piRNA biogenesis (usually in position 1-4 in the 5' to 3' direction). Red line = U, y-axis is the z-score, and x-axis is the distance toward the trailing the first U.

2.5.8 Decision making based on In-silico sRNA characterization

Based on results from the pipeline, a choice can be made regarding the best sRNA class to exploit for RNAi experiment of technology. For example, in the WCR sRNA analysis we performed here, we observed more piRNA loci based on the defined criteria relative to siRNA loci (Fig 2.3B). However, the siRNA loci appear to have greater read abundance. Widespread piRNA loci in this example could mean that piRNAs are more active in WCR, and that RNAi reagents based on piRNAs might elicit more gene silencing.

2.5.9 Steps to design RNAi constructs

Standard sequence overlap extension (SOEing) PCR reaction should be used to amplify gene of interest. sRNA of interest may be designed into the primer sequences for SOEing. Synthesis is then followed by in vitro transcription. Upon completing the above steps, the dsRNAs are ready to apply to insects or animal model of interest through oral, injection or soaking delivery methods. The incubation time required for dsRNA processing as well as delivery amount is dependent on the insect and target of interest being investigated and should be optimized. Total RNA extraction should be done using TRIzol or tri-reagent followed by gene expression analysis to observe knockdown efficiency using untreated animals and target dsRNA without sRNA as controls.

CHAPTER III - ASSESSING EXOGENOUS SRNA MODALITIES IN THE TWO SPOTTED SPIDER MITES

3.1 Introduction

Organisms distant from *Drosophila* show distinct sRNA biology, which is most apparent in chelicerates such as the two-spotted spider mite, *Tetranychus urticae* (*T. urticae*), a pest affecting hundreds of different crops and ornamental plants. These mites can rapidly develop resistance to conventional pesticides, making this organism a prime candidate for new pest control technologies such as RNAi (Dermauw et al., 2013).

A consequence of the rapidly evolving nature of siRNAs is inconsistency in RNAi efficacy across arthropods. For example, long dsRNA feeding is effective in beetles, but not in moths and butterflies (Ivashuta et al., 2015). Exploiting sensitivity to dsRNA in beetles has already yielded commercial products to combat the western corn root worm (Zukoff & Zukoff, 2017). However, similar products are not yet available for other pests. This suggests that differences in RNAi response is a major complication that limits application of gene silencing technology. In such an application, animals will need to ingest molecules like long dsRNA, which adds further barriers such as destruction of molecules in gut lumen prior to entry into cells (Spit et al., 2017). Moreover, preventing incidental entrance of exogenous RNA into RNAi pathways is likely a desirable trait, especially for polyphagous arthropods which ingest myriad RNA structures and sequences. Thus, characterization of exogenous RNA processing is beneficial for identifying candidate RNAs for technology development. In this study, methods appropriate for non-model organisms that lack genetic tools are used for dissecting sRNA biogenesis. The framework

of the approach here would be valuable for examining RNA processing in other non-model organisms.

The sRNA effectors of RNAi are particularly amenable to high-throughput sequencing approaches due to their short length. Fragmentation is not part of sequencing library creation, so RNA ends created by RNases like Dcr are preserved. Also due to sequential ligation of adapters strand is preserved. By analyzing read sizes, nucleotide content, and overlaps between complementary, alternate-strand mapping reads; class and biogenesis patterns can be elucidated. In arthropods, Dcr products (siRNAs and miRNAs) are typically 19-23 nt long. The RNase III domain of Dcr causes staggered cleavage of dsRNAs that leaves 2-nt 3' overhangs between strands of siRNA duplexes. In contrast, piRNAs are represented by longer 25-30 nt reads. Biogenesis of piRNAs involves cleavage by PIWI proteins guided by the activity of pre-existing piRNAs. Ago and PIWI proteins may exhibit slicer activity which cuts the base of target transcript 10 bases from the 5' end of the bound sRNA. After PIWI cleavage, the fragmented transcript is converted into a new piRNA directly or by alternative nucleases. Each of these biogenesis modes can be determined through sequencing data analysis independent of genetic tools as described earlier in the previous chapter.

To evaluate efficiency of feeding-based RNAi in the two-spotted spider mite, *T. urticae*, an approach was used that combines a purification method alongside detailed analysis of sRNA sequencing libraries. Several studies have tested *T. urticae* for sensitivity to dsRNA, finding saturating exposure like soaking in dsRNA solution is required for robust phenotypes (Suzuki et al., 2017a; Yoon et al., 2018). This suggests further investigation of sRNA populations in *T. urticae* could lead to superior RNAi techniques.

Furthermore, like other chelicerae arthropods mites harbor RNA-dependent RNA polymerases (RdRPs), a factor in nematodes and plants that potentiates RNAi through amplifying dsRNA. However, there is little functional evidence that animal RdRPs, other than those found in nematodes, have a role in RNAi pathways that interact with exogenous RNAs (Dalmay et al., 2000; Pinzón et al., 2019a; C. Zhang & Ruvkun, 2012). Additionally, spider mite siRNAs have a central role in genome surveillance unlike in *D. melanogaster* (Grbić et al., 2011; Mondal et al., 2018).

By applying this strategy, we find a variety of exogenous RNAs are processed into sRNAs following ingestion by spider mites. Interestingly, the most active configuration is not the expected long dsRNA. Further investigation of the sRNA library revealed that various plant produced RNAs are converted into sRNAs after ingestion. However, very few seem to associate with regulatory complexes, something that is affected by feeding behavior and prior exposure to dsRNA. This suggests a scenario where foreign RNAs appear to transit the spider mite RNAi pathway but are diverted from regulatory complexes, which is modulated by the status of the mite physiology.

3.2 Method

3.2.1 Mite handling

Mites were reared on two plants: *Phaseolus vulgaris* or *A. thaliana*. Mites were removed from plant leaves by gentle tapping and collected into a microfuge tube. Both juveniles and adults were kept. For column experiments, a volume of 100 µl packed mites was used, which is approximately 500 animals. Mite soaking was done by adding 200 µL of a 160 ng/µL RNA solution to collected animals. Soaking lasted overnight (15-16 hours) followed by rinsing in PBS (pH 7.4), and 0.1% Tween 20 solution. Afterwards, mites were

collected on a paper towel and allowed to dry before being placed onto a *P. vulgaris* leaf inside of a large Petri dish at room temperature. After five days, mites were collected, flash frozen and stored at -80˚C.

3.2.2 RNA synthesis

For the initial experiment, two dsRNAs were generated from cloned fragments of *T. urticae* actin and Green Fluorescent Protein (GFP) from *Aequorea victoria* (accession numbers CAEY01002033.1 and FJ172221.1, respectively). Long actin and GFP dsRNAs were both approximately 350-nt long and created using the MEGAscript™ in vitro transcription kit (Ambion). Templates for dsRNA synthesis were PCR products amplified with primers encoding T7 RNA polymerase promoter sequences on their 5' ends. Following in vitro transcription, lithium chloride precipitation was performed to purify products, and the resulting RNAs were annealed through heating and gradual cooling.

Four synthetic structures were designed based on RNAs shown to enter miRNA pathways. Templates for these structures were created by annealing long, ~100 nt, oligonucleotides. The resultant dsDNAs encoded a T7 promoter at the 5' end followed by the structured RNA sequence. RNAs were synthesized using the MEGAshortscript™ in vitro transcription kit (Ambion). The synthesis products were purified by phenol chloroform extraction, and ethanol precipitation.

3.2.3 HiTrap Q FF column enrichment

Column enrichment of sRNA containing complexes used 1 mL HiTrap Q FF columns (GE Life Sciences) as previously described (N. C. Lau et al., 2006). Briefly, columns were equilibrated following manufacturer instructions. First, 5 mL of start buffer (20 mM HEPES-KOH, pH 7.9) was applied and passed through the column (1 mL/minute).

Then 5 mL of elution buffer (20 mM HEPES-KOH, pH 7.9, 1 M NaCl), and 10 mL start buffer (20 mM HEPES-KOH, pH 7.9, 100 mM KOAc) were applied sequentially. Collected animals were washed several times with PBS, pH 7.4, followed by flash freezing and grinding with a mortar and pestle. 1 mL chilled binding buffer (20 mM HEPES-KOH pH 7.9, 100 mM KOAc, 0.2 mM EDTA, 1.5 mL $MgCl_2$, 10% glycerol, 0.2% PMSF, 1 mM DTT, 1X Roche EDTA-free protease inhibitor cocktail) was mixed with the pulverized animals. The lysate was then clarified by centrifugation. Cleared lysate was applied to the column and passed at a flow rate of 1 mL/minute. The column was washed with binding buffer followed by elution buffer (binding buffer with 300 mM KOAc). An equal volume of acid phenol-chloroform was added to the tube, followed by rocking at room temperature. Following phase separation, RNAs were isopropanol precipitated and resuspended in 30 µL of ddH_2O. Total RNAs extracted from mites used TRI reagent and followed manufacturer protocols.

3.2.4 RT-qPCR

Total RNA from spider mite samples was used for cDNA synthesis (Thermo Fisher T-7 kit) using random hexamer primers. cDNAs were used in qPCR assays with SYBR Green real-time PCR master mix (Thermo Fisher), following manufacturer protocols. In the qPCR assay, actin transcripts were normalized by assessing levels of 18S ribosomal RNA. Primer sequences for actin and were previously published (Suzuki et al., 2017a).

3.2.5 RNA Sequencing and Analysis

RNA obtained from both column extraction and total RNA extraction were sequenced from an Illumina TrueSeq Small RNA Library Prep Kit library. Sequencing occurred on an Illumina Nextseq500 or Miseq machine using a single-read 50 base pair

protocol. Raw sequence data were initially processed using Fastx toolkit to remove adapter sequences (Hannon, n.d.; Langmead et al., 2009). Libraries were normalized for number of reads by subsampling with the Seqtk program. For experiments looking at dsRNA processing 50M reads were used when comparison of plant derived RNAs 14M were used. Mapping of reads to mite and plant genomes used the Bowtie program with -a -v0 -m200 parameters to find all perfect alignments while also excluding extremely low complexity sequences (Langmead et al., 2009). Mapping exclusively to actin and GFP sequences used -v0 for only perfect matching RNAs, and chloroplast alignments used -v 0 -a to uncover all perfect alignments.

Alignments were then processed with Samtools and Bedtools to find regions of expression based on read density and merging of juxtaposed features (H. Li et al., 2009; Quinlan & Hall, 2010). Bedtools was also used to quantify alignments per feature and determine the intersection of genomic loci with library alignments. All mite and plant sequences as well as their annotations were taken from public databases (Antoniewski, 2014). Size distributions were calculated from alignments converted back to "fastq" format, which were then parsed to determine number of reads corresponding to lengths between 15-3-nt long.

A Python-based algorithm was used to find overlapping read pairs that represent Dcr cleavage produced sRNA duplexes. sRNAs of 15-31 nts were used to query target pairs, also 15-31 nts long, that exhibited an overlap of 2 nts less than the size of query RNA. The number of reads for each pair was plotted to show a heatmap matrix. Visualizations were created using the R packages ggplot2, qqman, Pheatmap, and Sushi (D. Turner, 2018; Kolde, 2012; Phanstiel et al., 2014; Wickham, 2009). All mite

sequencing data can be accessed for the BioProject ID: PRJNA591169. Public libraries used for plant mapping are: SRR7738374 (*P. vulgaris*), and SRR7947145 (*A. thaliana*).

3.3 Results

Following previously reported methods, we sought to induce maximum gene silencing by soaking animals in dsRNA target to spider mite actin and GFP. Following exposure to dsRNA, animals were processed into a lysate and sRNAs were isolated with the Hi Trap approach described above. Following sequencing, sRNA reads were aligned to actin and GFP, which showed significant accumulation of 18-21 nt RNAs at the target region (Fig 3.1A). For untreated conditions from Hi Trap enrichment or total RNA the accumulation was absent with very few alignments in the enriched sample and many apparent degradation fragments for total RNA (Fig 3.1A).

We then sought to verify Dcr processing of long dsRNA-derived siRNAs by examining overlaps between complementary pairs of sRNA reads (Grbić et al., 2011). Specifically, the abundance of sRNA pairs with 2nt overhangs, characteristic of RNase III cleavage, were summed and plotted in a matrix of pair lengths (Fig 3.1B-C). This reveals the sizes of sRNAs found in duplexes that have a signature of Dcr cleavage. In total ~45% of actin mapping reads showed signs of Dcr activity consistent with the centrality of this enzyme in the processing of dsRNA, and further validates the purification strategy for examination of functional siRNAs. Common pairs were the expected 18-21 nt length and tended to be offset by 1-2nts (Fig 3.1B-C). Actin siRNAs showed more diversity in pair lengths compared to GFP siRNAs, suggesting capture of target cleavage products by RNAi pathways occurs when a target transcript is present. We also observed many antisense reads mapping outside the target region that were not present in the total RNA library. This

potentially could be the result of RdRP activity producing antisense transcripts in response siRNA targeting (Fig 3.1D). We then inspected 5' end sequence biases of the actin read in the dsRNA flanking regions (Fig 3.1E-F). While there is little sequence similarity in sense mapping reads, antisense have clear 5' end "T" and "A" 10 position. This is highly suggestive of piRNAs, which exhibit these features because of ping-pong processing. In spider mite, siRNAs and piRNAs cooperate for genome surveillance (Dalmay et al., 2000). These results suggest that similar event may extend to siRNAs derived from exogenous sources, which likewise appear to be able to trigger piRNA production.

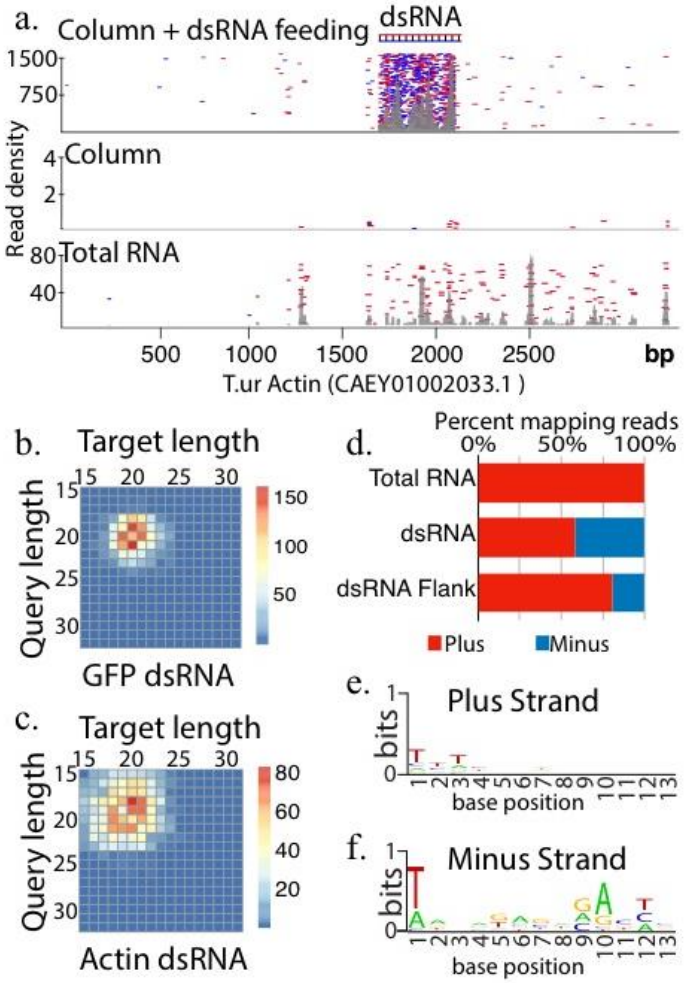


Figure 3.1 *Fate of ingested synthetic long dsRNA*

Figure 3.1 (continued)

(A) Alignment of reads to the Actin mRNA (CAEY01002033.1) sequenced from column purified RNAs following feeding with dsRNA (top panel), no feeding after column purification (middle panel), and total RNA from unfed mites (bottom panel). Red dashes are positive strand mapping reads, while blue represent negative strand mapping reads. Grey, overlaid density plots represent total read accumulations. B and C show number of Dcr sRNA sequencing read pairs of different sizes derived from dsRNA targeted to GFP, B or Actin, C. Pairs are defined by overlapping by 2nt less than the length of the query strand of the pair. This simulates the 2nt overhangs found after Dcr processing of dsRNA. Duplexes are biases toward short (18-22nt) pairs that show asymmetry by one or two bases, D. Percent of plus or minus strand mapping in Total RNA (from bottom panel of A), dsRNA targeted region following feeding and column purification (top panel A), or flanking region when dsRNA is fed, and column purification is used (top panel A). E and F are seqlogo plot showing sequence bias for reads mapping to dsRNA flanking region (top panel A) (E) or minus strand (F) of Actin.

In a separate analysis, members in our group showed that structured plant derived RNAs enter spider mite RNAi pathways (Mondal et al., 2021). On this basis we investigated the activity of short hairpin RNAs that mimic distinct, known miRNA-type biogenesis (Fig 3.2A). The ability of four different configurations of synthetic hairpin RNAs to trigger silencing of Actin was tested alongside long dsRNA also targeted to Actin by soaking in *in vitro* synthesized molecules. qPCR was used to quantify target knockdown (Fig 3.2B). The short hairpins were designed to transit canonical miRNA biogenesis (shRNA), Ago processed short loop RNAs that mimic miR-451 with and without a G-C clamp at the hairpin base (SL1 & SL2), and a G-C clamp stabilized RNA with a loop sequence complementary to actin (Loop) (Winter et al., 2013; Yang et al., 2010). GC clamps were added to the structure bases of the RNAs to increase stability as this high energy fold is a challenging substrate for RNases (Kowalinski et al., 2016). After soaking, relative abundance of actin transcripts was assessed showing the reported ~40% reduction for long dsRNA (Suzuki et al., 2017b). Three of the structured RNAs showed a greater degree of knockdown with a reduction of transcript accumulation reaching nearly 60%, suggesting that the siRNA pathway of *T. urticae* may not be the optimal mode of RNAi to exploit for gene silencing (Fig 3.2B).

We also sequenced samples after feeding with the SL2 (Fig 3.2C) and shRNA (Fig 3.2D) RNAs. SL2 derived sRNAs were more abundant but showed smaller fragments than shRNA, consistent with being an atypical RNAi substrate (Fig 3.2E-F). For both RNAs, however, cleavage sites were non-uniform, and for SL2 there was little evidence slicer-mediated precursor processing was occurring as seen with miR-451, but rather standard Dcr processing, suggesting that additional engineering could yield an even more superior gene silencing molecule. Nevertheless, many reads were complementary to Actin that were able to trigger knockdown. This together with the appearance of long dsRNA-derived siRNAs shows that a variety of ingested RNAs enter spider mite RNAi pathways, and that the designated siRNA/long dsRNA pathway may not be the most potent option for eliciting knockdown.

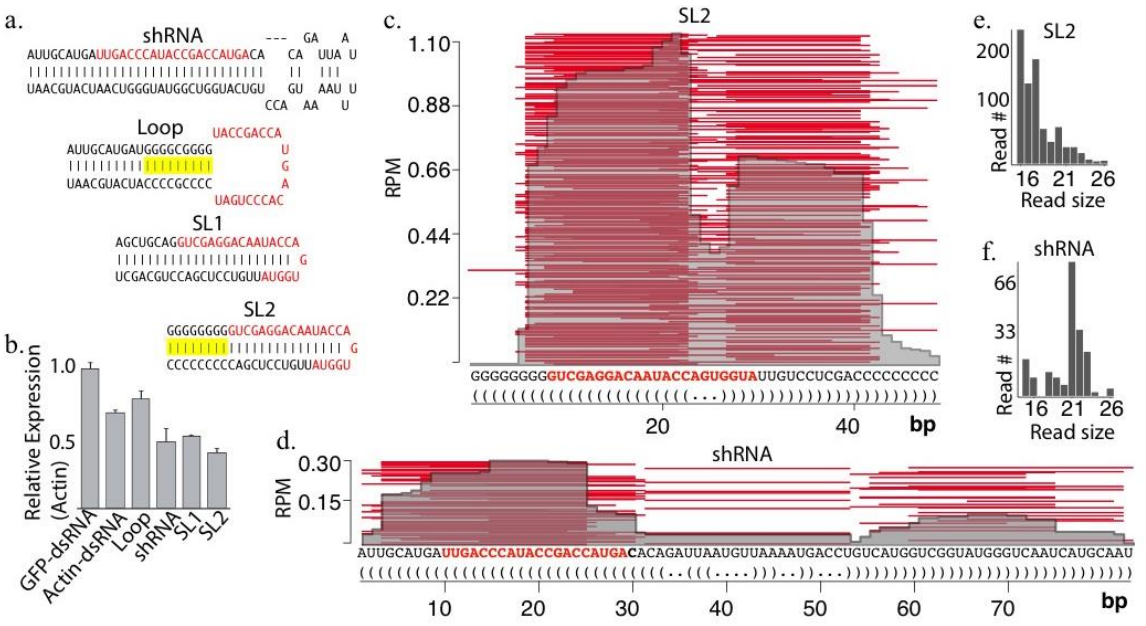


Figure 3.2 *Fate of synthetic short hairpin RNAs fed to spider mites.*

(A) Actin targeted structured short hairpin RNAs fed to animals via soaking. Designs mimic modes of miRNA biogenesis: Dcr processed (shRNA), AGO processed short loop RNAs (SL1, SL2), and loop derived RNAs (Loop). Red sequence indicates portion targeted to Actin. Yellow highlights show G/C "clamp". B shows Percent expression of target gene (Actin) following ingestion of long

Figure 3.2 (continued)

double-stranded RNA (dsRNA) or short RNAs in A determined by quantitative real time polymerase chain reaction (qRT-PCR) and normalized to rRNA. Error bars are standard deviation. C-D are Alignment data of RNAs sequenced from animals fed SL2, C or shRNA, D to the synthetic RNA sequences. Red sequence shows portion targeted to Actin. E-F are Absolute number of reads mapping to short hairpin RNAs SL2, E and shRNA, F.

3.4 Discussion

Using the strategy described here we were able to provide a comprehensive view of environmental RNA processing in the two-spotted spider mite. Several aspects of sRNA biogenesis emerged that are not found in model organisms to include involvement of ping pong piRNA-like processing of transcripts targeted by dsRNA (Mondal et al., 2020). Weak RdRP potential activity was also observed, however, it did not yield a significant population of siRNAs. This was apparent by the lack of transgenerational siRNA inheritance. We were also able to show single-strand structured RNAs are processed by Dcr into sRNAs, which guided the design of several short hairpin RNAs that have greater gene silencing efficacy compared to long dsRNA.

Our analysis has important implications for RNAi strategies for mite plant pests. While we found a wide range of ingested RNAs can access mite RNAi pathways, phenotypes are hard to induce in these animals with RNAi. We propose that this apparent contradiction may indicate that RNAi pathways in mites have another function: metabolism of double stranded or structured RNAs. Identity of transcripts converted into sRNAs by mites is highly dynamic and tied to feeding state. Moreover, for an animal with hundreds of plant hosts, conversion of ingested RNAs into multi-target gene regulators would be maladaptive. Preferential destruction of sRNAs based on origin is seen in *D. melanogaster* (Flynt et al., 2009; Westholm et al., 2012). The mirtron subclass of miRNAs as well as siRNAs derived from latent viruses are poorly recruited to Ago complexes (Flynt et al.,

2009; Westholm et al., 2012). Navigating such a fate for sRNAs will be critical to effective RNAi technology.

The approach we describe here is species agnostic. There are many species that would be excellent candidates for new pest control technologies where RNAi efficacy is unclear. By evaluating a range of substrates both synthetic and dietary using the strategy we applied to mites, better approaches could be developed. The synthetic hairpin RNAs that showed greater activity in spider mites are likely processed via miRNA biogenesis a much more conserved mechanism than arthropod siRNAs. A similar type of RNAi trigger may be a better choice for animals that have apparent resistance to long dsRNA. The biochemical and computational approach here would confirm this is situation in one of these species.

3.5 Conclusion

This chapter documents steps that can be taken to optimize sRNA selection for dsRNA design purposes to potentiate RNAi. Also, processing of exogenous RNAs and in the two-spotted spider mite is covered here. Insights were possible for pairing a species agnostic biochemical method with sequencing analysis that documented sRNA biogenesis patterns. From this, we were able to observe environmental RNA processing affected by possible RdRP activity and integration of siRNA targeting with piRNA processing. siRNAs acting upstream of piRNAs had been previously observed to occur with endogenous pathways. Also, by characterizing dietary RNAs from small structured RNAs were found to produce significant knockdown by exploiting miRNA function and not siRNA. These can lead to a better-informed design of RNAi for mite pest species and

provides a framework for evaluating environmental RNA processing in other non-model organisms.

CHAPTER IV – SPORADIC ACTIVITY OF AMPHIOXUS RDRP IS ASSOCIATED WITH SOMATIC GENOME MAINTENANCE THROUGH RNAI

4.1 Abstract

RNAi is a pervasive gene regulatory mechanism in eukaryotes based on the action of multiple classes of sRNA. Some organisms incorporate RdRPs activity into RNAi pathways either through amplifying precursors or by directly generating sRNAs. While this enzymatic activity has prominent roles in plants and fungi, involvement of RdRP in RNAi pathways of many animals is controversial. In this study, we investigate contribution of RdRP to sRNA generation in *B. floridae*, a species of amphioxus. With RdRP transcripts expressed in a variety of organs and tissues and at different stages of development, we sought to detect RdRP activity by inhibiting RNA Pol II's activity and analyzing potentially RdRP synthesized nascent RNAs. Using this approach, several loci were found to give rise to siRNAs generated from Pol II independent transcripts, potentially being derived from RdRP activity. The putative RdRP derived siRNAs were observed to target genes and mobile genetic elements. The low level of RdRP activity reported here suggests a scenario where RdRP function has become marginal, consistent with their loss in several animal lineages such as vertebrates and explains the challenges of detecting a role in creation of dsRNAs. This observation shows a framework for how an essential gene regulation mechanism slowly decays on its way to being lost.

4.2 Relevance

RNAi (RNAi) mechanisms are widespread in eukaryotes, and in some animals is made more complex by a family of RdRP. Although present in many eukaryotes, a function for RdRPs have not been identified in animals other than *C. elegans*. Amphioxus species,

a group of basal chordates, encodes the most RdRP homolog of all higher animals. However, analysis of sRNAs in the amphioxus *Branchiostoma lanceolatum* (*B. lanceolatum)* reported RdRPs in this group are unlikely to participate in RNAi. Given the intimate connection between eukaryotic RdRPs and RNAi, it is likely that at least a remnant role in RNAi may remain for amphioxus RdRPs.

Further, amphioxi occupy a very important position in animal phylogeny making them the closest living relatives of vertebrate ancestors. As a result, numerous studies on the evolution of higher chordates have been done with these animals to explore evolution and genome wide gene duplication events observed in different vertebrate genomes (Dehal & Boore, 2005). Due to the association of eukaryotic RdRPs in the siRNA partway especially as relating to TE control, studies to unravel the role of these proteins may reveal a new perspective about how new genes possibly evolved from vertebrate ancestors. To this end, in this chapter, we re-examined the involvement of RdRP in the amphioxus species, *B. floridae*, through curation of nascent transcripts in Pol II inhibited animals. We identified 21 loci that show nascent transcript generation when Pol II is inhibited. The sRNAs produced shows complementarity primarily to mobile genetic elements as well as some genes suggesting a potential role for amphioxus RdRPs in RNAi-mediated genome surveillance.

4.3 Introduction

RdRPs contribute to RNAi (RNAi) pathways in many eukaryotes, with roles in regulating genic transcripts, TEs, and viruses. While function of these proteins has been well-established in fungi and plants, their role in animals, other than nematodes, is relatively unexplored. In plants and fungi, RdRPs participate in RNAi through production

of long dsRNA that becomes a substrate of Dcr for processing into siRNAs. In *Arabidopsis thaliana* (*A. thaliana*) RdRPs, SGS2 and SGS3 are implicated in PTGS during anti-viral defense while RdR2 mediates RNA directed DNA methylation and histone modification at telomeres (Mourrain et al., 2000; Willmann et al., 2011a). Fungal RdRPs such as QDE1 have been implicated in control of genic transcripts and TEs through siRNA production (Cogoni & Macino, 1999; L. Li et al., 2010).

The *C. elegans* genome encodes four RdRP genes: *Ego-1, rrf-1, rrf-2,* and *rrf-3*. *Rrf*-3 behaves like the RdRPs in plants and fungi, producing dsRNAs which become Dcr substrates (Aoki et al., 2007). The other three (*Ego*-1, *rrf*-1, and *rrf*-2) produce siRNAs directly through a Dcr independent (*de novo*) mechanism, which appears restricted to nematodes. *De novo* siRNA producing RdRPs possess a proline-rich loop that facilitates non-processive synthesis of 5' tri-phosphorylated siRNAs (Sarkies et al., 2015b). In addition, *de novo* synthesis does not require priming due to the presence of an extra loop rich in aromatic residues found in the RdRPs which facilitate a stronger template interaction by enabling the initiating guanine (G) to be placed at the active site (Ferrer-Orta et al., 2006). While little is known about the behavior of individual genes, apparent RdRP-mediated amplification of siRNA for gene silencing has been observed in chelicerate arthropods such as mites and ticks as well (Mondal et al., 2018; Sarkies et al., 2015b).

Ancient eukaryote progenitors are thought to have had three groups of RdRPs; RdRP-α, RdRP-β and RdRP-γ, with the α and β groups present in animals (Zong et al., 2009b). All metazoan superphyla have independently loss RdRP (Zong et al., 2009b). For instance, RdRPs are encoded in the cephalochordates and urochordates, but are absent in vertebrate genomes. Also, in the Spiralia, RdRP exists only in bivalves, but is otherwise

absent in this group. Likewise, in ecdysozoan Chelicerae arthropods possess RdRP, while the remainder of species belonging to Pancrustacea do not. In addition to the distribution of RdRP in animals, the number of homologs in different lineage is also labile. Cnidarians possess at least four homologs (only two possess complete RdRP domain and active site motif) in *N. vectensis*, chelicerates five, and hemichordates three. Cephalochordates have the greatest number of RdRPs with six homologs all of which possess active site motif and four having full length ERD at least in *B. floridae*. These complex conservation patterns create difficulty in identifying shared functions.

A recent study investigated the involvement of RdRP in generating sRNA species in *B. lanceolatum* and found no concrete evidence of contribution to RNAi (Pinzón et al., 2019b). Here we re-examine RdRP function in a different amphioxus species, *B. floridae* through searching for evidence of RdRP product derived siRNAs. We utilize metabolic labelling followed by NGS. This technique uses 4-thiouridine (4sU) to label nascent transcripts after inhibition of Pol II activity with alpha-amanitin. We observed regions with nascent siRNAs produced in Pol II inhibited animals, suggesting the possibility of RdRP processivity. Upon further query, we found that majority of the siRNAs target TEs suggesting a role in transposon control for RdRP produced siRNAs in amphioxus.

4.4 Materials and Methods

4.4.1 Protein sequence identification

Protein sequences were obtained from the following databases: NCBI (https://www.ncbi.nlm.nih.gov), ENA (https://www.ebi.ac.uk/ena/browser/) and WormBase (wormbase.org). Predicted *B. floridae* proteins were obtained from NCBI after performing local Blast (tblastn) with protein orthologs from *C. elegans*. Gene IDs were

identified from genomic loci of the protein orthologs in *B. floridae* using integrative genome viewer (IGV) or through NCBI Blast search. *B. lanceolatum* RdRPs were obtained through Hmmer search on the ensemble website (https://metazoa.ensembl.org) by using RdRP IDs from Pinzon *et al*., 2019 (Pinzón et al., 2019b). All amino and nucleic acid sequence IDs associated with this study are available in Table 4.1. Further, RNAi factors in amphioxus alongside other eukaryotes are shown in scheme 4.1.

Table 4.1 *Protein IDs associated with this chapter*

S/N	Organism	Protein	Protein ID	Source
1	*O. dioica*	RdRP-1	CBY14941	NCBI
2	*O. dioica*	RdRP-2	CBY12568.1	NCBI
3	*B. floridae*	RdRP-1	XP_035694482.1	NCBI
4	*B. floridae*	RdRP-2	XP_035692999.1	NCBI
5	*B. floridae*	RdRP-3	XP_035696221.1	NCBI
6	*B. floridae*	RdRP-4	XP_035683931.1	NCBI
7	*B. floridae*	RdRP-5	XP_035694481.1	NCBI
8	*B. floridae*	RdRP-6	XP_035673963.1	NCBI
9	*C. elegans*	Ego-1	CE27140	WormBase
10	*C. elegans*	*rrf*-1	CE46152	WormBase
11	*C. elegans*	*rrf*-2	CE23872	WormBase
12	*C. elegans*	*rrf*-3	CE45624	WormBase
13	*N. crassa*	QDE-1	XP_959047.1	NCBI
14	*N. crassa*	RdRP	CAD70515.1	NCBI
15	*N. crassa*	OR74A	XP_964248.3	NCBI
16	*A. thaliana*	RdRP-1	AAF79241.1	NCBI
17	*A. thaliana*	SGS2	AAG52184.1	NCBI
18	*A. thaliana*	RdRP-2	NP_192851.1	NCBI

Table 4.1 (continued)

19	*A. thaliana*	RdRP-3	NP_179581.2	NCBI
20	*A. thaliana*	RdRP-4	NP_179582.2	NCBI
21	*A. thaliana*	RdRP-5	NP_179583.3	NCBI
22	*A. tenebrosa*	RdRP-1	XP_031551119.1	NCBI
23	*A. tenebrosa*	RdRP-2	XP_031551081	NCBI
24	*A. tenebrosa*	RdRP-3	XP_031551118.1	NCBI
25	*A. tenebrosa*	RdRP-4	XP_031551102.1	NCBI
26	*A. tenebrosa*	RdRP-5	XP_031572042.1	NCBI
27	*A. tenebrosa*	RdRP-6	XP_031556213.1	NCBI
28	*A. tenebrosa*	RdRP-7	XP_031560564.1	NCBI
29	*A. tenebrosa*	RdRP-8	XP_031554636.1	NCBI
30	*B. lanceolatum*	RdRP3	BL09945	Ensemble
31	*B. lanceolatum*	RdRP4	BL23385	Ensemble
32	*B. lanceolatum*	RdRP5	BL27717	Ensemble
33	*B. lanceolatum*	RdRP6	BL02069	Ensemble
34	*B. belcheri*	RdRP1	XP_019640764.1	NCBI
35	*B. belcheri*	RdRP2	XP_019640931.1	NCBI
36	*B. belcheri*	RdRP3	XP_019645817.1	NCBI
37	*B. belcheri*	RdRP4	XP_019641825.1	NCBI
38	*B. belcheri*	RdRP5	XP_019640931.1	NCBI
39	*B. belcheri*	RdRP6	XP_019625576.1	NCBI

	RNAi Factors			
Animal	Dcr	AGO	PIWI	RdRP
Human	1	4	4	0
Flies	2	2	3	0
Worms	1	27	2	4
Amphioxus	1	2	3	6

Scheme 4.1 *RNAi factors in humans, fruit fly, worms, and amphioxus*

4.4.2 Renaming Amphioxus RdRPs for universal identification

To avoid confusion in naming and due to the absence of a standard nomenclature for RdRPs in this clade, we attempted to rename the RdRPs for the sack of consistency in this study as well as future studies. Using RdRPs reported by Pinzon *et. al.,* 2019 (Pinzón et al., 2019a) as a benchmark for alignment and protein ID naming, we have renamed the NCBI (for *B. floridae* and *B. belcheri*) and ensemble (*B. lanceolatum)* IDs to specific RdRPs (1-6). Also, for the sake of this study, other RdRPs named here have been assigned their respective gene IDs as shown in Table 4.2.

Table 4.2 *List of renamed RdRP Protein IDs used in this chapter*

RdRP #	*B. floridae*	*B. lanceolatum*	*B. belcheri*	*N. vectensis*	*O. dioica*
RdRP1	XP_035694482.1	BL19289	XP_019640764.1	XP_032240568.1	CAG5098211.1
RdRP2	XM_002596781.1	BL09945	XP_019640931.1	XP_032236239.1	CAG5087879.1
RdRP3	XM_002595575.1	BL23385	XP_019645817.1	XP_032234096.1	CAG5094182.1
RdRP4	XM_002595574.1	BL27717	XP_019641825.1	XP_032228801.1	
RdRP5	XM_002589548.1	BL07831			
RdRP6	XM_002591531.1	BL02069			

4.4.3 Phylogeny, domain architecture, and prediction of RdRP processivity

Verification of eukaryotic RdRP domain (ERD) in RdRP homologs from several animals was done using NCBI's conserved domain prediction tool as well as the protein family database tool, pfam (http://pfam.xfam.org). *B. floridae* RdRPs alongside RdRPs from *B. belcheri, B. lanceolatum, O. dioica, and N. vectensis* were used to construct a phylogenetic tree using default parameters of the MEGA11 software. Due to poor annotation of a dated *B. lanceolatum* genome, the previous work reported sequences that were incomplete. Giving a sense that only two of six amphioxus RdRPs have full length ERD domains. The recently assembled *B. lanceolatum* genome with 482 scaffolds enabled us to locally blast and obtained RdRP homologs with improved prediction for our analysis. 23 nucleic acid sequences (from *B. floridae*, *B. belcheri*, *N. vectensis,* and *O. dioica*) were aligned with MUSCLE.

Evolutionary tree was inferred using the general time reversible (*GTR*) model and gamma distribution with invariant sites (*G+I*) for rate among site estimation, phylogeny was estimated through maximum likelihood method with 500 bootstrap replications (Stecher et al., 2020; Tamura et al., 2021). Using the GenomeNet database

(https://www.genome.jp), processivity of *B. floridae* RdRPs were predicted through multiple sequence alignment (Clustal W). For this analysis, the default option was used with pairwise alignment set to slow and accurate. Non-processive motif found in some worms RdRPs were identified from the amino acid alignment to predict the nature of second strand synthesis in this animal.

4.4.4 Prediction of subcellular localization of eukaryotic RdRPs

Following ERD identification in *B. floridae* RdRPs, RdRP families that have the eukaryotic RdRP domain were tested for amino acid signals that enhance their localization within the cell. Prediction of potential subcellular locations where the proteins are likely to traffic more is possible using these signals. To begin, we obtained RdRP amino acid sequences from the following eukaryotic phyla: Urochordata (*O. dioica*), Cephalochordata (*B. floridae*), Nematoda (*C. elegans*), Cnidaria (*N. vectensis*), plants (*A. thaliana*). These were then queried for localization signals using the subcellular protein localization tool, WoLF PSORT (https://wolfpsort.hgc.jp) (Horton et al., 2007). In addition, we also seek to validate the predictions obtained from the tool by adding human GAPDH and histidine (H3) sequences to the analysis.

Following this, we seek to understand the enzymatic properties of each of *B. floridae*'s RdRPs. Through an open-source deep learning tool (http://www.cbrc.kaust.edu.sa/DEEPre) that predict enzyme function based on amino acid sequences, we analyzed the amino acid sequences to understand the enzymatic activity of *B. floridae*'s RdRPs (Y. Li et al., 2018).

4.4.5 Estimation of RdRP abundance in amphioxus transcriptome

RdRP transcript abundance across several stages of *B. floridae* development as well as several *B. lanceolatum* tissues were determined using the following approach. Publicly available transcriptomic datasets (RNA-sequencing datasets) generated from 32-64 cell stage, blastula, gastrula, neurula, larva, juvenile, adult male, and adult female *B. floridae* were obtained and each mapped to bowtie2 indexes of *B. floridae* RdRP mRNA. To observe the possibility of a tissue specific RdRP activity in amphioxus, transcriptomic datasets originating from epidermis (skin), muscle, gill, male, and female gonads of B. lanceolatum were obtained from NCBI and mapped to *B. lanceolatum* RdRP mRNA indices (Marlétaz et al., 2018). To obtain transcript abundance, the mapped reads from each stage was separately intersected with an annotation bed file of the six *B. floridae* RdRPs. The same steps were repeated with the transcriptomic data from the above B. lanceolatum tissues for RdRP 3-6. Further, primers were design to amplify the 6 RdRPs in the gill, and notochord of male *B. floridae* through reverse transcriptase PCR (RT-PCR) to verify transcript production (Table 4.3).

Table 4.3 *List of primers used in this chapter*

Primer name	Primer sequence (5' - 3')
B. floridae-RdR1qPCRf	CTCACCACAGAGCCGTTCTTC
B. floridae-RdR1qPCRr	CATGATGCGGCCATATTCTGGTG
B. floridae-RdR2qPCRf	CGTCACGTCAAGTCAAAGGAGG
B. floridae-RdR2qPCRr	GCCGTACCGGATTTGGAAGG
B. floridae-RdR3qPCRf	GGTCATCGTCACGCCGACAAG
B. floridae-RdR3qPCRr	GTCCTCGTCTCGGAACACCAC
B. floridae-RdR4qPCRf	GACAAGTTCAGGGACGCACAC

Table 4.3 (continued)

B. floridae-RdR4qPCRr	CGTACCGGATTTGGAAGGCAG
B. floridae-RdR5qPCRf	GTCACCACAGAACCGTTCTTCAG
B. floridae-RdR5qPCRr	CCAACATGTTCCGTCCGTACTC
B. floridae-RdR6qPCRf	CCGGACAGTGCTTCTTCCAAG
B. floridae-RdR6qPCRr	GGATGGTAGCACGGACACTTC
B. floridae-rRNAf	CCTGCCAGTAATCATATGCTTGTC
B. floridae-rRNAr	GCCACAGTTATCCATGTAGGATG

4.4.6 Global distribution of sRNA in Amphioxus

Adult male and female *B. floridae* were obtained from the Gulf Specimen Marine Laboratories (https://gulfspecimen.org) and total RNA was separately extracted from whole animal tissues using TRIzol RNA extraction method followed by isopropanol precipitation. RNA qualities were determined with the 2100 Agilent Bioanalyzer System. Small RNA libraries were generated and sequenced using the Illumina Nextseq 500 sequencing system. Sequenced libraries were clipped to remove adapter sequences and reads having sizes between 18 to 34 nucleotides were selected. Matured *B. floridae* microRNA (miRNA) sequences were obtained from the miRNA database (mirbase.org). They were then indexed, and sRNA libraries mapped to them with Bowtie to eliminate miRNAs (reads that mapped). All analysis associated with sRNAs in this study were performed using reads that failed to map to the known miRNA. The resulting cleaned reads from male and female libraries were then normalized to reads per million (RPM). In addition, read sizes in each library were characterized. Using the same approach as above, size distribution was estimated for sRNA libraries from female *B. belcheri* (*B. belcheri*) tissues generated in a previous study (Liao et al., 2017).

4.4.7 Dcr signatures and siRNA production centers in amphioxus

B. floridae Male and female libraries as well as the libraries from *B. belcheri* tissues; gill, skin, intestine, notochord, and ovary, were separately analyzed for overlapping read pairs with 2-nt 3' Dcr overhangs. First, the reads from the libraries were mapped to the respective reference genomes after which mapped reads were subject to an overlap query that uses a set of query strands to overlap a set of target strands in the same locus to produce read pairs that overlap perfectly with 2nt 3' overhang on each strand. This analysis was repeated for both adult and tissue libraries by first, calling loci where read abundance exceed a certain threshold (locus length > 200, read depth > 10) to identify all loci with high sRNA activity. All recovered loci potentially generating sRNAs were grouped into piRNA-like (25-30 nt) or siRNA-like (20-23nt) loci depending on indicated size and presence or absence of ping-pong or Dcr signatures respectively. All loci obtained based on the filtering condition were then visualized in IGV and candidate loci that appear to be sequencing artifacts were discarded.

4.4.8 Somatic sRNA quantification

To observe the distribution of sRNAs across loci with high Dcr activity in the tissues studied here, miRNA depleted tissue libraries were mapped to the genome and queried for overlapping read signatures. Using the filtration method described above, all loci where high read pair activity like Dcr processing were recovered and siRNA producing loci compared between tissues to access whether potential RdRP function can take place across tissues.

4.4.9 Metabolic labeling of nascent transcripts

To observe potential RdRP activity in amphioxus, we inhibited RNA Polymerase II (Pol II) activity and labelled nascent transcripts as follows: first, two males were placed in separate Falcon tubes containing 3.5mL filtered sea water. To the first tube, 20 µl 1mg/mL α-Amanitin was carefully added in a biosafety hood and incubated for 10-15 minutes to halt Pol II associated nascent RNA synthesis. Afterwards, 70 µl of 100 mM 4-thiouridine (4sU) was added to each tube (one with α-amanitin and the other without α-amanitin) for 40 minutes. Whole animals were then briefly placed on ice, and quickly rinsed 3-4 times with filtered salt water, cut into segments with a sterile blade before tissue homogenization. Total RNAs were extracted using TRIzol followed by isopropanol precipitation, quantification, sRNA library preparation and sequencing on the Illumina Nextseq 500 platform.

Following sequencing, adapter sequences were clipped from reads prior to mapping to a modified *B. floridae* reference genome (GCF_000003815.2) in which all thymidine were intentionally changed to cytosine's (T to C conversion). This was done to properly map nascent transcripts that have 4sU bases sequenced as Cs rather than Ts to the genome. Based on HE loci reported in Table 4.4, mapped reads were then analyzed in integrative genome viewer (IGV) for nascent transcript against the intact reference genome.

To further inspect specific characteristics of the predicted RdRP associated reads in one locus reported here as a potential RdRP locus with evidence of RdRP activity (NC_049996.1:5328725-5330257), reads from these loci were obtained from control male body library and assessed for alignment properties such as unique mapping or multi-

mapping as well as read size abundance. Furthermore, we queried the reads from these loci for Dcr signatures as well as individual nucleotide abundance across all reads.

4.4.10 Prediction of RdRP dependent sRNA targets

To predict the function of sRNAs that were recovered from the metabolic labelling experiment above, we sort to identify potential targets of RdRP associated siRNAs from the 21 predicted loci in *B. floridae*. First, coordinates corresponding to the 21 predicted loci were collected and sRNAs were mapped to the genome. Following this, genome matching reads were either mapped to indexes of *B. floridae* TEs, and several viral genomes that infect or are related to amphioxus herpesvirus or intersected against annotated genes. Amphioxus TEs were separated into seven categories: DNA, LTR, SINE, LINE, low complexity, simple repeats, and unknown. *B. floridae* Adintovirus genome along with OsHV-1 and abalone Herpesvirus which had been reported to be closely related to amphioxus associated Herpesvirus were separately indexed. Antisense reads were then mapped to the indices using default bowtie settings (Savin et al., 2010). Hits were obtained for all targets and duplicates removed. For gene targets, only reads that mapped to the genome with a single mismatch were used.

4.5 Results

4.5.1 RdRPs in Amphioxus are processive

To further explore RdRP function in amphioxus, we first compared conservation of this protein family in the three sequenced species: *B.la, B. belcheri,* and *B. floridae*. While the current *B. floridae* genome contains only 432 scaffolds (contig N50 = 50,413), *B. belcheri* and the older *B. lanceolatum* genomes for which the previous RdRP associated study is based respectively have 2308 and 10,247 scaffolds suggesting that *B. floridae* has

a superior genome assembly. As a result, we selected *B. floridae* as the model for our study. Prior work in *B. lanceolatum* found six homologs, however, some appear incomplete, potentially due to an incomplete genome assembly. We curated six RdRPs in *B. floridae*, consistent with prior reports in other lancelets. Only five have complete eukaryotic RdRP domain (ERD) (PF05183) though all contain an RdRP active site motif (DxDGD) (Fig 4.1A).

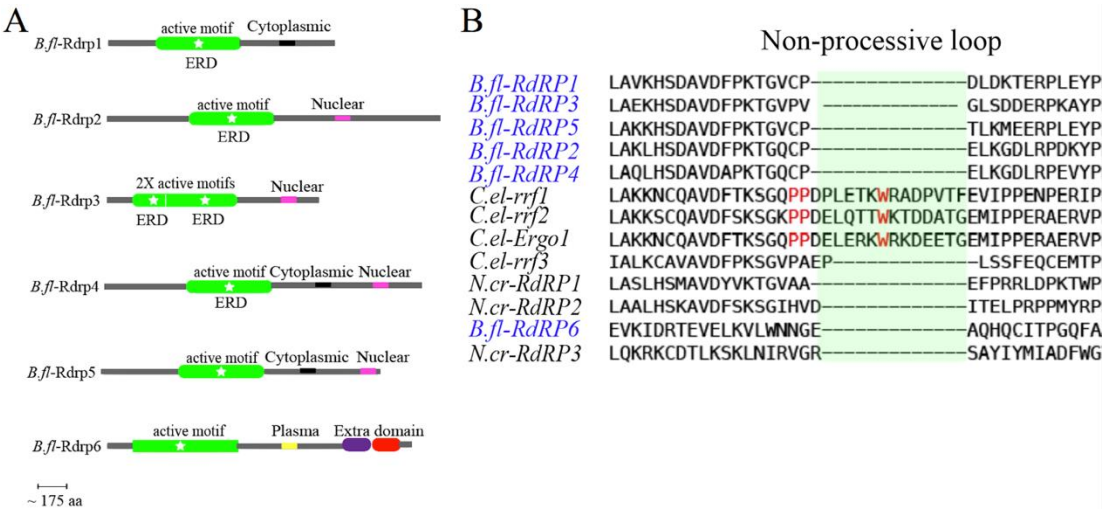


Figure 4.1 *Amphioxus RdRP domain structure and processivity.*

(A) Presence of the RdRP domain and active site motif in amphioxus RdRP. Green regions represent complete (smooth-end) or partial (flat-end) eukaryotic RdRP domain (ERD pfam ID = PF05183.12), white star indicates presence of the active site motif (DXDGD). RdRP6 possess two additional known domains, AAA_11 and AAA_12 respectively shown in purple and red. Gap in RdRP3 represents separation between duplicated ERD. (B) RdRP in amphioxus lack the C. elegans non-processive loop highlighted in green with red colored letters showing conserved proline/tryptophane residues. As a result, they are expected to behave like canonical eukaryotic RdRPs polymerizing long dsRNA Dcr substrate.

When we searched for the proline/tryptophane rich loop required for *de novo* synthesis in *C. elegans*, we found that *B. floridae* RdRPs alongside *rrf-3* and other eukaryotic RdRPs lack this loop (Fig 4.1B). Hence, unlike the specialized RdRPs in *C. elegans*, cephalochordates are expected to have processive RdRPs that polymerize long dsRNA that would be predicted to become Dcr substrates. Absence of *de novo* products

such as 5’ triphosphorylated reads was confirmed in the work of Pinzon et al. (Pinzón et al., 2019a), enabling us to observe only Dcr generated reads with 5’ monophosphates and 3’-hydroxyl termini characteristic of processive synthesis in this study. Further, by querying siRNAs for their ability to span exon-exon junction, the above referenced study did not detect any evidence to suggest that RdRP synthesize dsRNA that could be processed into siRNAs for RNAi purposes.

Next, we compared phylogenetic relationships between RdRPs from amphioxus species where full length RdRPs have been annotated to urochordate and cnidarian RdRPs. When the tree is rooted to a cnidarian RdRP, Amphioxus RdRPs clustered into five groups: group1&2 respectively consist of RdRP 1&3, group3&5 are mixed groups containing RdRPs 2,4, &5, while group4 contains a newly evolved group (RdRP6) seen clustering with tunicates and urochordates. Interestingly, homologs do not cluster clearly in the mixed groups, indicating that even in these sister species there is very recent gain and loss in sequences. This result suggests that while there is conservation of sequence between animal RdRPs, some RdRPs appear to be divergent even among sister species (Fig 4.2A).

4.5.2 Amphioxus RdRPs translocate between the cytoplasm and nucleus

To further investigate potential function of RdRPs, we focused on *B. floridae* homologs due to the more complete sequences offered by a superior genome assembly. First, trafficking motifs were investigated to understand localization and possible substrates where RdRPs are active. We queried localization signals of RdRPs, GAPDH and Histidine (H3) using WoLF PSORT tool and found that *B. floridae* RdRPs are predicted to traffic to several subcellular regions (Fig 4.2B). The majority appear to shuffle between the nucleus and cytoplasm. This appears true for all but RdRP6. RdRPs 1,3, &4

on the other hand tend to show bias either towards nuclear or cytoplasmic localization. Also, we observed that RdRPs 1 and 2 are preferentially localized to the cytoplasm and nucleus respectively. Surprisingly, we found that *B. floridae* RdRP6, unlike the other eukaryotic RdRPs show localization to the plasma membrane. This RdRP was observed to contain two extra domains (Fig 4.1A). We observed that elimination of these domains failed to impact its localization. Observation of *Actinia tenebrosa* (*A.te*) RdRP5 which is most like *B. floridae* RdRP6 did not show plasma localization. To further inquire about this behavior and why RdRP6 localizes differently from the other 5 RdRPs in *B. floridae*, we queried the amino acid sequences of the RdRPs using DEEPred (Y. Li et al., 2018).

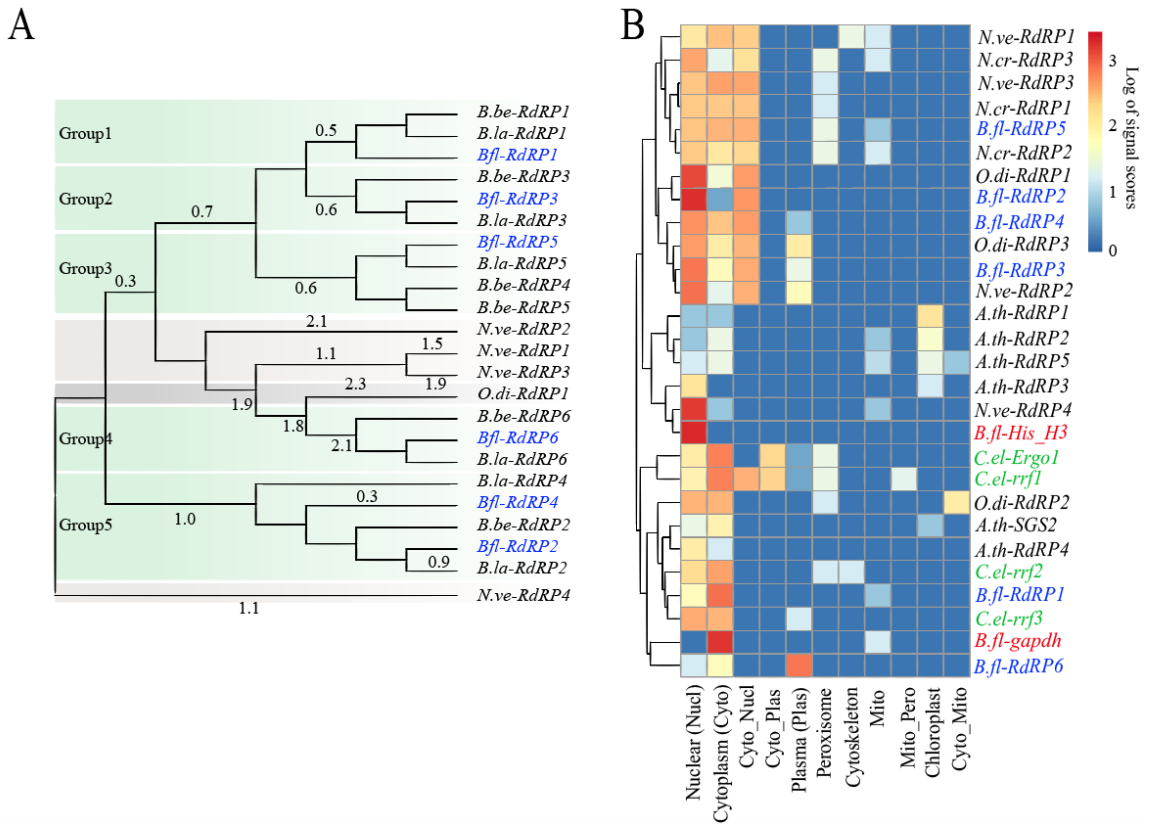


Figure 4.2 *Phylogeny and subcellular localization of some eukaryotic RdRPs.*

(A) Phylogenetic relationship of amphioxus RdRPs using tunicates as sister taxa and rooting the tree with N.ve-RdRP4 separates RdRPs in this clade into 5 groups. RdRP6 appear to be the most evolved. mRNAs sequences were aligned with MUSCLE algorithm and phylogeny estimated

Figure 4.2 (continued)

through Maximum likelihood. Only branch length values > 0.03 were shown. (B) Subcellular localization of RdRPs indicates eukaryotic RdRPs are preferentially localized to both cytoplasm and nucleus with B. floridae RdRP6 being an exception. (B. fl = B. floridae, O. di = O. dioica, C. el = C. elegans, A. th = A. thaliana, A. te = A. tenebrosa, N. cr = N. crassa, H. sa = H.sapiens, nucl = nuclear, cyto = cytoplasm, plas = plasma membrane, mito = mitochondria, pero = peroxisomes, cysk = cytoskeleton, extr = extracellular matrix, E.R = endoplasmic reticulum, Chlo = Chloroplast, Vac = Vacuole.

Based on the prediction, we found that most eukaryotic RdRPs belong to the enzyme class, transferases (EC# 2.7.7.48), carrying out nucleotidyltransferase function. However, the prediction showed that RdRP6 belong to a different enzyme class, hydrolases (EC# 3.6.4.13), which provide RNA helicases with ATP hydrolyzing activity and assist in unwinding RNAs using the energy derived from ATP breakdown, an activity demonstrated in some viral RdRPs (H. Li et al., 1999; Warrener et al., 1993). Further analysis of the amino acid sequences showed an AAA_11 (ATPase associated with various cellular activities) domain found in RdRP6 is solely responsible for this activity (Bar-Nun & Glickman, 2012; Beyer, 1997). Removal of the AAA_11 domain restores RdRP6 classification as a transferase. We tested RdRP5 in *A.te* to predict its enzymatic activity since it equally has two extra domains like *B. floridae* RdRP6. Unlike RdRP6, *A.te* RdRP5 lack the hydrolase activity. Further inspection of *A.te* RdRP5 revealed the presence of a partial AAA_11 domain. Taken together, we propose that Amphioxus RdRP6 in addition to having an RdRP activity possess an additional role in unwinding template RNA secondary structures downstream as polymerization proceeds, a role like replicative helicases and nsp13 in viruses (Chen et al., 2020; Kaplan & O'Donnell, 2002).

Further, amphioxus Dcr possess the relevant motifs and residues in its helicase and RNase III domains required for dsRNA cleavage and siRNA production as observed in *D. melanogaster* Dcr2, human Dcr, worm Dcr, *A. thaliana* Dcr2, Dcr3 and Dcr4, as well as

C. elegans Dcr1 (Hohn & Vazquez, 2011; Muhammad et al., 2019; Welker et al., 2010). We found that *B. floridae's* Dcr is likely to behave like *C. elegans* and human Dcr1 and possibly *D. melanogaster* Dcr2 as *D. melanogaster* Dcr1 appear to have undergone a single amino acid conservative substitution (Leucine to Isoleucine) that is an essential residue in the RNase IIIb domain (Fig 4.3A). All motifs necessary for ATP hydrolysis activity in the helicase domain were found to be like those in fruit fly, humans, and *C. elegans*. Although, human Dcr have been shown to not require ATP hydrolysis to be functional (H. Zhang et al., 2002). Also, we observed that lancelet Dcr possess the conserved Mg^{+2} coordinating residue in the RNase III domain (Kidwell et al., 2014). Moreover, overlapping function have been reported for the single Dcr encoded in both worms and humans. Therefore, a dual role for miRNA and siRNA maturation seems likely for lancelet Dcr (Qiu et al., 2017; Su et al., 2009b; H. Zhang et al., 2002; R. Zhang et al., 2018).

4.5.3 RdRPs in Amphioxus are transcriptionally active

Having seen the differences in RdRP localization, we inquired whether there is developmental or tissue specificity to RdRP expression. Next, we queried expression profile across developmental stages for all genes in the animal to observe how the RdRPs expression level ranked in comparison to other genes. We found that RdRP6 ranked higher than the median gene expression of 3.71 (log FPKM) whereas the other RdRPs ranked close to the median. RdRP expression here is therefore considered to be transcriptionally active which increases the possibility that they are translated into functional proteins (Fig 4.3B). Following this, we queried the transcription of the RdRPs across transcriptomic datasets of several *B. floridae* developmental stages as well as multiple *B. lanceolatum* tissues to understand where these polymerases could be acting. We found that RdRPs have

expression across development increasing significantly as the animal matures (Fig 4.3C). While RdRP6 was seen to be expressed across several stages, RdRPs 1, 3, and 4 expressions were detected in unfertilized egg suggesting the possibility of maternal deposition. Data from *B. lanceolatum* showed variable expression by tissue. All RdRPs are abundant in the gill with RdRP6 expressing strongly in the gut and neural tube as well, and others (RdRPs, 1, 3 and 5) expressed in the epidermis (Fig 4.3D). The abundant expression in gills was verified through RT-PCR (Fig 4.4). Taken together, our observation revealed that RdRPs are expressed in different tissues, and therefore have pervasive roles in gene regulation, and are not restricted to a phenomenon, like gametogenesis.

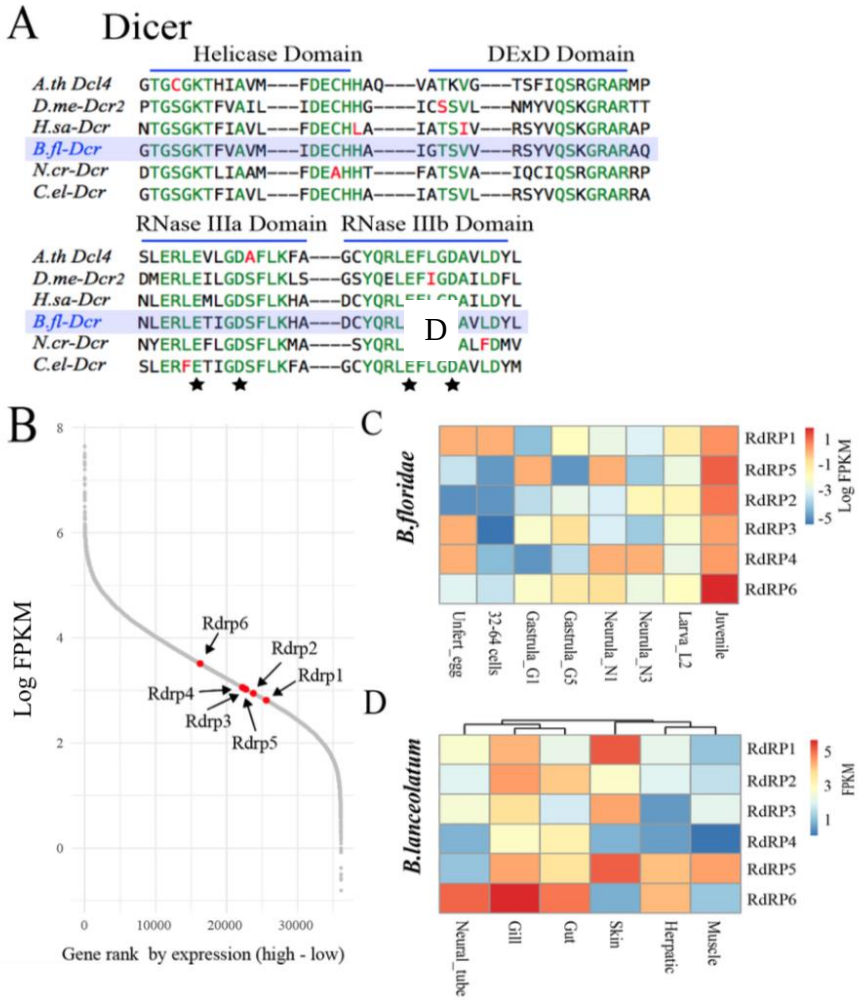


Figure 4.3 *RdRPs and Dcrs can potentiate siRNA pathway in amphioxus.*

B. floridae Dcr show similarity to D. melanogaster and human Dcrs respectively in its helicase and RNase IIIb domain suggesting a possible ATP-independent cleavage of dsRNA in this animal Residues colored in green are either conserved throughout the animals observed or in 5 of the 6 animals. Amphioxus Dcr is highlighted in purple and residues required for divalent metal ion (Mg+2) coordination are conserved in all 6 Dcrs observed (asterisk). B represents ranked expression profile of all annotated genes in B. floridae to show how B. floridae RdRPs rank among other genes. C and D are gene expression (log FPKM) profile of 6 B. floridae RdRPs respectively across developmental stages (B. floridae) and adult tissues (B. lanceolatum).

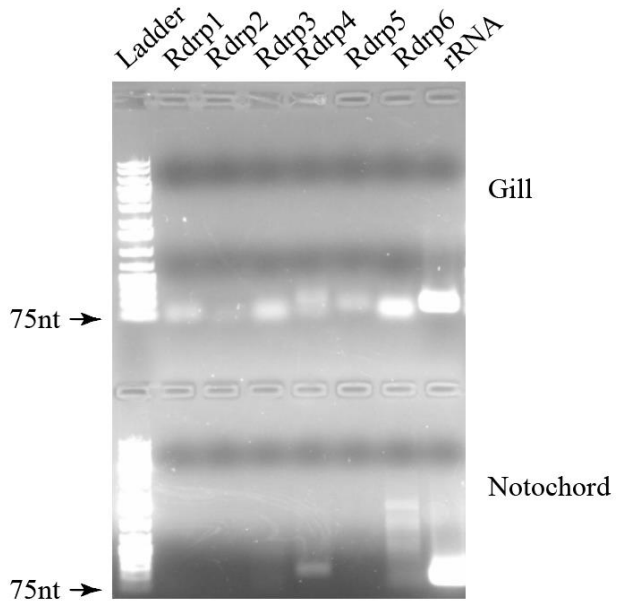


Figure 4.4 *RT-PCR agarose gel showing male amphioxus RdRPs in Gill (top) and Notochord (bottom)*

4.5.4 Globally, Dcr-like products are restricted to somatic tissues in amphioxus

Our observation of Amphioxus's RdRP expressions across different developmental stages and in several tissues and organs, led us to analyze the sRNAome in this animal to determine if and how the RNAi pathways are influenced by RdRP. Prior, we searched whether amphioxus Dcr possess the relevant motifs and residues in its helicase and RNase III domains required for dsRNA cleavage and siRNA production as observed in *D. melanogaster* Dcr2, human Dcr, worm Dcr, *A. thaliana* Dcr2, Dcr3 and Dcr4, as well as *C. elegans* Dcr1 (Hohn & Vazquez, 2011; Muhammad et al., 2019; Welker et al., 2010). We found that *B. floridae*'s Dcr is likely to behave like *C. elegans* and human Dcr1 and possibly *D. melanogaster* Dcr2 as the residue in *D. melanogaster* appear to have undergone a single amino acid conservative substitution (Leucine to Isoleucine) in a conserved residue of the RNase IIIb domain (Fig 4.3A) Moreover, overlapping function have been reported

for the single Dcr encoded in both worms and humans as mentioned previously. Therefore, a dual role for miRNA and siRNA maturation seems likely for *B. floridae* Dcr (Su et al., 2009b; R. Zhang et al., 2018).

Briefly, we sequenced sRNAs from male and female adult *B. floridae* and obtained publicly available sRNA sequencing libraries from adult female *B. belcheri* for; gill, intestine, epidermis (skin), muscle, nerve-cord, notochord, and ovary. We then mapped and excluded all reported miRNA (miRbase.org) from the sRNA libraries (two mismatches allowed). Non-miRNA matching reads where then mapped to the respective genomes (allowing no mismatches). Our analysis revealed two distinct sRNA peaks corresponding to siRNA (21-23nt) and piRNA (28-31nt) lengths in *B. floridae* adult whole-body libraries and *B. belcheri* tissues (Fig 4.5A and 4.5B).

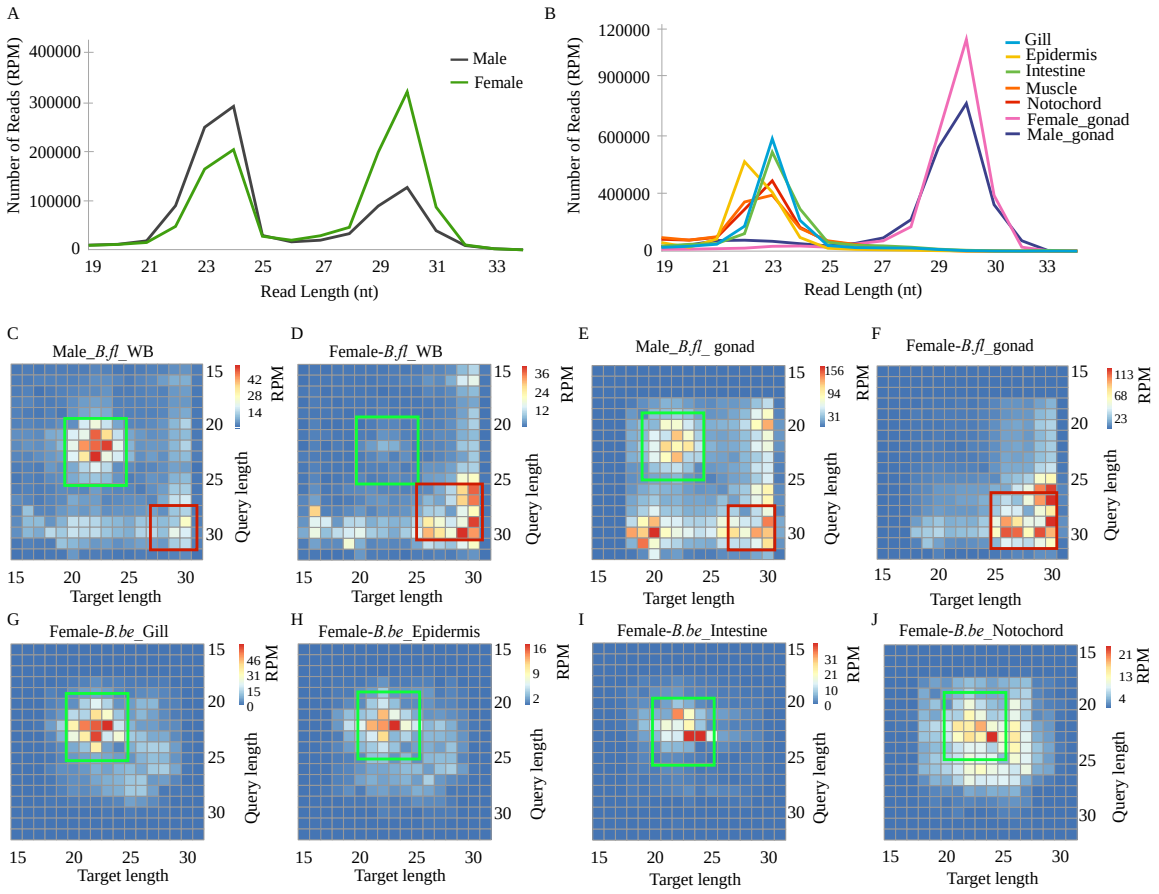


Figure 4.5 *Global read size distribution and Dcr signature in amphioxus sRNA libraries.*

(A) miRNA depleted adult male and female WB libraries (B. floridae). Peaks correspond to 21-24 nt (siRNA-like) and 28-31 nt (piRNA-like) sized reads are highlighted in green and red squares. Similarly, the recovered peaks were also found in the tissues (B). A strong bias towards 21-24nt sRNAs was observed in somatic tissues while majority of sRNAs in the ovary are in the piRNA-like peak. Overall indication of read pairs possessing 2nt 3' overhangs characteristic of Dcr cleavage are observed in male and female WB, as well as in male gonad (green square) (C-E). Also, ping-pong signatures are observed in male and female Wb, as well as in male and female gonad libraries (red squares). Female somatic tissue (B. belcheri) libraries show only Dcr signatures (G-J). Expression values are in reads per million (RPM).

Further, we found abundant piRNA expression in the female library compared to the male. This is possibly due to the presence of a piRNA cluster resembling the *flamenco*-like (*flam*) locus that we recovered from the female *B. belcheri* library (Fig 4.6). The reverse of this is seen in the siRNA peaks between the male and female as siRNA-sized reads were more prevalent in the male whole-body (WB) library (Fig 4.5A). While there

appear to be variation in siRNA-like population in all tissue libraries, gill, skin, and intestine libraries appear to have a higher siRNA-sized reads compared to the rest of the tissues observed (Fig 4.5B). Likewise, increased production of siRNA-like reads in the gill could be consistent with the abundance of RdRP expression that we observed in this tissue.

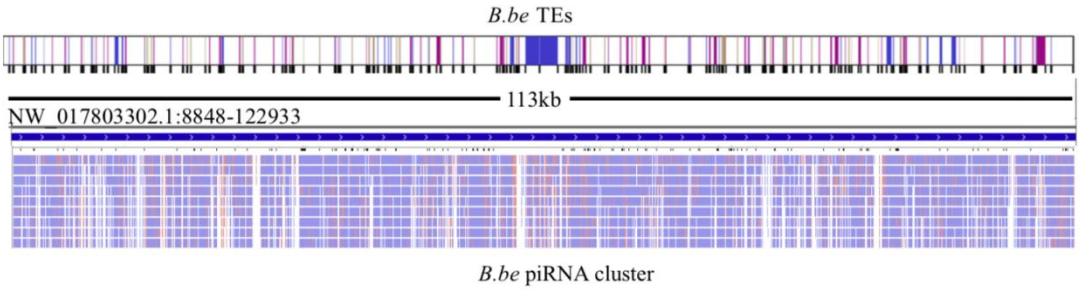


Figure 4.6 *Amphioxus giant piRNA cluster dominates whole body libraries and ovaries.*

To further investigate whether the reads from the sRNA libraries (20-23nt) are generated by Dcr, we queried each library for presence of overlapping reads with 3' 2nt overhangs characteristic of Dcr cleavage. In the WB libraries, we found that dual strand Dcr signatures are more abundant in the male library relative to the female library (centers of Fig 4.5C-D respectively) consistent with the read size distribution in Fig 4.5A. Accumulation of both somatic and germline siRNA reads likely account for the high siRNA peak observed in the male WB library. It is likely that siRNAs of RdRP origin also have a role in protecting the male germline genome alongside piRNAs. Immediately, one would think that siRNAs associated with RdRPs if present in this animal would have a bias towards the male species. This is however not the case because while apparent dual strand reads appear to be poorly represented in female WB library, somatic read pair overlaps of Dcr origin are enriched in the female gill, skin, intestine, and notochord libraries (Fig 4.5G-J) but not ovary (Fig 4.5F). Therefore, the poorly detected siRNAs in the female WB library (Fig 4.5D) are possibly due to overrepresentation of piRNAs in the library which masked

sequencing of the sRNA populations creating the bias observed in this library. This prediction is supported by the overwhelming abundance of somatic siRNAs when individual tissues are considered for sRNA library creation. Likewise, the piRNA peak observed in the WB female library potentially represents germline piRNAs, consistent with the piRNA accumulation we observed in the female gonad (Fig 4.5F).

4.5.5 Characterization of sRNA producing loci in amphioxus

Having seen that Dcr like reads are preferentially produced in the somatic tissues, we sort to identify and characterize differences is individual Dcr associated hot spots. Briefly, mapped reads from tissues libraries were filtered to identify regions where dual strand (dsRNAs) are abundant (hot regions) using a criterion that take locus length and read coverage into account. We obtained the following number of loci generating dual strand reads in the different tissues: gill (120), skin (109), intestine (99), muscle (162), nerve cord (94), notochord (259), ovary (507). Gill, skin, and notochord sRNA distribution by loci is shown in Fig 4.7A-C respectively. To observe loci with the most abundant siRNAs, we compared the three libraries and found that 14 of the hot regions are common to all three tissues (Fig 4.7D). Further, we observed that while these regions are common to these tissues, reads (RPM) originating from the 14 loci appear to be more abundant in the notochord (Fig 4.7D).

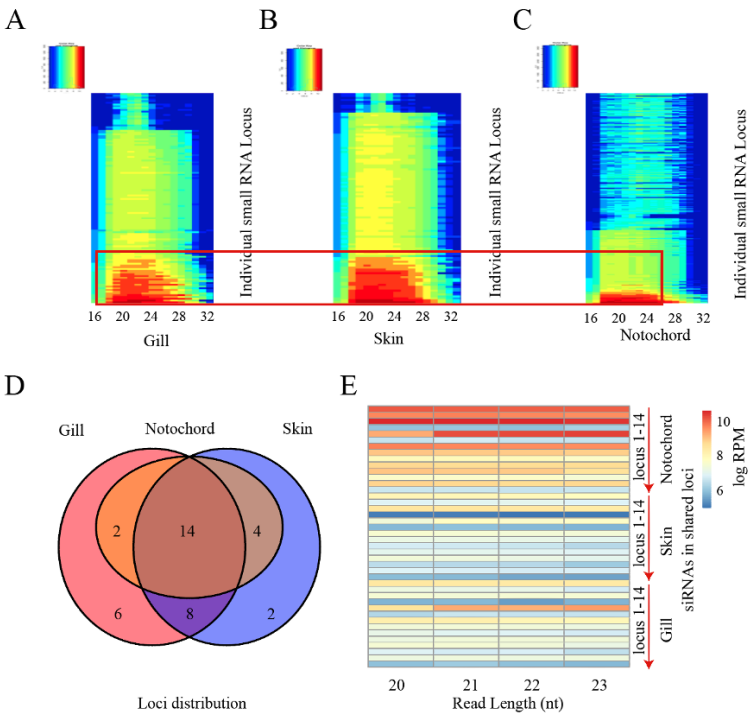


Figure 4.7 *Somatic siRNA producing loci in amphioxus.*

Heatmaps show normalized read count from region of high Dcr activity in gill (A), skin (B), and notochord (C). The region highlighted in red was compared between the three tissues (D). E shows expression of sRNAs in the intersecting region compared in D. Red arrows show direction of loci count (1-14).

4.5.6 RdRP synthesizes long dsRNA from Pol II generated transcripts

To identify potential RdRP-associated sRNAs, males amphioxi were targeted. Using the approach and criteria described above for recovering loci with dual strand reads possessing Dcr signature, we identified 338 loci comprising of 64 high confidence and 274 low confidence calls. Confidence here was determined using size, overlap reads, 2-nt 3' overhang Dcr signature, and read abundance features of each locus. We hypothesized that some of these loci could potential implicate RdRP in dsRNA synthesis in this animal.

To further investigate if any of the 64 loci identified have RdRP involvement in siRNA production, we sort to find evidence of nascent RdRP activity through inhibition of

RNA Pol II using α-amanitin (Aman) and subsequent addition of 4sU to label nascent transcripts. Following a brief inhibition of Pol II and addition of 4sU, sRNA sequencing libraries were generated (Fig 4.8A). We then inspected and compared the 64 loci recovered from the male WB library in Pol II inhibited and Pol II active libraries (control). We detected nascent RNA production in some of the loci. 21 of the 64 genomic loci in the libraries showed adequate nascent transcript formation. α-amanitin inhibition of Pol II was confirmed in one of the 21 loci (NC_049983.1:16490001-16490612) evident by absence of nascent transcripts in Pol II inhibited library (Fig 4.8B and Fig 4.8D).

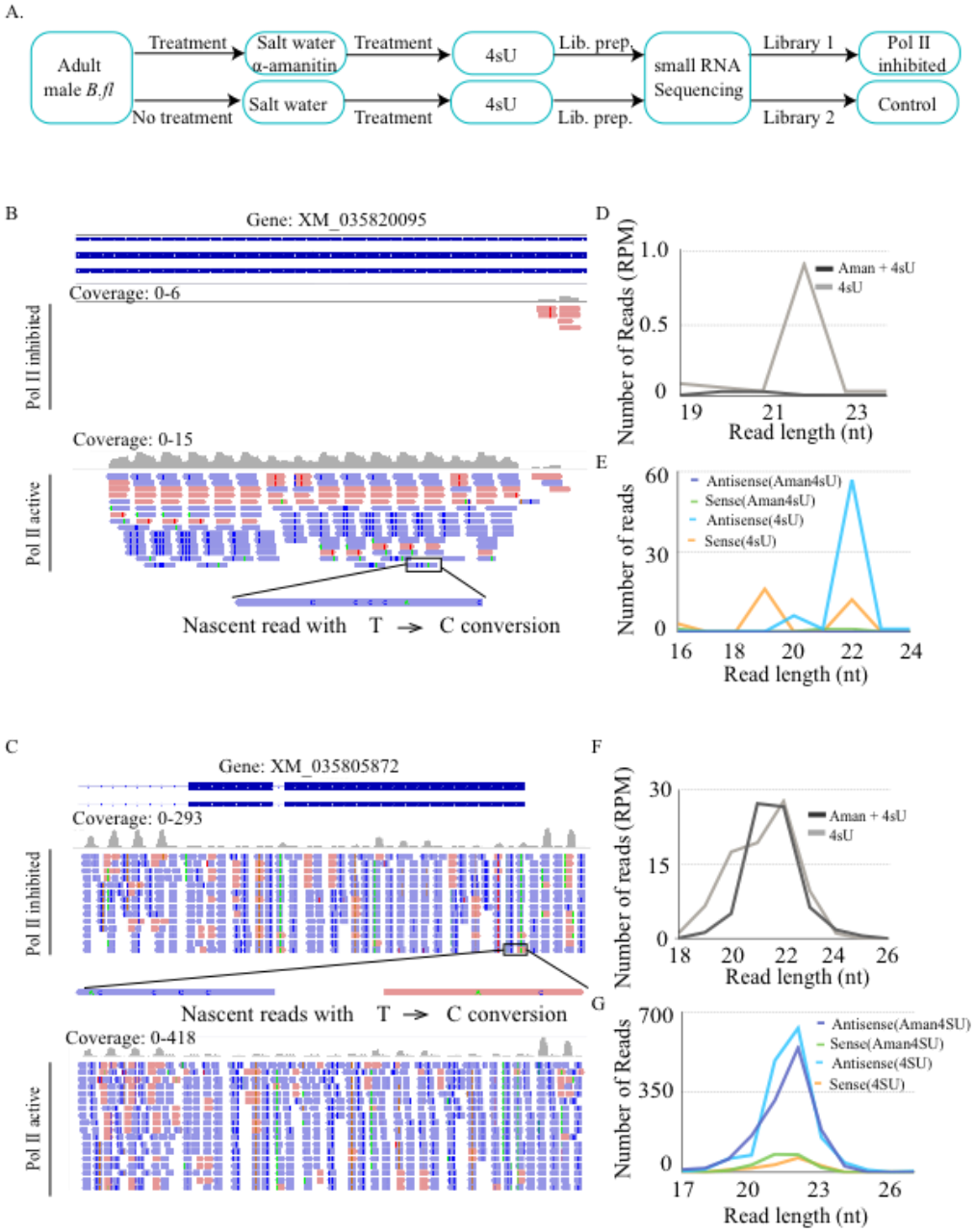


Figure 4.8 *Nascent transcripts potentially synthesized by amphioxus RdRP were detected in the absence of RNA Pol II.*

(A) Summary of experimental workflow for nascent transcript labeling. (B) An example locus showing successful inhibition Pol II. This locus is a non-RdRP target locus only showing transcripts in Pol II active library. C is an example RdRP target locus showing nascent transcripts despite Pol II inhibition. Blue and red arrows represent antisense, and sense reads respectively with magnified region showing Thiamin to Cytosine conversion in nascent transcripts. Respective size distribution as well as antisense/sense read distribution in the non-RdRP, and potential RdRP target locus are shown in (D-E & FG).

In addition, the reads from this Pol II target locus show a narrow range of read size accumulation (21-23nt) with many of the reads mapping uniquely to the transcript of origin (Fig 4.9B-D, 4.9A-B). RNA-fold analysis revealed that the Pol II transcript folded back onto itself to generate a dsRNA Dcr substrate (Fig 4.9C-D).

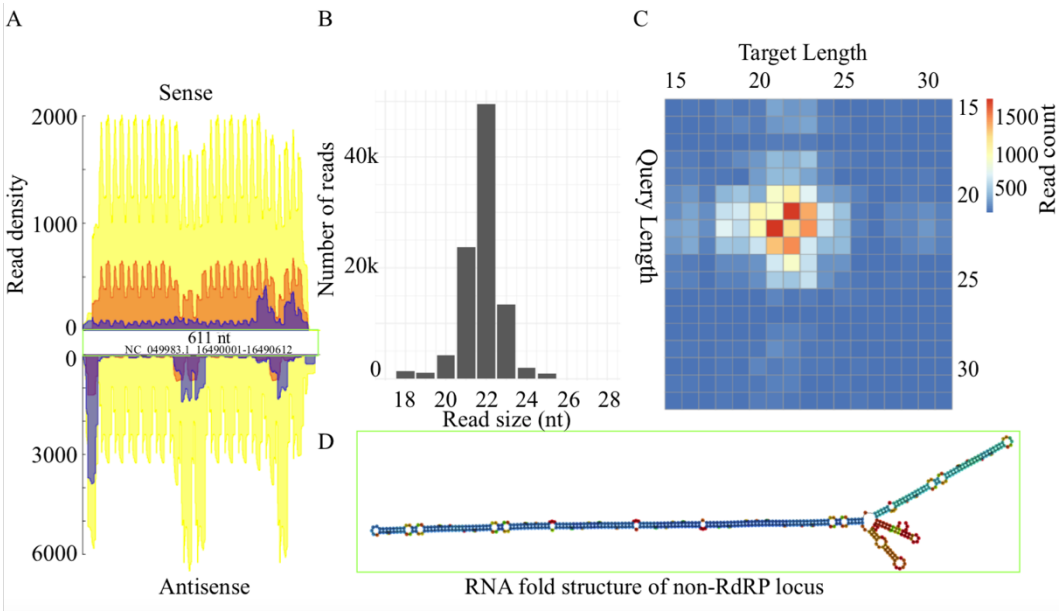


Figure 4.9 *Example non RdRP target locus observed after Pol II inhibition.*

(A) Density plot shows reads from example non RdRP targeted locus in untreated adult male WB library devoid of miRNA. (B) Distribution of reads in the proposed locus show a narrow size range (21-23). (C) Dual strand read accumulation of Dcr cleaved fragments characterized by 2-nt 3' overhangs especially in reads between 21-23 nt (center) are seen to be produced from the template (>600 bp) in this locus. (D) RNAfold structure of the template region highlighted in green in 4.9A.

Also, we show here another locus (NC_049996.1:5328725-5330257) with nascent sRNAs believed to be derived from RdRP polymerization in both Pol II inhibited and Pol II active libraries (Fig 4.8C). The originating transcript when folded shows extremely poor complementarity in the sequence (Fig 4.10E). As a result, we conclude that the second strand was generated from a Pol II transcript. Further, we observed an abundant as well as broader distribution of read sizes (20 – 24 nt) generated by Dcr suggesting an inconsistent

sRNA cleavage pattern in reads from the proposed RdRP target (Fig 4.8F). In addition, we found that majority of these reads are antisense to the template strand and originated from an un-spliced genic transcript (Fig 4.8G). The mechanism resulting in excessive antisense reads here is unclear. Observation of the proposed RdRP target locus in the male WB library confirms that RdRPs in amphioxus are indeed capable of generating dsRNA from a longer template (1532 nt), consistent with our earlier suggestion that amphioxus RdRPs have a processive nature (Fig 4.10A-C).

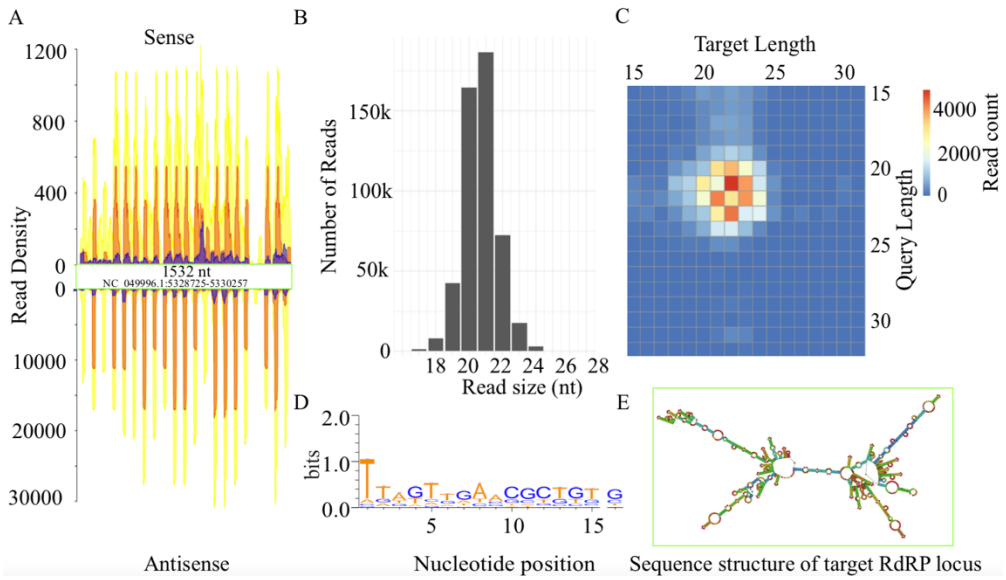


Figure 4.10 *RdRP generated dsRNA acts as substrates for Dcr.*

(A) Density plot shows reads from the proposed RdRP targeted locus in untreated adult male WB library devoid of miRNA. (B) Distribution of reads in the proposed locus show a wide size range (19-23). (C) Dual strand read accumulation of Dcr cleaved fragments characterized by 2-nt 3' overhangs especially in reads between 21-23 nt (center) are seen to be produced from the long template (>1500 bp) in the RdRP targeted locus. (D) Through the Weblogo algorithm, a strong bias for 5' U was observed in reads from this locus confirming that indeed, Dcr did cleave the long dsRNA generated by RdRP into siRNAs. Also, absence of A bias at the 10nt position confirms that

Figure 4.10 (continued)

these read pairs are not piRNAs. In A, purple regions = unique mapping reads, yellow regions = reads of all sizes, and orange regions = multi-mapping reads. (E). RNA fold structure of the template region highlighted in green in 4.10A.

Table 4.4 *List of 21 loci generating some nascent transcript in both libraries*

Loci #	Contig	Start	End
1	NC_049979.1	15272270	15273335
2	NC_049979.1	25226208	25226309
3	NC_049979.1	25264169	25266811
4	NC_049979.1	29807522	29813592
5	NC_049979.1	33605840	33609421
6	NC_049980.1	2894323	2896181
7	NC_049980.1	29923719	29926380
8	NC_049981.1	23868226	23870278
9	NC_049981.1	28427022	28428939
10	NC_049981.1	30449173	30450738
11	NC_049981.1	7225788	7227336
12	NC_049982.1	24640481	24642668
13	NC_049983.1	21625903	21626943
14	NC_049983.1	21893042	21898991
15	NC_049990.1	22695163	22696221
16	NC_049993.1	9914923	9916001
17	NC_049994.1	15744501	15745663
18	NC_049994.1	16004550	16005624
19	NC_049995.1	5814866	5815938
20	NC_049996.1	5328701	5330325
21	NC_049996.1	4431671	4432712

4.5.7 RdRP synthesized dsRNA acts as substrate for siRNA production

To further explore the properties of siRNAs suggested here to be of RdRP origin from Pol II inhibited library, we analyzed reads from the proposed RdRP target locus in the untreated *B. floridae* male WB library. Read representation in this library reflect what was observed in the treated library in addition to more read abundance resulting from a greater sequencing depth (Fig 4.10A). In addition, an overlapping signature suggests that Dcr acts on dsRNA generated by RdRPs from this locus (Fig 4.10B-C). This activity is evident in the continuity of sRNAs produced from the proposed locus. Sequence logo algorithm further confirms that the sRNAs produced here were generated by Dcr as evident by the presence of a 5' Uridine bias (Fig 4.10D). The absence of a 5' Adenine (A) or Guanine (G) nucleotide bias from the siRNAs in the proposed RdRP target locus commonly associated with secondary and *De novo* RdRP synthesis re-emphasizes the processive nature of Amphioxus RdRPs.

4.5.8 Mobile elements constitute targets for RdRP generated siRNAs in Amphioxus

To probe the function of antisense siRNAs made from the 21 proposed loci where nascent transcripts are generated from, we searched for genes, TEs, and viral genomes that are trans-regulated by these siRNAs. We found that when antisense siRNAs are mapped with two mismatches allowed for amphioxus TEs and two/three mismatches for the three viruses, over 26% (1032 reads) of the reads mapped to TEs with majority matching DNA repeats while 2.21% (87 reads) mapped to the viral genomes used here (Fig 4.11A and 4.11B respectively).

We observed that majority of the virus matching antisense reads target Abalone Herpesvirus (54%) when 3 mismatches are permitted. This is possibly due to the genome

size of the virus as the smaller viral genome (lancelet Adintovirus) only showed 6% matching reads. However, we found that the virus mapping rate decreases when the number of mismatches is reduced. Further, we obtained gene annotation targets by mapping antisense sRNA to amphioxus genome and intersecting mapped reads with a *B. floridae* annotation file. The intersection generated 390 hits for all siRNAs from the 21 regions predicted. Filtering out genes that are targeted by 1-19 reads resulted in a total of 28 genes (Table 4.5), of which over 95% are trans-targeted (Fig 4.11C). These results suggest that the predicted RdRP associated siRNAs potentially have a role in controlling TEs, and genic transcripts.

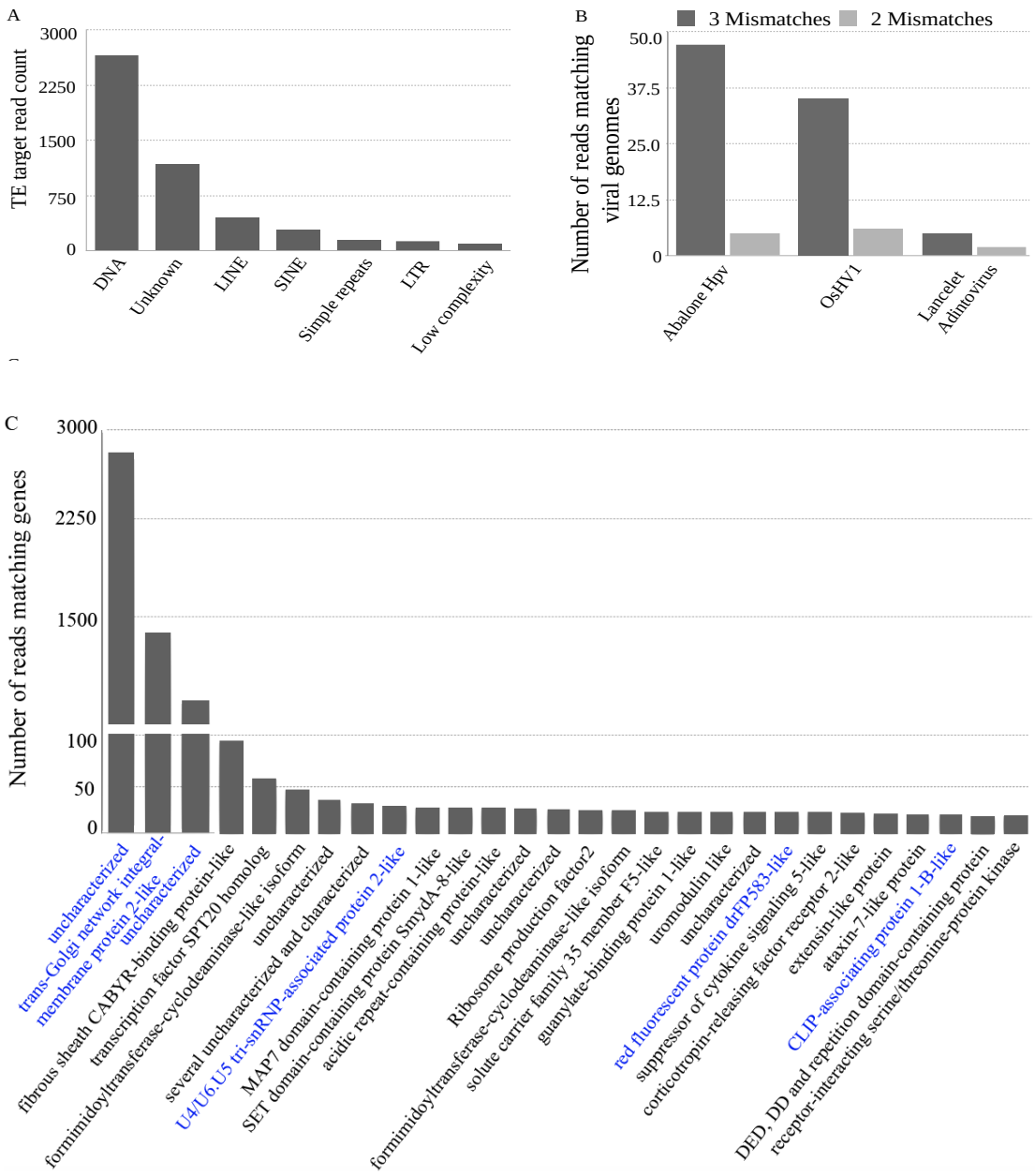


Figure 4.11 *Predicted RdRP generated siRNAs target Amphioxus TE's, and genes.*

(A) Various mobile genetic elements are targeted by siRNAs from the proposed RdRP target locus with DNA elements being the most regulated. (B) Ability of the proposed siRNAs to match genomes of 3 available genomes associated with Amphioxus were tested. While reads poorly aligned to viral genomes, some matching was observed when the number of mismatches was increased to 3. Targets where number of reads matching genes exceeds a threshold (19 reads) are shown in (C). Black colored labels represent genes that are trans-regulated while blue colors include some of the genes where proposed RdRP generated siRNAs originated.

Table 4.5 *List of genes targeted by reads predicted to be RdRP-associated siRNAs*

Contig	Start	End	gene
NC_049980.1	29866841	29950601	unconventional myosin-IXAa-like
NC_049983.1	18261724	18328450	synapsin-3-like
NC_049979.1	10545238	10590718	F-box/WD repeat-containing protein 7-like
NC_049981.1	30448843	30489115	U4/U6.U5 tri-snRNP-associated protein 2-like
NC_049986.1	4959598	4998894	vam6/Vps39-like protein
NC_049992.1	18306060	18344346	agrin-like
NC_049991.1	20259241	20295336	TATA-binding protein-associated factor 172-like
NC_049985.1	25855143	25888561	nuclear mitotic apparatus protein 1-like
NC_049984.1	22711632	22742645	hydroxyproline dehydrogenase-like
NC_049994.1	17587262	17613438	KAT8 regulatory NSL complex subunit 1-like
NC_049984.1	2372801	2398730	mRNA (2'-O-methyladenosine-N(6)-)-methyltransferase-like
NC_049979.1	10067780	10092947	putative N-acetylated-alpha-linked acidic dipeptidase
NC_049982.1	3897803	3921052	uncharacterized
NC_049979.1	2477430	2500476	syntaxin-5-like
NC_049983.1	5035822	5058803	centromere protein P-like
NC_049996.1	5309421	5331826	uncharacterized
NC_049985.1	21177075	21199291	nucleoporin p54-like
NC_049984.1	11007776	11027719	DED, DD and repetition domain-containing
NC_049996.1	5890711	5910384	solute carrier family 35 member F5-like
NC_049988.1	16509033	16528341	corticotropin-releasing factor receptor 2-like
NC_049984.1	23101811	23120676	uncharacterized
NC_049983.1	18293324	18310376	metalloproteinase inhibitor 3-like
NC_049982.1	24632273	24647502	red fluorescent protein drFP583-like
NC_049983.1	12405773	12420907	methionine-R-sulfoxide reductase B3-like
NC_049994.1	4654811	4668930	ataxin-7-like protein
NC_049983.1	16823001	16836281	SET domain-containing protein SmydA-8-like
NC_049979.1	29803644	29816381	uncharacterized
NC_049979.1	21076581	21088592	uncharacterized
NC_049982.1	29646482	29658342	tolloid-like protein 2
NC_049984.1	24028514	24040295	guanylate-binding protein 1-like
NW_02336574	32527	44271	uncharacterized
NC_049979.1	3779751	3791118	suppressor of cytokine signaling 5-like
NC_049988.1	16775378	16784238	solute carrier family 40 member 1-like
NC_049980.1	19381673	19389356	5-formyltetrahydrofolate cyclo-ligase-like
NC_049987.1	12158073	12164671	Ribosome production factor2
NC_049995.1	18648682	18654051	receptor-interacting serine/threonine-protein kinase 2-like

Table 4.5 (continued)

NC_049979.1	29818695	29823788	extensin-like protein
NC_049979.1	10582158	10587045	pneumococcal serine-rich repeat protein-like
NC_049983.1	16487601	16492163	acidic repeat-containing protein-like
NC_049985.1	6007143	6011464	FTCD-like isoform
NC_049979.1	29825965	29829697	fibrous sheath CABYR-binding protein-like
NC_049983.1	16825705	16828031	MAP7 domain-containing protein 1-like
NC_049996.1	5328274	5330072	trans-Golgi network integral membrane protein like
NC_049985.1	6007143	6007329	FTCD-like
NC_049985.1	6007431	6007550	FTCD-like isoform

4.6 Discussion

To date, animal RdRP's only known function is involvement in RNAi. A previous study investigated RdRPs in amphioxus, where no evidence was recovered to suggest its involvement in RNAi. Through curation of sRNAs in developmental stages and tissues, based on terminal sRNA modifications, nucleotide bias and siRNA size distribution, characteristic *de novo* or secondary RdRP products were not recovered in amphioxus. As a result, the authors suggested that RdRPs in amphioxus are not involved in RNAi despite encoding six homologs of the protein in their genome, the most for any animal. While the approach undertaken in the above study seeks to identify siRNAs of RdRP origin across entire sRNA through a holistic inspection of the amphioxus sRNAome, it is easy to miss subtle RdRP activity especially bearing in mind that many animals elicit RNAi without the need for the protein. Their activity may therefore be small enough that the bulk of sRNAs in the animal mask their recovery.

Eukaryotic RdRPs can carry out second strand RNA synthesis either through a processive or non-processive means. Understanding the nature of processivity in Amphioxus RdRPs was crucial to the sequencing approach to undertake for our study. In addition to the absence of *de novo* or secondary siRNAs reported in the work of Pinzon *et*

al. (Pinzón et al., 2019a), our investigation confirms that RdRPs in this animal lack the putative loop required for *de novo* or non-processive synthesis. As a result, we applied a sequencing approach that recovers 5' monophosphorylated reads characteristic of Dcr cleavage as we have shown that amphioxus Dcr possess the necessary motifs in its helicase domain for siRNA synthesis.

In contrast to the holistic approach used in the aforementioned study, the present study interrogates individual genomic loci in search of vestigial RdRP activity. In addition to possessing relevant domain (ERD) and active site motif (DxDGD) required for processivity, Amphioxus RdRPs are expressed under normal conditions as evident in transcriptomic datasets, and RT-PCR. As a result, it is more likely than not that we uncover some activity or function with the most likely being involvement in RNAi. Curation of individual genomic loci where Dcr processing are elevated revealed that any 1 of 64 genomic loci in Amphioxus has the potential to serve as RdRP target for dsRNA synthesis. Inhibition of RNA Pol II followed by metabolic labelling of nascent transcripts with 4sU in Amphioxus enabled us to observe nascent transcripts in these loci. We uncovered 21 regions where nascent RNAs are generated in the absence of Pol II. The sRNAs from these genomic loci where then used for downstream target identification.

Eukaryotic genomes are known to be rich in mobile genetic elements like TEs. These elements have been a major driving force, contributing to genetic diversity and evolution. Through RNAi, sRNAs including those of RdRP origin have been implicated in maintenance of genome integrity by regulation of TE activities (Gent et al., 2009). Our search for TE matching reads by the proposed RdRP generated siRNAs showed that TEs

are indeed target of control. We believe that RdRP generated siRNAs offer protection and maintenance of somatic genome integrity through control of TEs.

Furthermore, subcellular localization of RdRPs in Amphioxus showed that as opposed to worm non-processive RdRPs (*rrf1, rrf2,* and Ego1), most Amphioxus RdRPs have localization patterns like worms *rrf-3*. This enabled us to question whether conservation of sequences could potentially influence function in this animal. For eukaryotic and viral RdRPs, the nucleus and cytoplasm represent the most frequent subcellular locations. Also, unlike many DNA viruses, replication cycles of most RNA viruses involve an RNA intermediate, and in most cases, the replication takes place in the host cell cytoplasm. For example, to achieve translation in negative strand RNA viruses and dsDNA viruses, the genes encoded must be transcribed respectively into a positive strand RNA and an mRNA generating substrates for RdRPs. Also, most viruses that replicate in the cytoplasm tend to compartmentalize replication of their genome making use of subcellular compartments like the mitochondria, Golgi apparatus, plasma membrane and others. This act enhances host immune evasion by the viruses (Melanie et al., 2014). Surprisingly, eukaryotic RdRP proteins that we analyzed here appear to be found in several of these compartments. *B. floridae* RdRPs appears to be localized more in the cytoplasm, nucleus, mitochondria, and plasma membrane. This signals invariably suggests a role for eukaryotic RdRP in the life cycle of viruses. Despite this pattern, we failed to recover significant viral genome matching reads like we see with TE and genic transcript matching. Although using more viral genomes especially RNA viruses in the analysis could improve the number of reads matching them. However, it was reported that *de novo* assembly of non-genome mapping reads identified several non-viral genomes. Further, the few reads

mapping to viruses may just be enough to interfere with the life cycle of the viruses used here. This sporadic activity therefore suggests an association between RdRP localization and their potential involvement in viral control. More studies will be necessary to establish why eukaryotic RdRPs are preferentially localized in similar compartments as viruses.

4.7 Conclusion

Taken together, our findings in this work suggest that a remnant function seem likely for RdRPs in Amphioxus despite their gradual disappearance from animal genomes. As the protein continues to disappear from eukaryotic systems, it is reasonable to assume that their function will likewise gradually fade until it is completely replaced by either a novel or an advanced mechanism like we have witnessed in numerous lineages where RNAi is elicited in the completely absence of the protein. Based on our findings, we assume here that Amphioxus RdRPs have a vestigial role in RNAi associated with somatic genome protection against mobile genetic elements.

CHAPTER V – POTENTIAL RdRP MEDIATED VIRAL CONTROL IN OYSTERS AT ELEVATED TEMPERATURES

5.1 Relevance

Genetic tools and RNAi technology possess enormous potential for creating new traits or upgrading already existing ones. These genetic technologies are however not sufficiently available for many non-model animals like oysters, warranting the need for basic research into RNA biology of this animals.

Oyster reefs represent one of few saltwater ecosystems that can provide physical and direct estimate for determining marine ecosystems condition due to the presence of only few oyster species making up its structure. Throughout the world's estuaries, the last two centuries have seen well over 85% loss in oyster reefs and the ecosystem related values such as water filtration, food for other marine animals, stabilization of the shoreline as well as coastal defense and other services they provide (Beck et al., 2011; Board OS, 2010; Newell, 2004). This decline is particularly striking as the remaining native oysters in the wild can be traced to only five ecoregions including the Southern Gulf of Mexico. Oyster fishery in this region accounts for the second largest annual catch between 1995 – 2004 resulting in slow but consistent decline in reef ecosystems (Beck et al., 2011). This decline has sparked off efforts for shellfish reef ecosystem restorations by the global community. Despite restoration efforts such as conservation, fishery management and introduction of non-native species, excessive fishery and diseases remain prevalent, slowing down current method of reef restoration.

5.2 Introduction

RNAi in bilaterian animals have been actively studied. However, no comprehensive study has been conducted to examine sRNA biogenesis in lophotrochozoans, making them an exception. Lophotrochozoan genomes appear to encode an interesting RNAi pathway configurations and factors. Across the subphyla, members appear to have miRNAs, piRNAs, a single Dcr enzymes, and in bivalves, four RdRP homologs. A similar configuration and RNAi pathway components have been reported previously from our lab in mite species, which have tremendous significance in development of RNAi-based technology, genome defense, and genome evolution. Of relevance to this chapter are the RdRP proteins encoded in the oyster's genome, specifically in the *Crassostrea* genus. To this end, this chapter aims to explore whether RdRP has a role in the endogenous RNAi pathway against viruses or genome defense in this animal. Unraveling RdRP as a functional component of the animal's RNAi pathway could have potential in understanding sRNA pathways and their eventual exploitation for designing RNAi technologies in marine bivalves.

Further, marine-related industries are a major aspect of the Southern Mississippi economy as the eastern oyster fishery market are estimated to have an estimate of about $80 M annually. Despite this and the fact that reef ecosystem faces risk of endangerment in the future, the existence of genetic technologies for manipulating or developing potential RNAi technologies for use to derive new traits in these organisms are lacking. Decline in oyster reefs due to abiotic stresses and pathogens, could be ameliorated with gene manipulation approaches.

5.3 Materials and Methods

5.3.1 Oyster primary cell culture establishment for in-vivo biochemical studies

To assess cells for biochemical tests during RdRP study in this chapter, we established a protocol for primary cell culture. To begin, a custom aquarium with a cooling device, a filter device, and an aeration system built into it was constructed. The aquarium system was regularly cleaned and observed for changes in nitrate, ammonia, pH, and salinity changes as well as removal of dead oysters. Following this, the eastern oysters (*C. virginica*) were obtained from gulf coast and transported to the lab on ice. Oysters were kept in the aquarium until needed for cell culture.

To establish primary cells, oysters were obtained from the in-house aquarium and carefully shucked. Following this, hemolymph was carefully extracted, filtered, and saved in a 15mL tube for later use. Tissues of interest, namely, gill, mantle, palp, and gonad were cut and sired twice for 20-30 minutes each time in a decontamination buffer comprising of 1X antibiotic-antimycotic (10,000 units/mL of penicillin, 10,000 µg/mL of streptomycin, and 25 µg/mL of Gibco Amphotericin B, with Gibco ID 15240-062, Thermo Fisher Scientific) in filtered salt water (FSW). Following this, 5-8 explants were excised aseptically from each tissue and cultured in 2-3 mL of culture media in a 36 mm culture dish. The culture media comprised of 50% FSW and 50% Leibovitz L-15 media (Gibco 11415064, Thermo Fisher Scientific) supplemented with 1-2% filtered hemolymph saved from earlier step. Antibiotic was then added to a final concentration of 1X. Tissues were cultured at room temperature overnight to two days or until cells reach 45-60% confluency.

To observed cell morphology, we fixed ovarian cells on a slide and stained the nuclei and cytoskeleton respectively with DAPI and phalloidin 568 (Thermo Fisher

Scientific catalog numbers D1306 and A12380). Stained slides where then blotted and viewed with a confocal microscopy.

5.3.2 Recovery of RNAi proteins and analysis

To identify proteins relevant for RNAi, we performed both local and NCBI blast searches against genomes of *C. gigas* as well as *C.vi* using proteins from other animals. For RdRP sequence recovery, we used *rrf*-3,2, and 1. For AGO, and PIWI protein sequence identification, we used *D. melanogaster* AGO1 and AGO2, PIWI1, AGO3 and Aub amino acid sequences.

Following this, AGO, PIWI and RdRPs were compared through phylogenetic analysis with homologs from six sub-phyla across the lophotrochozoan phylum including Gastropoda, Cephalopoda, Brachiopoda, Polychaeta, and Clitellata (leeches) using MUSCLE for amino acid sequence alignment and phylogeny estimation with the maximum likelihood method in the MEGA7 software (Kumar et al., 2016).

5.3.3 Estimation of oyster RdRP abundance in transcriptomic datasets

After identifying RdRPs in the genome, we estimated transcription levels of the proteins alongside other RNAi factors using transcriptomic data deposited in NCBI (Xu et al., 2014). Briefly, authors from the referenced study either dissected 2-3-year-old adult oysters (*C. gigas*) and obtained various tissues listed below or obtained oyster samples at different stages in development for transcriptomic sequencing. We obtained the transcriptomic data and applied *in silico* methods described in chapter II for transcript quantification. To this end, we quantified these transcripts across different stages of development and tissues. Respectively, egg, morula, blastula, gastrula, trochophore larva 1, pediveliger, and spat for developmental stages and L-palps, D-gland, hemolymph,

mantle, gill, M-gonad, F-gonad, adductor muscle, and D-shape for tissues. Likewise, we quantified sRNAs from the stages and tissues listed above. Following quantification, reads were normalized respectively to FPKM and RPM for stage and tissue libraries. In addition, we estimated the overlapping read pairs in the sRNA libraries to observe reads generated either through the ping-pong cycle or by Dcr.

To estimate RdRP abundance in oysters challenged with oyster herpesvirus type1 (OsHV-1), we obtained transcriptomic data of oysters exposed to OsHV-1 infected oysters kept at 21°C, 26°C and 29°C (for 0H, 48H, and 96H) from the NCBI database (T. J. Green et al., 2014). Reads were trimmed with trimmomatic and mapped with bowtie2 to the current version of the pacific oyster (*C. gigas*) genome (GCA_902806645.1) (Bolger et al., 2014; Langmead & Salzberg, 2012). Mapped reads were then intersected with annotation file containing coordinates with mRNA in the 3rd column using bedtools multicov. Following this, we compared overall RdRP expression to the entire gene sets as described in chapter II. In addition, we estimated the amount of some genes relevant for antiviral response including Tol-like receptor 1 (TLR1), myeloid differentiation factor 88 (MYD88) and retinoic acid-inducible gene (RIG 1 and 4) all of which have been shown to be responsive in the presence of dsRNA viruses (T. Green & Speck, 2018). FPKM values of the entire gene set where obtained and ranked from the most to the least expressed gene. Based on rank and expression level (Log FPKM)), gene expression was plotted using the ggplot function in R. Further, differential gene expression analysis was performed using the transcriptomic data mentioned above to ascertain differences in gene expression among immune related genes and RdRPs at various points of exposure (0H and 96H) to OsHV-1 infection and at the different temperatures (21°C, 26°C, and 29°C). Further, differential

expressions were determined using the DESeq2 package in R (Love et al., 2014) for low, mid, and high (21°C, 26°C, and 29°C) temperatures transcriptomic data at 0H, 48H, and 96H post infection.

5.4 Results

5.4.1 RNAi factors in bivalves

Bivalves like oysters have a unique set of RNAi factors in comparison to other lophotrochozoan members (Fig 5.1A). Animals in this clade encode relevant RNAi proteins including homologs of AGO (2), Dcr (1), PIWI (3) and RdRPs (2-4). Analysis of these mix of RNAi components suggest the potentially to identify various endogenous sRNA biogenesis pathways in this animal. In our analysis, we observed that AGOs in this animal are found clustering with *D. melanogaster* AGO1 which is known to be implicated in fly's miRNA recruitment (Fig 5.1B). However, we described in chapter II, the potential for similar AGO proteins to have overlapping functions as reported for several plant, human and worm AGOs. Also, a single Dcr with relevant helicase residues can efficiently process dsRNA as seen also in the above-mentioned eukaryotes (Su et al., 2009b). Further, unlike in worms where certain RdRPs (*rrf*-1, 2 and EGO1) can initiate synthesis *de novo* as described in chapter I and II, RdRPs in oysters are seen clustering with worm *rrf*-3 which is known to behave differently from the *de novo* RdRPs of the worm, *C. elegans* (Fig 5.1C). This clustering mean we can predict the polymerization mechanism in this animal like what we observed for amphioxus RdRPs in chapter II where the proteins lack the non-processive loop (Fig 5.1D).

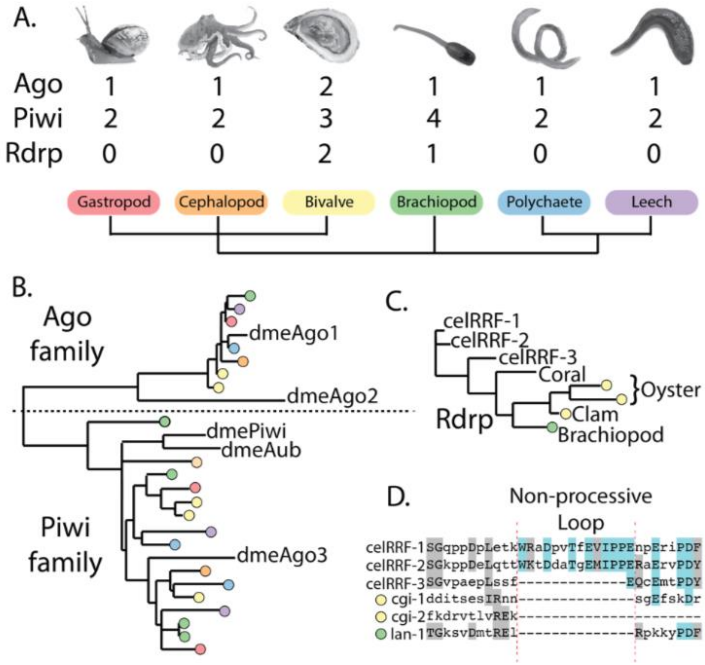


Figure 5.1 *RNAi factors in Lophotrochozoans.*

(A) RNAi factors in Lottia gigantea (gastropod), Octopus bimaculoides (cephalopod), Crassostrea gigas (bivalve), Lingula anatina (brachiopod), Capitella teleta (Polycheate), and Helobdella robusta (leech). (B) Phylogenetic comparison of lophotrochozoan Ago/PIWI's to D. melanogaster. Color scheme represent species depicted in part "A". Homology of brachiopod, bivalve, and coral RdRP to C. elegans RdRP is shown in C. (A) Multiple sequence alignment showing absence of non-processive loop in oyster, C. elegans rrf-3, and brachiopod RdRP.

5.4.2 Stage and tissue specific expression of RNAi factors during development of *C. gigas*

Analysis RNAi factors in the pacific oyster transcriptome shows an interesting expression pattern between stages during development and adult tissues. While PIWI class of proteins as well as the miRNA AGO1 appear to be highly expressed during development, AGO2 and the two RdRPs analyzed were found to show more expression in adult tissues with RdRP2 expressing in the trochophore larva stage (Fig 5.2A). This pattern

creates an opposing expression profile of RNAi factors in oysters between development and adult tissues (Fig 5.2B).

The above expression pattern suggests that there could be a similar opposing expression in sRNA profile across these libraries. As expected, sRNA profiling shows the expected pattern with piRNA like reads being abundant across developmental stages and siRNA/miRNA sized reads populating the adult tissues (Fig 5.2C). Further, the expression profile between these factors has immense potential in implicating oyster RdRPs in sRNA biology in adult tissues while suggesting an embryogenesis role for oyster PIWIs. Whether or not this role is associated with RNAi remains undetermined.

Using these same datasets, we examined read pair overlaps to understand the biogenesis of the sRNAs. Unexpectedly, we observe the ping-pong signature across all stages and tissues in all size ranges (Fig 5.3). While PIWI expression is highest in embryos and gonads it's expression can be detected in other stages/tissues such that it is sufficient to explain functioning ping-pong biogenesis. In fact, this is one of the most striking examples of somatic piRNAs observed to date. We also found evidence of Dcr processing in gills and digestive tissues, which correlates with high expression of RdRP1 (Fig 5.2A). Hemolymph and female gonads stand out from other samples in that they cluster with somatic tissue in terms of RNAi factor expression yet have clear accumulation of longer size sRNAs (Fig 5.2C).

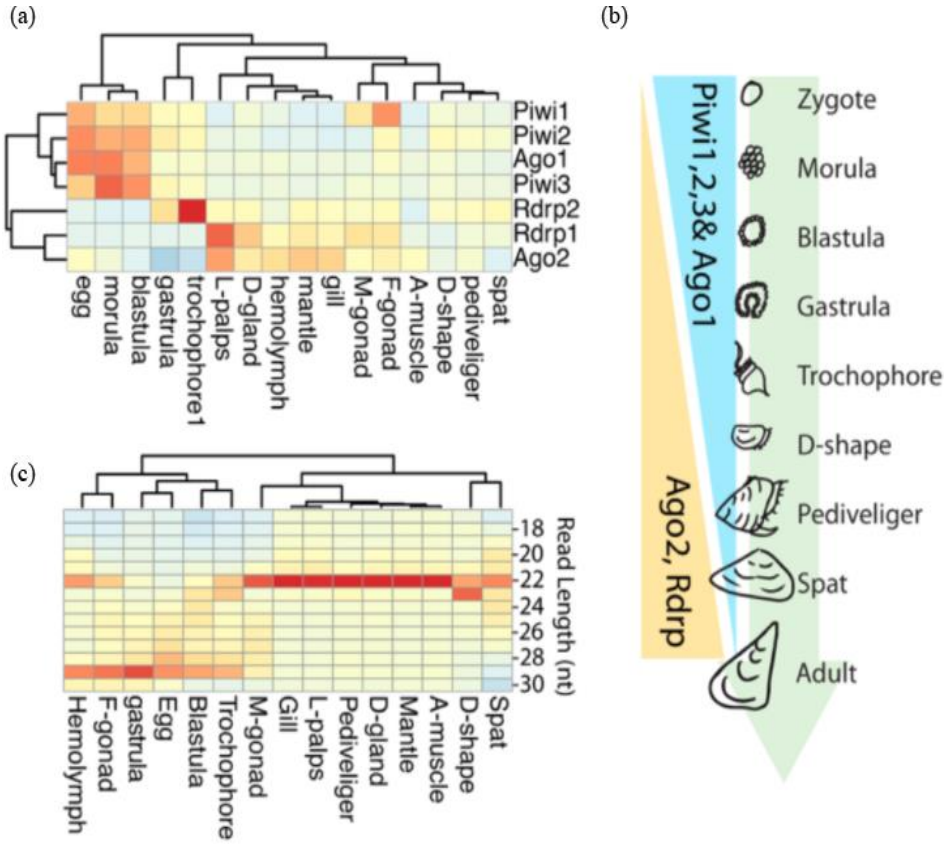


Figure 5.2 *Analysis of RNAi factors and sRNA populations during development and adult tissues of C. gigas.*

(A) Per gene normalized expression of RNAi factors (FPKM). (B) Diagram showing switch of Ago/PIWI and RdRP expression during development. (C) Size distribution of sRNA normalized for stage and tissue (RPM).

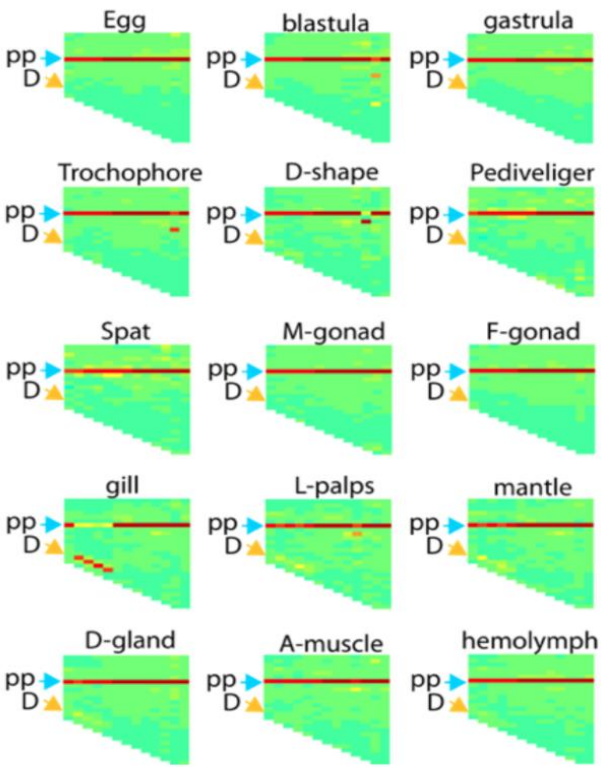


Figure 5.3 *Identity of sRNA population in C. gigas. Overlap probabilities of read pairs in developmental stages and adult tissues.*

Heatmap columns correspond to sRNAs ranging from 18 nt (right) to 30nt (left). Rows are specific overlaps beginning with 4 nt at top to 30 nt at bottom. (Example on the right) "pp" indicates 10-nt overlap ping pong signature. "D" indicates Dcr processing.

5.4.3 Elevated response of RdRP in OsHV-1 challenged oysters at high temperatures

To observe how RdRPs in oysters respond to viral presence, we quantified gene expression of all gene set as described above. Firstly, we observed ranking of expression of key antiviral response genes including TRL1, MYD88, and RIG1 in infected oysters kept either at 21°C or 29°C and at the start of the experiment (0H) or at the end of the experiment (96H post-infection) (Fig 5.4, A-B) (T. J. Green et al., 2014). Also, we observed four RdRPs and their ranking among other genes (Fig 5.4, B-C). Rankings of

these proteins were observed with respect to the median gene expression of all genes at 21°C and 29°C.

Based on this analysis, TLR1 was observed to have an expression that exceeds the median expression level at 21°C and at both 0H and 96H. MYD88 was however seen to exceed this median only at the 96th H, showing that the gene became elevated upon infection. Also, RIG1 showed a slight increase at the 96th H post-infection but did not exceed the median level (Fig 5.4A). Similarly, in the 29°C oysters, TLR1 had its normal expression above median as seen in the 21°C at both 0H and 96H. MYD88 and RIG1 were observed to express almost at the same rate and below the median expression value at both 0H and 96 H of the experiment. This observation shows that MYD88 expression was reduced greatly from above median at 21^{o}C to below median at 29°C (Fig 5.4B) so that its expression is slightly higher at 0H. Taken together, the expression level here suggests a preference for antiviral response at a lower temperature (21°C). Fluctuations in temperature as well as age have been reported to have impact on antiviral genes response when oysters are challenged with viruses (T. J. Green et al., 2014; T. Green & Speck, 2018).

Further, our analysis of RdRP ranking and expression level revealed that at least two of the four oyster RdRPs ranked high in the gene ranking order with RdRP 1 having high expression levels at 21°C (0H and 96H) that exceeds the median expression level. RdRP2 on the other hand failed to exceed median expression. However, we observed an increase in RdRP2 expression level at 96H post infection (Fig 5.4C). Furthermore, observation of RdRP expression at 29°C post-infection revealed a fascinating difference in expression between RdRP2 at 0H and 96H post-infection such that its expression exceeded

the median value slightly at 96H post-infection. We also observed a difference in rank for RdRP1 at 96H post infection although its expression at both time points is still above median level (Fig 5.4D). RdRP3 and 4 however, ranked poorly and failed to cross the median line. Taken together, this rank-dependent gene expression analysis suggest that oyster RdRP expression is higher at elevated temperature (29°C) and increases as viral infection surge, suggesting a potential role for RdRP1 and RdRP2 in viral life cycle at elevated temperatures where innate antiviral response genes are insufficient.

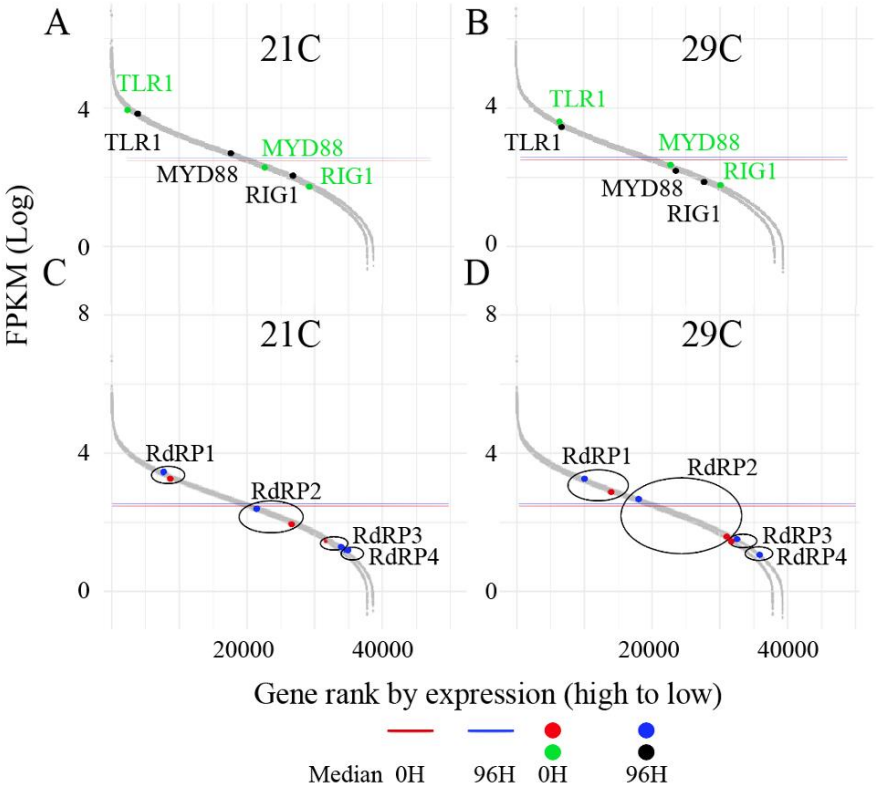


Figure 5.4 *Response of immune genes and RdRPs in oysters challenged with OsHV-1.*

(A and B) shows how 3 antiviral genes rank among all genes based on log (FPKM) at onset (0H) and post-infection (96H) respectively at 21°C and 29°C. (C and D) show similar expression profile rank in A and B with RdRPs highlighted instead of antiviral genes. Circled dotes represent each RdRP in both time points (0H and 96H post-infection).

To further, observe differential expression of immune genes versus RdRPs at the different temperatures, we performed differential gene expression analysis by comparing fold changes (Log2 Fold Change or l2fc) in expression during the exposure periods of 0H/48H and 0H/96H at 21°C, 26°C, and 29°C (Fig 5.5). The result of this analysis revealed that at 21°C and at either 2 days (48H) or 4 days (96H) post-infection, immune genes including RIG1, MYD88, TLR1, and TLR4 showed some level of difference in expression compared to the start of the experiment (0H) (Fig 5.5A). These expressions are however not statistically significant enough to conclude that the genes are differentially expressed (l2fc < 2 and adjusted p-value or padj > 0.05). RdRPs are also poorly differentially expressed at this temperature and exposure periods except for RdRP4 (l2fc > 2 and padj = NA) (Fig 5.5B). Without an adjusted p-value, this estimate cannot be considered statistically significant. Further, observation of the genes at a slightly higher temperature (26°C) revealed that the immune genes observed here are maximally expressed with RIG1 and TLR1 having fold change values above 3 when compared to the 0H of the same temperature. Similarly, RdRPs 1&2 were shown to have fold changes of ~2 and ~4 respectively at both 48H and 96H post infection. RdRPs 3&4 were not considered as padj values were not reported. At 29°C, we observed a fold change of RIG1 and TLR1 expressions at both 48H and 96H (Fig 5.5A). On the other hand, RdRP2 appeared to maintain its expression at this temperature (29°C) increasing from ~2 to ~4 fold from 48H to 96H post infection (Fig 5.5B). Taken together, we revealed here a possible alternating viral response mechanism in oysters involving immune response genes and RdRPs during different temperature exposures throughout the year.

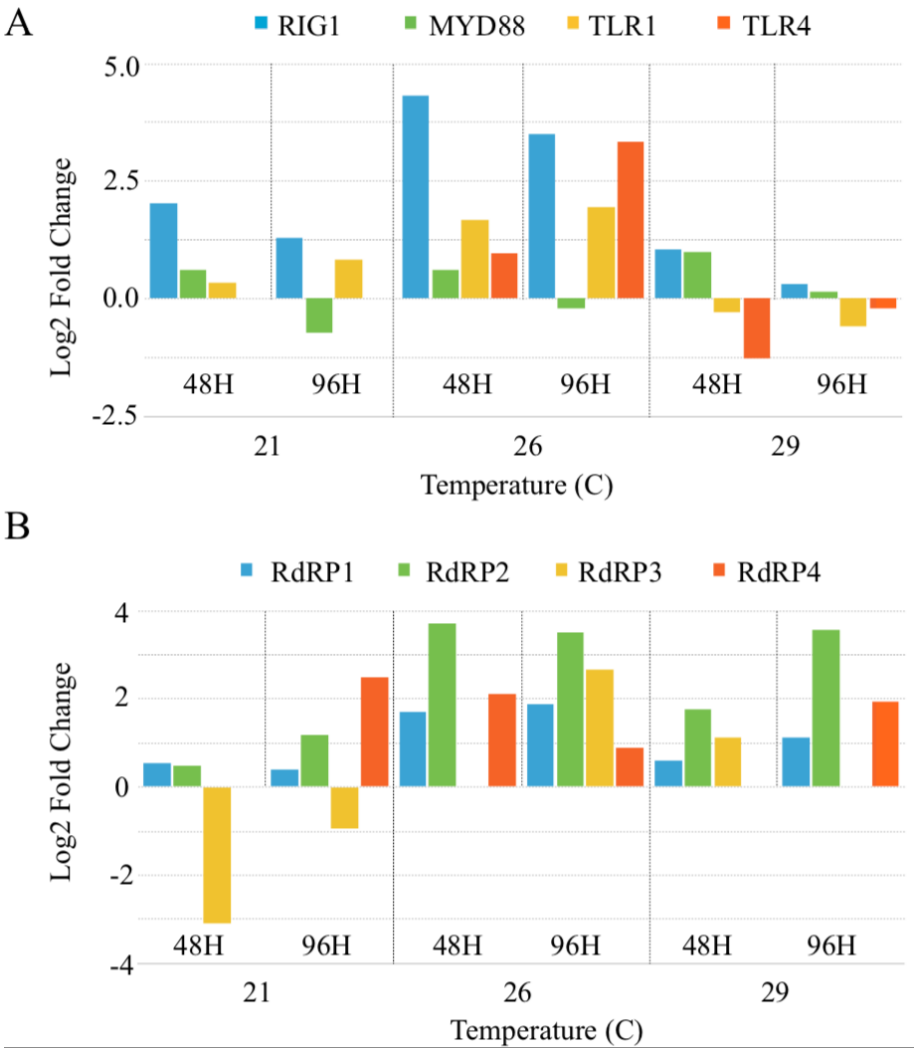


Figure 5.5 *Differential expression analysis of immune vs RdRP genes showing fold change data*

A and B respectively show fold change information of immune related genes and four oyster RdRPs. Y-axis show log2 fold change values while x-axis show temperature in Celsius.

5.5 Conclusion

In this chapter, we have explored various RNAi factors in the pacific oysters and deduced how expression profiles can influence sRNA profiles in this animal. Further, RdRP expression profiles were found to be surprisingly abundant in host tissues coinciding with evidence of short read processing that comprise of both miRNAs and potentially

siRNAs. Tissue specific observation of read pairs produced by Dcr show that gills (greatest), L-palp and mantle are enriched in Dcr signatures when compared to other tissues or embryogenic stages. These 3 tissues are also known to be the major tissues exposed to the surrounding increasing their exposure to environmental nucleic acids or viruses. Surprisingly, these tissues also show a high expression of RdRP1 and 2, suggesting a possible connection between RdRP and environmental viruses.

To further, explore this connection, we queried RdRP expression in oysters challenged with OsHV-1 and observed increased responses in two of the RdRPs at 96H post infection. Although oysters are known to possess an innate antiviral response pathway, their activity have been known to be impacted by temperature fluctuations and age (T. J. Green et al., 2014; T. Green & Speck, 2018). The results shown here suggest a scenario where RdRPs preferentially responds to viral infections at elevated temperatures when immune genes are unable to perform in this animal. To conclude, we predict that this pattern suggests a scenario where elevated temperatures hamper antiviral genes in oysters, paving the way for temporary control of viral mRNA translation as they enter the cytoplasm from the nucleus (for DNA viruses) or from direct release into the cytoplasm (for RNA viruses) until appropriate temperatures are restored. We therefore suggest here that oyster RdRPs have a role in viral life cycles especially at elevated temperatures.

CHAPTER VI – CONCLUSION AND FUTURE DIRECTIONS

6.1 Small RNA characterization pipeline can improve design and efficacy of RNAi

In this chapter, the need for establishing an in-silico sRNA characterization pipeline is addressed that could impact RNAi experimental design. This approach can be particularly useful for researchers with minimal to no expertise in using computational approaches to understand gene functions or to define gene regulatory mechanisms especially in animals where biochemical methods can be challenging or absent.

To test pipeline functionality, we applied the methods accordingly to a high-throughput sequencing data of the WCR and characterized loci with high production of siRNAs, and piRNAs based on a defined parameter that can be adjusted by users as need. The characterized sRNA from a region of high Dcr activity could potentially enter the RNAi pathway more effectively than randomly selected sRNA candidate. Further, to ensure that sRNA characterization can impact RNAi efficacy, we tested the ability of a characterized siRNA to down regulate actin transcripts in the two-spotted spider mite, *T. urticae* through engineering of siRNA into actin dsRNA. Compared to un-altered dsRNA, several configurations of engineered dsRNA were shown to greatly influence actin expression confirming that candidate siRNA selection can influence RNAi efficacy. Therefore, we have shown that this engineering approach can have great implications for other researchers to induce better gene knockdown.

6.1.1 Future direction

While this work has important implication for RNAi as we have shown that different configuration impacts gene knockdown, and efficacy can be greatly improved through testing of a wide array of dsRNA configuration and selecting the most efficacious

model for RNAi. Also, while the dsRNA delivery method we used here influenced RNA, more efficient methods need to be developed as it is quite expensive to use the soaking method on a large scale such as a farmland infested by insect pests.

6.2 Sporadic activity of Amphioxus RdRP is associated with RNAi

In eukaryotes, RdRPs have one main function, RNAi. While these proteins have an essential role that have been shown to impact development in worms and other eukaryotes, their role in a host of other animals where they are encoded seem to be poorly explored. Also, animal RdRPs are highly labile. As a result, conservation of homologs is non-uniform throughout the different animal phyla where they are encoded. Among chordates, cephalochordate members such as amphioxus which have been the focal point in a variety of studies exploring the evolutionary events resulting in vertebrate divergence from their chordate ancestors. These animals represent the closest extant relatives to chordate ancestors. Surprisingly, despite the evolutionary events that resulted in RdRP loss from various animal lineages, amphioxus appear to possess the most RdRP homologs (6 RdRPs) whereas, some Urochordate members only encode half this number in their genome.

Further, it had been reported that amphioxus RdRPs do not have a role in RNAi. However, despite the important niche (relative to vertebrates) and number of RdRPs Amphioxus encode alongside possessing relevant RNAi components, we believe a vestigial role seem likely, warranting the need for more investigation. In this dissertation, we used both biochemical and computational methods to investigate individual locus in search of potential RdRP synthesized reads that enter Dcr pathway. We observed in this work that RdRPs generate long dsRNA that a cleaved by Dcr. Further, RdRP-associated siRNAs appear to have a role in TE repression especially in the somatic tissues. This

suggests a potential role for Amphioxus RdRP in somatic genome protection from mobile elements, implicating the proteins in the RNAi pathway of the animal. In conclusion, although RdRPs have been completely replaced by other RdRP-independent mechanism for genome protection or viral dsRNA processing as seen in fruit flies and humans, their function in in animals where they are encoded have not been completely abrogated. This function as we have shown here is tied to RNAi.

6.2.1 Future direction

While the nascent transcript labeling method we employed here was able to recover a handful of loci predicted to be targeted by RdRP, we believe the labelling was not efficient enough. More optimization of labelling conditions is therefore needed to recover more target loci and possibly expand on the discussion about RdRPs role in amphioxus RNAi. Efficiency can either be in the way of increased 4sU (the labelling molecule) or generation of primary cells or cell line followed by labelling of nascent transcripts on cells rather that whole animal as used here.

6.3 Potential RdRP mediated viral control in oysters at elevated temperatures

Oysters constitute sources of reef life, food for both saltwater animals as well as for humans, revenue source among others. Owing to fishery pressures and diseases impacting oysters in the wild, several global efforts are currently in place to ensure their survival. Although these efforts are ongoing, genetic methods that can be employed to engineer oysters to withstand diseases are inadequate or lacking.

To explore genetic tools for this animal, adequate understanding of RNAi components as well as endogenous sRNA biogenesis pathways are relevant. This necessitated why we explored RdRPs in this animal and their potential implication in viral

life cycle. In this work, we showed that Bivalves comprise one of few animals in the lophotrochozoan subphylum where RdRPs are encoded. Upon observing transcript expression, we found an opposing expression between RNAi components and sRNA population. Specifically, PIWI and AGO1 expression during embryogenesis mirrors piRNA production in corresponding tissues while RdRPs and AGO2-like protein are preferentially transcribed in the adult tissues alongside short, sRNAs (~22nt). Further investigation revealed that tissues namely, gill, L-palp, and mantle which are often exposed to the surrounding water produced sRNAs with read length and read pair characteristic of Dcr cleavage. This read accumulation is somewhat in line with expression of RdRP in these tissues, suggesting the potential involvement of RdRPs in their synthesis.

Also, we showed in this work that RdRP expression in the pacific oyster is elevated at higher temperatures in oysters infected with OsHV-1 while some of the antiviral genes such as MYD88 that had higher expression at lower temperature showed lower expression levels at higher temperature. To conclude, we predict here that this pattern suggests a scenario where elevated temperatures hamper antiviral genes in oysters, paving the way for temporary control of viral mRNA translation as they enter the cytoplasm from the nucleus (for OsHV-1 a DNA viruses) or from direct release into the cytoplasm (for RNA viruses) until appropriate temperatures are restored. In addition, identifying a role for oyster RdRPs have great implication as they can be exploited as drivers of transgenerational RNAi technologies.

6.3.1 Future directions

To verify if indeed oyster RdRPs have a role in viral control at elevated temperatures, more in-vitro studies need to be conducted. This may include a cell-free in-

vitro studies that uses recombinant RdRP with a viral ssRNA mimic generated through in-vitro transcription or an *in vivo* study at elevated temperatures where oysters are challenged with viruses and relevant antiviral immune response genes are downregulated.

Following several methods to optimize conditions that eliminate contamination and allow cells to grow, we successfully established a protocol that further research for tissue specific cell-type biochemical studies will be based (Fig 6.1). With this method, cells can be cultured and used for short term studies as we have done with several tissues in this chapter. Further, confocal microscopy image was able to show intact cells recovered from the primary cell culture establishment with the nuclei and actin cytoskeleton properly stained (Fig 6.2).

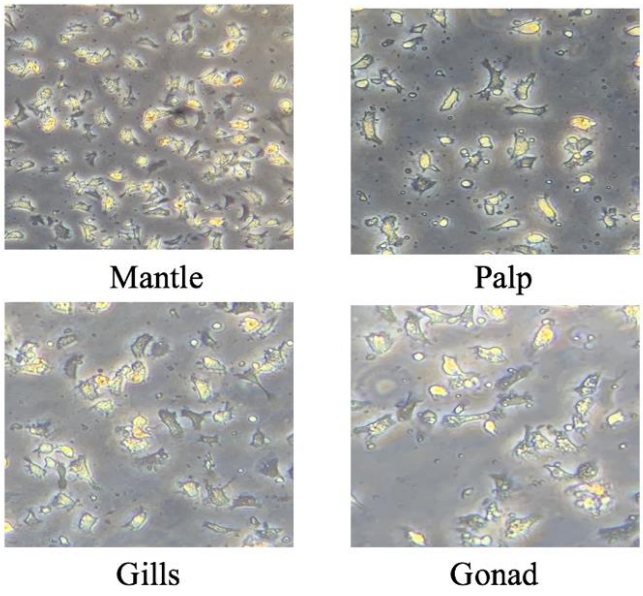


Figure 6.1 *Primary cell culture established from eastern oyster tissues*

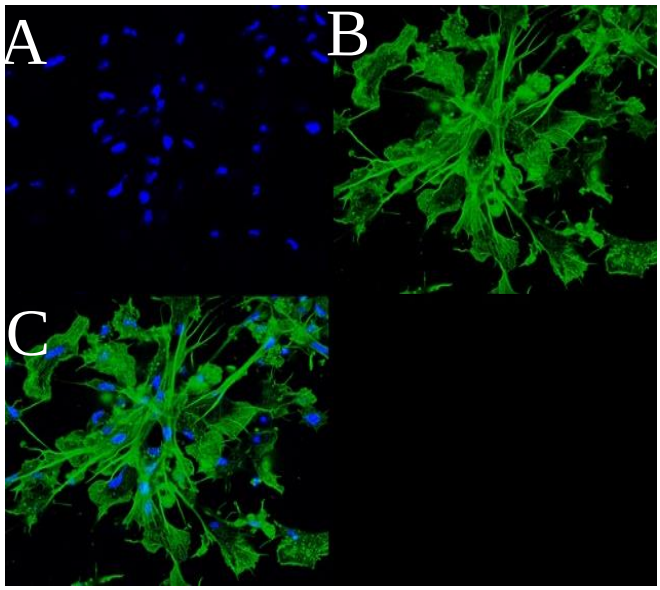


Figure 6.2 *Confocal microscopy showing Oyster ovary primary cell morphology.*

(A) Nuclei staining with Dapi (B) Cytoskeletal staining with Phalloidin (C) A and B merge

REFERENCES

Almeida, M. V., Andrade-Navarro, M. A., & Ketting, R. F. (2019). Function and Evolution of Nematode RNAi Pathways. *Non-Coding RNA*, *5*(1), 8. https://doi.org/10.3390/ncrna5010008

Almeida, M. V., Dietz, S., Redl, S., Karaulanov, E., Hildebrandt, A., Renz, C., Ulrich, H. D., König, J., Butter, F., & Ketting, R. F. (2018). <scp>GTSF</scp> -1 is required for formation of a functional <scp>RNA</scp> -dependent <scp>RNA</scp> Polymerase complex in *Caenorhabditis elegans*. *The EMBO Journal*, *37*(12). https://doi.org/10.15252/embj.201899325

Antoniewski, C. (2014). *Computing siRNA and piRNA Overlap Signatures* (pp. 135–146). https://doi.org/10.1007/978-1-4939-0931-5_12

Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., & Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *The EMBO Journal*, *26*(24), 5007–5019. https://doi.org/10.1038/sj.emboj.7601910

Bar-Nun, S., & Glickman, M. H. (2012). Proteasomal AAA-ATPases: Structure and function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1823*(1), 67–82. https://doi.org/https://doi.org/10.1016/j.bbamcr.2011.07.009

Bassanezi, R. B., Lopes, S. A., de Miranda, M. P., Wulff, N. A., Volpe, H. X. L., & Ayres, A. J. (2020). Overview of citrus huanglongbing spread and management strategies in Brazil. *Tropical Plant Pathology*, *45*(3), 251–264. https://doi.org/10.1007/s40858-020-00343-y

Beck, M. W., Brumbaugh, R. D., Airoldi, L., Carranza, A., Coen, L. D., Crawford, C., Defeo, O., Edgar, G. J., Hancock, B., Kay, M. C., Lenihan, H. S., Luckenbach, M.

W., Toropova, C. L., Zhang, G., & Guo, X. (2011). Oyster Reefs at Risk and Recommendations for Conservation, Restoration, and Management. *BioScience*, *61*(2), 107–116. https://doi.org/10.1525/bio.2011.61.2.5

Beyer, A. (1997). Sequence analysis of the AAA protein family. *Protein Science : A Publication of the Protein Society*, *6*(10), 2043–2058. https://doi.org/10.1002/pro.5560061001

Blumenfeld, A. L., & Jose, A. M. (2016). Reproducible features of small RNAs in *C. elegans* reveal NU RNAs and provide insights into 22G RNAs and 26G RNAs. *RNA*, *22*(2), 184–192. https://doi.org/10.1261/rna.054551.115

Board OS. (2010). *Ecosystem concepts for sustainable bivalve mariculture*. National Academies Press.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolognesi, R., Ramaseshadri, P., Anderson, J., Bachman, P., Clinton, W., Flannagan, R., Ilagan, O., Lawrence, C., Levine, S., Moar, W., Mueller, G., Tan, J., Uffman, J., Wiggins, E., Heck, G., & Segers, G. (2012). Characterizing the Mechanism of Action of Double-Stranded RNA Activity against Western Corn Rootworm (Diabrotica virgifera virgifera LeConte). *PLOS ONE*, *7*(10), e47534-. https://doi.org/10.1371/journal.pone.0047534

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of

Transposon Activity in Drosophila. *Cell*, *128*(6), 1089–1103. https://doi.org/https://doi.org/10.1016/j.cell.2007.01.043

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., & Croce, C. M. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15524–15529. https://doi.org/10.1073/pnas.242606799

Cerutti, H., & Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics*, *50*(2), 81–99. https://doi.org/10.1007/s00294-006-0078-x

Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., Maruthi, K., Eng, E. T., Vatandaslar, H., Chait, B. T., Kapoor, T. M., Darst, S. A., & Campbell, E. A. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, *182*(6), 1560-1573.e13. https://doi.org/https://doi.org/10.1016/j.cell.2020.07.033

Cogoni, C., & Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature*, *399*(6732), 166–169. https://doi.org/10.1038/20215

D. Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, *3*(25), 731. https://doi.org/10.21105/joss.00731

Dahlberg, J. E., & Lund, E. (1998). Functions of the GTPase Ran in RNA export from the nucleus. *Current Opinion in Cell Biology*, *10*(3), 400–408. https://doi.org/https://doi.org/10.1016/S0955-0674(98)80017-3

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., & Baulcombe, D. C. (2000). An RNA-Dependent RNA Polymerase Gene in Arabidopsis Is Required for Posttranscriptional Gene Silencing Mediated by a Transgene but Not by a Virus. *Cell*, *101*(5), 543–553. https://doi.org/https://doi.org/10.1016/S0092-8674(00)80864-8

de Farias, S. T., dos Santos Junior, A. P., Rêgo, T. G., & José, M. v. (2017). Origin and Evolution of RNA-Dependent RNA Polymerase. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00125

Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology*, *3*(10), e314–e314. https://doi.org/10.1371/journal.pbio.0030314

Dermauw, W., Wybouw, N., Rombauts, S., Menten, B., Vontas, J., Grbić, M., Clark, R. M., Feyereisen, R., & van Leeuwen, T. (2013). A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences*, *110*(2). https://doi.org/10.1073/pnas.1213214110

Elmayan, T., Balzergue, S., Béon, F., Bourdon, V., Daubremet, J., Guénet, Y., Mourrain, P., Palauqui, J.-C., Vernhettes, S., Vialle, T., Wostrikoff, K., & Vaucheret, H. (1998). Arabidopsis Mutants Impaired in Cosuppression. *The Plant Cell*, *10*(10), 1747–1757. https://doi.org/10.1105/tpc.10.10.1747

Ferrer-Orta, C., Arias, A., Escarmís, C., & Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. *Current Opinion in Structural Biology*, *16*(1), 27–34. https://doi.org/https://doi.org/10.1016/j.sbi.2005.12.002

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(6669), 806–811. https://doi.org/10.1038/35888

Flynt, A., Liu, N., Martin, R., & Lai, E. C. (2009). Dicing of viral replication intermediates during silencing of latent *Drosophila* viruses. *Proceedings of the National Academy of Sciences*, *106*(13), 5270–5275. https://doi.org/10.1073/pnas.0813412106

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, *40*(1), 37–52. https://doi.org/10.1093/nar/gkr688

Gent, J. I., Schvarzstein, M., Villeneuve, A. M., Gu, S. G., Jantsch, V., Fire, A. Z., & Baudrimont, A. (2009). A *Caenorhabditis elegans* RNA-Directed RNA Polymerase in Sperm Development and Endogenous RNA Interference. *Genetics*, *183*(4), 1297–1314. https://doi.org/10.1534/genetics.109.109686

Graham, J., Gottwald, T., & Setamou, M. (2020). Status of Huanglongbing (HLB) outbreaks in Florida, California and Texas. *Tropical Plant Pathology*, *45*(3), 265–278. https://doi.org/10.1007/s40858-020-00335-y

Grbić, M., van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Thi Ngoc, P. C., Ortego, F., Hernández-Crespo, P., Diaz, I.,

Martinez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R. M., Djuranović, S., … van de Peer, Y. (2011). The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature*, *479*(7374), 487–492. https://doi.org/10.1038/nature10640

Green, T. J., Montagnani, C., Benkendorff, K., Robinson, N., & Speck, P. (2014). Ontogeny and water temperature influences the antiviral response of the Pacific oyster, Crassostrea gigas. *Fish & Shellfish Immunology*, *36*(1), 151–157. https://doi.org/10.1016/j.fsi.2013.10.026

Green, T., & Speck, P. (2018). Antiviral Defense and Innate Immune Memory in the Oyster. *Viruses*, *10*(3), 133. https://doi.org/10.3390/v10030133

Guo, S., & Kemphues, K. J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, *81*(4), 611–620. https://doi.org/https://doi.org/10.1016/0092-8674(95)90082-9

Gurung, C., Fendereski, M., Sapkota, K., Guo, J., Huang, F., & Guo, Y.-L. (2021). Dicer represses the interferon response and the double-stranded RNA-activated protein kinase pathway in mouse embryonic stem cells. *Journal of Biological Chemistry*, *296*, 100264. https://doi.org/https://doi.org/10.1016/j.jbc.2021.100264

György, H., Juanita, M., E, P. A., Éva, B., Thomas, T., & D, Z. P. (2001). A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. *Science*, *293*(5531), 834–838. https://doi.org/10.1126/science.1062961

Hackenberg, M., Rodríguez-Ezpeleta, N., & Aransay, A. M. (2011). miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Research*, *39*(Web Server issue), W132–W138. https://doi.org/10.1093/nar/gkr247

Hammond, S. M., Bernstein, E., Beach, D., & Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, *404*(6775), 293–296. https://doi.org/10.1038/35005107

Han, B. W., Wang, W., Zamore, P. D., & Weng, Z. (2015). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics (Oxford, England)*, *31*(4), 593–595. https://doi.org/10.1093/bioinformatics/btu647

Hannon, G. J. (n.d.). *FASTX-Toolkit*. Http://Hannonlab.Cshl.Edu/Fastx_toolkit.

Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R., & Meister, G. (2013). Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nature Structural & Molecular Biology*, *20*(7), 814–817. https://doi.org/10.1038/nsmb.2577

Head, G. P., Carroll, M. W., Evans, S. P., Rule, D. M., Willse, A. R., Clark, T. L., Storer, N. P., Flannagan, R. D., Samuel, L. W., & Meinke, L. J. (2017). Evaluation of SmartStax and SmartStax PRO maize against western corn rootworm and northern corn rootworm: efficacy and resistance management. *Pest Management Science*, *73*(9), 1883–1899. https://doi.org/https://doi.org/10.1002/ps.4554

Hohn, T., & Vazquez, F. (2011). RNA silencing pathways of plants: Silencing and its suppression by plant DNA viruses. *Biochimica et Biophysica Acta (BBA) - Gene*

Regulatory Mechanisms, *1809*(11), 588–600. https://doi.org/https://doi.org/10.1016/j.bbagrm.2011.06.002

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, *35*(suppl_2), W585–W587. https://doi.org/10.1093/nar/gkm259

Huang, J., Wang, F., Argyris, E., Chen, K., Liang, Z., Tian, H., Huang, W., Squires, K., Verlinghieri, G., & Zhang, H. (2007). Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. *Nature Medicine*, *13*(10), 1241–1247. https://doi.org/10.1038/nm1639

Huang, X., Fejes Tóth, K., & Aravin, A. A. (2017). piRNA Biogenesis in Drosophila melanogaster. *Trends in Genetics : TIG*, *33*(11), 882–894. https://doi.org/10.1016/j.tig.2017.09.002

Ivashuta, S., Zhang, Y., Wiggins, B. E., Ramaseshadri, P., Segers, G. C., Johnson, S., Meyer, S. E., Kerstetter, R. A., McNulty, B. C., Bolognesi, R., & Heck, G. R. (2015). Environmental RNAi in herbivorous insects. *RNA*, *21*(5), 840–850. https://doi.org/10.1261/rna.048116.114

Iyer, L. M., Koonin, E. v, & Aravind, L. (2003). Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Structural Biology*, *3*(1), 1. https://doi.org/10.1186/1472-6807-3-1

J, M. I., Kaihong, Z., Fei, L., Adrian, R., N, B. A., Zacheus, C. W., D, A. P., & A, D. J. (2006). Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*, *311*(5758), 195–198. https://doi.org/10.1126/science.1121638

Ji-Joon, S., K, S. S., J, H. G., & Leemor, J.-T. (2004). Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science*, *305*(5689), 1434–1437. https://doi.org/10.1126/science.1102514

Kaplan, D. L., & O'Donnell, M. (2002). DnaB Drives DNA Branch Migration and Dislodges Proteins While Encircling Two DNA Strands. *Molecular Cell*, *10*(3), 647–657. https://doi.org/https://doi.org/10.1016/S1097-2765(02)00642-1

Kidwell, M. A., Chan, J. M., & Doudna, J. A. (2014). Evolutionarily Conserved Roles of the Dicer Helicase Domain in Regulating RNA Interference Processing. *Journal of Biological Chemistry*, *289*(41), 28352–28362. https://doi.org/10.1074/jbc.M114.589051

Kolde, R. (2012). *Pheatmap: pretty heatmaps*. *61*(915).

Kowalinski, E., Kögel, A., Ebert, J., Reichelt, P., Stegmann, E., Habermann, B., & Conti, E. (2016). Structure of a Cytoplasmic 11-Subunit RNA Exosome Complex. *Molecular Cell*, *63*(1), 125–134. https://doi.org/10.1016/j.molcel.2016.05.028

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Kwak, P. B., & Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature Structural & Molecular Biology*, *19*(2), 145–151. https://doi.org/10.1038/nsmb.2232

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., & Kingston, R. E. (2006). Characterization of the piRNA Complex from Rat Testes. *Science*, *313*(5785), 363–367. https://doi.org/10.1126/science.1130164

Lau, P.-W., Guiley, K. Z., De, N., Potter, C. S., Carragher, B., & MacRae, I. J. (2012). The molecular architecture of human Dicer. *Nature Structural & Molecular Biology*, *19*(4), 436–440. https://doi.org/10.1038/nsmb.2268

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993a). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. https://doi.org/https://doi.org/10.1016/0092-8674(93)90529-Y

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993b). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. https://doi.org/https://doi.org/10.1016/0092-8674(93)90529-Y

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., & Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, *425*(6956), 415–419. https://doi.org/10.1038/nature01957

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., & Carthew, R. W. (2004). Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell*, *117*(1), 69–81. https://doi.org/https://doi.org/10.1016/S0092-8674(04)00261-2

Li, C., Vagin, V. v, Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L. W., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng, Z., & Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, *137*(3), 509–521. https://doi.org/10.1016/j.cell.2009.04.027

Li, H., Clum, S., You, S., Ebner, K. E., & Padmanabhan, R. (1999). The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *Journal of Virology*, *73*(4), 3108–3116. https://doi.org/10.1128/JVI.73.4.3108-3116.1999

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Chang, S., & Liu, Y. (2010). RNA interference pathways in filamentous fungi. *Cellular and Molecular Life Sciences : CMLS*, *67*(22), 3849–3863. https://doi.org/10.1007/s00018-010-0471-y

Li, S., & Patel, D. J. (2016). Drosha and Dicer: Slicers cut from the same cloth. *Cell Research*, *26*(5), 511–512. https://doi.org/10.1038/cr.2016.19

Li, Y., & Kowdley, K. v. (2012). MicroRNAs in common human diseases. *Genomics, Proteomics & Bioinformatics*, *10*(5), 246–253. https://doi.org/10.1016/j.gpb.2012.07.005

Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L., & Gao, X. (2018). DEEPre: sequence-based enzyme EC number prediction by deep learning. *Bioinformatics (Oxford, England)*, *34*(5), 760–769. https://doi.org/10.1093/bioinformatics/btx680

Liao, X., Yang, L., Chen, X., & Chen, J. (2017). Identification of microRNA expression profiles in the gill, intestine and hepatic caecum of Branchiostoma belcheri. *Protein & Cell*, *8*(4), 302–307. https://doi.org/10.1007/s13238-016-0365-3

Löfgren, S. E., Frostegård, J., Truedsson, L., Pons-Estel, B. A., D'Alfonso, S., Witte, T., Lauwerys, B. R., Endreffy, E., Kovács, L., Vasconcelos, C., Martins da Silva, B., Kozyrev, S. v, & Alarcón-Riquelme, M. E. (2012). Genetic association of miRNA-146a with systemic lupus erythematosus in Europeans through decreased expression of the gene. *Genes and Immunity*, *13*(3), 268–274. https://doi.org/10.1038/gene.2011.84

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luo, S., & Lu, J. (2017). Silencing of Transposable Elements by piRNAs in Drosophila: An Evolutionary Perspective. *Genomics, Proteomics & Bioinformatics*, *15*(3), 164–176. https://doi.org/https://doi.org/10.1016/j.gpb.2017.01.006

Makeyev, E. v., & Bamford, D. H. (2002). Cellular RNA-Dependent RNA Polymerase Involved in Posttranscriptional Gene Silencing Has Two Distinct Activity Modes. *Molecular Cell*, *10*(6), 1417–1427. https://doi.org/10.1016/S1097-2765(02)00780-3

Marlétaz, F., Firbas, P. N., Maeso, I., Tena, J. J., Bogdanovic, O., Perry, M., Wyatt, C. D. R., de la Calle-Mustienes, E., Bertrand, S., Burguera, D., Acemel, R. D., van

Heeringen, S. J., Naranjo, S., Herrera-Ubeda, C., Skvortsova, K., Jimenez-Gancedo, S., Aldea, D., Marquez, Y., Buono, L., … Irimia, M. (2018). Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature*, *564*(7734), 64–70. https://doi.org/10.1038/s41586-018-0734-6

Medley, J. C., Panzade, G., & Zinovyeva, A. Y. (2021). microRNA strand selection: Unwinding the rules. *Wiley Interdisciplinary Reviews. RNA*, *12*(3), e1627–e1627. https://doi.org/10.1002/wrna.1627

Melanie, S., Thomas, S., Thomas, D., & A, G. R. (2014). DNA Virus Replication Compartments. *Journal of Virology*, *88*(3), 1404–1420. https://doi.org/10.1128/JVI.02046-13

Mohn, F., Handler, D., & Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812–817. https://doi.org/10.1126/science.aaa1039

Mohn, F., Sienski, G., Handler, D., & Brennecke, J. (2014). The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell*, *157*(6), 1364–1379. https://doi.org/https://doi.org/10.1016/j.cell.2014.04.031

Mondal, M., Brown, J. K., & Flynt, A. (2020). Exploiting somatic piRNAs in Bemisia tabaci enables novel gene silencing through RNA feeding. *Life Science Alliance*, *3*(10), e202000731. https://doi.org/10.26508/lsa.202000731

Mondal, M., Mansfield, K., & Flynt, A. (2018). siRNAs and piRNAs collaborate for transposon control in the two-spotted spider mite. *RNA (New York, N.Y.)*, *24*(7), 899–907. https://doi.org/10.1261/rna.065839.118

Mondal, M., Peter, J., Scarbrough, O., & Flynt, A. (2021). Environmental RNAi pathways in the two-spotted spider mite. *BMC Genomics*, *22*(1), 42. https://doi.org/10.1186/s12864-020-07322-2

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Jouette, D., Lacombe, A.-M., Nikic, S., Picault, N., Rémoué, K., Sanial, M., Vo, T.-A., & Vaucheret, H. (2000). Arabidopsis SGS2 and SGS3 Genes Are Required for Posttranscriptional Gene Silencing and Natural Virus Resistance. *Cell*, *101*(5), 533–542. https://doi.org/https://doi.org/10.1016/S0092-8674(00)80863-6

Muhammad, T., Zhang, F., Zhang, Y., & Liang, Y. (2019). RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. *Cells*, *8*(1), 38. https://doi.org/10.3390/cells8010038

Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant Cell*, *2*(4), 279–289. https://doi.org/10.1105/tpc.2.4.279

Newell, R. I. E. (2004). Ecosystem influences of natural and cultivated populations of suspension-feeding bivalve molluscs: a review. *Journal of Shellfish Research*, *23*, 51+. https://link.gale.com/apps/doc/A118543914/AONE?u=anon~b937cabe&sid=googleScholar&xid=e5b425d3

Peter, J. O., Santos-Ortega, Y., & Flynt, A. (2022). Guiding RNAi Design Through Characterization of Endogenous Small RNA Pathways. In L. M. Vaschetto (Ed.), *RNAi Strategies for Pest Management: Methods and Protocols* (pp. 33–47). Springer US. https://doi.org/10.1007/978-1-0716-1633-8_4

Phanstiel, D. H., Boyle, A. P., Araya, C. L., & Snyder, M. P. (2014). Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, *30*(19), 2808–2810. https://doi.org/10.1093/bioinformatics/btu379

Pinzón, N., Bertrand, S., Subirana, L., Busseau, I., Escrivá, H., & Seitz, H. (2019a). Functional lability of RNA-dependent RNA polymerases in animals. *PLOS Genetics*, *15*(2), e1007915. https://doi.org/10.1371/journal.pgen.1007915

Pinzón, N., Bertrand, S., Subirana, L., Busseau, I., Escrivá, H., & Seitz, H. (2019b). Functional lability of RNA-dependent RNA polymerases in animals. *PLOS Genetics*, *15*(2), e1007915-. https://doi.org/10.1371/journal.pgen.1007915

Qiu, Y., Xu, Y., Zhang, Y., Zhou, H., Deng, Y.-Q., Li, X.-F., Miao, M., Zhang, Q., Zhong, B., Hu, Y., Zhang, F.-C., Wu, L., Qin, C.-F., & Zhou, X. (2017). Human Virus-Derived Small RNAs Can Confer Antiviral Immunity in Mammals. *Immunity*, *46*(6), 992-1004.e5. https://doi.org/https://doi.org/10.1016/j.immuni.2017.05.006

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramaseshadri, P., Segers, G., Flannagan, R., Wiggins, E., Clinton, W., Ilagan, O., McNulty, B., Clark, T., & Bolognesi, R. (2013). Physiological and Cellular Responses Caused by RNAi- Mediated Suppression of Snf7 Orthologue in Western Corn Rootworm (Diabrotica virgifera virgifera) Larvae. *PLOS ONE*, *8*(1), e54270-. https://doi.org/10.1371/journal.pone.0054270

Rampersaud, G. C., & Valim, M. F. (2017). 100% citrus juice: Nutritional contribution, dietary benefits, and association with anthropometric measures. *Critical Reviews in Food Science and Nutrition*, *57*(1), 129–140. https://doi.org/10.1080/10408398.2013.862611

Randall, G., Panis, M., Cooper, J. D., Tellinghuisen, T. L., Sukhodolets, K. E., Pfeffer, S., Landthaler, M., Landgraf, P., Kan, S., Lindenbach, B. D., Chien, M., Weir, D. B., Russo, J. J., Ju, J., Brownstein, M. J., Sheridan, R., Sander, C., Zavolan, M., Tuschl, T., & Rice, C. M. (2007). Cellular cofactors affecting hepatitis C virus infection and replication. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(31), 12884–12889. https://doi.org/10.1073/pnas.0704894104

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, *81*, 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Romano, N., & Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular Microbiology*, *6*(22), 3343–3353. https://doi.org/https://doi.org/10.1111/j.1365-2958.1992.tb02202.x

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., & Bartel, D. P. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. *Cell*, *127*(6), 1193–1207. https://doi.org/10.1016/j.cell.2006.10.040

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., & Siomi, M. C. (2009). A regulatory circuit for piwi by the

large Maf gene traffic jam in Drosophila. *Nature*, *461*(7268), 1296–1299. https://doi.org/10.1038/nature08501

Saito, K., & Siomi, M. C. (2010). Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Developmental Cell*, *19*(5), 687–697. https://doi.org/https://doi.org/10.1016/j.devcel.2010.10.011

Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N. M., Schiffer, P. M., Kraus, C., Taylor, M. J., Koutsovoulos, G., Blaxter, M. L., & Miska, E. A. (2015a). Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PLOS Biology*, *13*(2), e1002061. https://doi.org/10.1371/journal.pbio.1002061

Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N. M., Schiffer, P. M., Kraus, C., Taylor, M. J., Koutsovoulos, G., Blaxter, M. L., & Miska, E. A. (2015b). Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biology*, *13*(2), e1002061–e1002061. https://doi.org/10.1371/journal.pbio.1002061

Savin, K. W., Cocks, B. G., Wong, F., Sawbridge, T., Cogan, N., Savage, D., & Warner, S. (2010). A neurotropic herpesvirus infecting the gastropod, abalone, shares ancestry with oyster herpesvirus and a herpesvirus associated with the amphioxus genome. *Virology Journal*, *7*(1), 308. https://doi.org/10.1186/1743-422X-7-308

Schiebel, W., Pélissier, T., Riedel, L., Thalmeir, S., Schiebel, R., Kempe, D., Lottspeich, F., Sänger, H. L., & Wassenegger, M. (1998). Isolation of an RNA-Directed RNA

Polymerase–Specific cDNA Clone from Tomato. *The Plant Cell*, *10*(12), 2087–2101. https://doi.org/10.1105/tpc.10.12.2087

Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., & Zamore, P. D. (2003). Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell*, *115*(2), 199–208. https://doi.org/https://doi.org/10.1016/S0092-8674(03)00759-1

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte Jr, D., & Mello, C. C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. *Cell*, *150*(1), 65–77. https://doi.org/10.1016/j.cell.2012.06.015

Shu, B., & Gong, P. (2016). Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proceedings of the National Academy of Sciences*, *113*(28). https://doi.org/10.1073/pnas.1602591113

Sigova, A., Rhind, N., & Zamore, P. D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in *Schizosaccharomyces pombe*. *Genes & Development*, *18*(19), 2359–2367. https://doi.org/10.1101/gad.1218004

Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., & Maine, E. M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Current Biology*, *10*(4), 169–178. https://doi.org/10.1016/S0960-9822(00)00323-7

Spit, J., Philips, A., Wynant, N., Santos, D., Plaetinck, G., & vanden Broeck, J. (2017). Knockdown of nuclease activity in the gut enhances RNAi efficiency in the Colorado potato beetle, Leptinotarsa decemlineata , but not in the desert locust, Schistocerca gregaria. *Insect Biochemistry and Molecular Biology*, *81*, 103–116. https://doi.org/10.1016/j.ibmb.2017.01.004

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution*, *37*(4), 1237–1239. https://doi.org/10.1093/molbev/msz312

STEIN, P., SVOBODA, P., ANGER, M., & SCHULTZ, R. M. (2003). RNAi: Mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA*, *9*(2), 187–192. https://doi.org/10.1261/rna.2860603

Su, H., Trombly, M. I., Chen, J., & Wang, X. (2009a). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development*, *23*(3), 304–317. https://doi.org/10.1101/gad.1749809

Su, H., Trombly, M. I., Chen, J., & Wang, X. (2009b). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development*, *23*(3), 304–317. https://doi.org/10.1101/gad.1749809

Suzuki, T., Nunes, M. A., España, M. U., Namin, H. H., Jin, P., Bensoussan, N., Zhurov, V., Rahman, T., de Clercq, R., Hilson, P., Grbic, V., & Grbic, M. (2017a). RNAi-based reverse genetics in the chelicerate model Tetranychus urticae: A comparative analysis of five methods for gene silencing. *PLOS ONE*, *12*(7), e0180654. https://doi.org/10.1371/journal.pone.0180654

Suzuki, T., Nunes, M. A., España, M. U., Namin, H. H., Jin, P., Bensoussan, N., Zhurov, V., Rahman, T., de Clercq, R., Hilson, P., Grbic, V., & Grbic, M. (2017b). RNAi-based reverse genetics in the chelicerate model Tetranychus urticae: A comparative analysis of five methods for gene silencing. *PLOS ONE*, *12*(7), e0180654. https://doi.org/10.1371/journal.pone.0180654

Svobodova, E., Kubikova, J., & Svoboda, P. (2016). Production of small RNAs by mammalian Dicer. *Pflügers Archiv - European Journal of Physiology*, *468*(6), 1089–1102. https://doi.org/10.1007/s00424-016-1817-6

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Thivierge, C., Makil, N., Flamand, M., Vasale, J. J., Mello, C. C., Wohlschlegel, J., Conte, D., & Duchaine, T. F. (2012). Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nature Structural & Molecular Biology*, *19*(1), 90–97. https://doi.org/10.1038/nsmb.2186

Tijsterman, M., Okihara, K. L., Thijssen, K., & Plasterk, R. H. A. (2002). PPW-1, a PAZ/PIWI Protein Required for Efficient Germline RNAi, Is Defective in a Natural Isolate of C. elegans. *Current Biology*, *12*(17), 1535–1540. https://doi.org/10.1016/S0960-9822(02)01110-7

Tolia, N. H., & Joshua-Tor, L. (2007). Slicer and the Argonautes. *Nature Chemical Biology*, *3*(1), 36–43. https://doi.org/10.1038/nchembio848

Vasale, J. J., Gu, W., Thivierge, C., Batista, P. J., Claycomb, J. M., Youngman, E. M., Duchaine, T. F., Mello, C. C., & Conte, D. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proceedings of the National Academy of Sciences*, *107*(8), 3582–3587. https://doi.org/10.1073/pnas.0911908107

Vogel, E., Santos, D., Mingels, L., Verdonckt, T.-W., & Broeck, J. vanden. (2019). RNA Interference in Insects: Protecting Beneficials and Controlling Pests. *Frontiers in Physiology*, *9*. https://www.frontiersin.org/article/10.3389/fphys.2018.01912

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., & Martienssen, R. A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science*, *297*(5588), 1833–1837. https://doi.org/10.1126/science.1074973

Vrbsky, J., Akimcheva, S., Watson, J. M., Turner, T. L., Daxinger, L., Vyskot, B., Aufsatz, W., & Riha, K. (2010). siRNA–Mediated Methylation of Arabidopsis Telomeres. *PLoS Genetics*, *6*(6), e1000986. https://doi.org/10.1371/journal.pgen.1000986

Warrener, P., Tamura, J. K., & Collett, M. S. (1993). RNA-stimulated NTPase activity associated with yellow fever virus NS3 protein expressed in bacteria. *Journal of Virology*, *67*(2), 989–996. https://doi.org/10.1128/JVI.67.2.989-996.1993

Welker, N. C., Pavelec, D. M., Nix, D. A., Duchaine, T. F., Kennedy, S., & Bass, B. L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. *RNA (New York, N.Y.)*, *16*(5), 893–903. https://doi.org/10.1261/rna.2122010

Westholm, J. O., Ladewig, E., Okamura, K., Robine, N., & Lai, E. C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA*, *18*(2), 177–192. https://doi.org/10.1261/rna.030627.111

Wickham, H. (2009). *ggplot2*. Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D. (2011a). The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. *The Arabidopsis Book*, *9*, e0146–e0146. https://doi.org/10.1199/tab.0146

Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D. (2011b). The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. *The Arabidopsis Book*, *9*, e0146. https://doi.org/10.1199/tab.0146

Winter, J., Link, S., Witzigmann, D., Hildenbrand, C., Previti, C., & Diederichs, S. (2013). Loop-miRs: active microRNAs generated from single-stranded loop regions. *Nucleic Acids Research*, *41*(10), 5503–5512. https://doi.org/10.1093/nar/gkt251

Xu, F., Wang, X., Feng, Y., Huang, W., Wang, W., Li, L., Fang, X., Que, H., & Zhang, G. (2014). Identification of Conserved and Novel MicroRNAs in the Pacific Oyster Crassostrea gigas by Deep Sequencing. *PLoS ONE*, *9*(8), e104371. https://doi.org/10.1371/journal.pone.0104371

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R. M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G. A., Liu, C.-G., Croce, C. M., & Harris, C. C. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, *9*(3), 189–198. https://doi.org/https://doi.org/10.1016/j.ccr.2006.01.025

Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K. D., Jeffrey, K. L., Chandwani, R., Papapetrou, E. P., Sadelain, M., O'Carroll, D., & Lai, E. C. (2010). Conserved vertebrate *mir-451* provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proceedings of the National Academy of Sciences*, *107*(34), 15163–15168. https://doi.org/10.1073/pnas.1006432107

Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C.-C. G., Tolia, N. H., Joshua-Tor, L., Mitani, S., Simard, M. J., & Mello, C. C. (2006). Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. *Cell*, *127*(4), 747–757. https://doi.org/https://doi.org/10.1016/j.cell.2006.09.033

Yoon, J.-S., Sahoo, D. K., Maiti, I. B., & Palli, S. R. (2018). Identification of target genes for RNAi-mediated control of the Twospotted Spider Mite. *Scientific Reports*, *8*(1), 14687. https://doi.org/10.1038/s41598-018-32742-2

Zhang, C., & Ruvkun, G. (2012). New insights into siRNA amplification and RNAi. *RNA Biology*, *9*(8), 1045–1049. https://doi.org/10.4161/rna.21246

Zhang, H., Kolb, F. A., Brondani, V., Billy, E., & Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *The EMBO Journal*, *21*(21), 5875–5885. https://doi.org/10.1093/emboj/cdf582

Zhang, R., Jing, Y., Zhang, H., Niu, Y., Liu, C., Wang, J., Zen, K., Zhang, C.-Y., & Li, D. (2018). Comprehensive Evolutionary Analysis of the Major RNA-Induced Silencing Complex Members. *Scientific Reports*, *8*(1), 14189. https://doi.org/10.1038/s41598-018-32635-4

Zhou, C. (2020). The status of citrus Huanglongbing in China. *Tropical Plant Pathology*, *45*(3), 279–284. https://doi.org/10.1007/s40858-020-00363-8

Zong, J., Yao, X., Yin, J., Zhang, D., & Ma, H. (2009a). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene*, *447*(1), 29–39. https://doi.org/https://doi.org/10.1016/j.gene.2009.07.004

Zong, J., Yao, X., Yin, J., Zhang, D., & Ma, H. (2009b). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene*, *447*(1), 29–39. https://doi.org/https://doi.org/10.1016/j.gene.2009.07.004

Zukoff, S. N., & Zukoff, A. L. (2017). Host Recognition Responses of Western (Family: Chrysomelidae) Corn Rootworm Larvae to RNA Interference and Bt Corn. *Journal of Insect Science*, *17*(2). https://doi.org/10.1093/jisesa/iex022