5-2021

Investigating piRNA expression and function in regenerating tissue of segmented annelid Capitella teleta

Beatriz Schueng Zancanela

The University of Southern Mississippi

Investigating piRNA expression and function in regenerating tissue of segmented annelid *Capitella teleta*

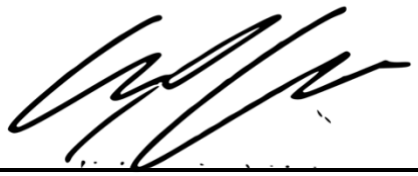
by

Beatriz Schueng Zancanela

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

Spring 2021

Approved by:



Alex Flynt, Ph.D., Thesis Advisor,
School of Biological, Environmental and Earth Sciences

Jacob Schaefer, Ph.D., Director,
School of Biological, Environmental and Earth Sciences

Ellen Weinauer, Ph.D., Dean
Honors College

ABSTRACT

Regeneration can be observed virtually in all animals and previous studies have identified numerous genes involved in this process. In some invertebrates, the P-element induced wimpy testis (PIWI) genes appear to have an essential role. One such organism is *Capitella teleta*, a widely accepted model for annelid development. PIWI proteins are associated with small non-coding RNA called PIWI-interacting RNAs or piRNAs, which are involved in transposon silencing in the germline cells of many animals. It was previously believed that these proteins were only expressed in germline cells, however, recent studies have shown expression in somatic tissues as well. The function of piRNAs in somatic tissues is not well-understood. In this study, RNA samples were collected from the regenerating tissue of *C. teleta* and sequenced to explore the function of piRNAs in this process. Analysis of these data indicated that piRNAs are targeting genes involved in apoptosis and the repression of transposable elements. These results demonstrate that piRNAs influence the regeneration process in annelids through both promoting cell survival and maintaining genome integrity.

Keywords: piRNA, Capitella teleta, GO analysis, PIWI, Transposons

DEDICATION

I would like to dedicate this thesis publication to my sister, who has always believed in me and pushed me to be a better scholar, to my parents for their support in all the endeavors I have chosen to pursue, and lastly, to my best friend Isabelle McKellar for always staying by my side and for pushing me to be the best version of myself.

ACKNOWLEDGMENTS

I would like to thank Dr. Alex Flynt, my thesis advisor, for the opportunity to pursue research in his lab under a topic of high interest to me. I would also like to thank him for all his guidance and help throughout this semester in all aspects of my academic career. I would also like to give a special thank you to the graduate student, Sweta Khanal for your guidance, help, and patience in teaching and explaining to me all of the laboratory techniques and computational analysis protocols. Finally, I would like to thank the Honors College for the opportunity to conduct such in-depth and important research as an undergraduate student.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

TBL3	Transducin Beta Like 3
TTC27	Tetratricopeptide Repeat Domain 27
RACK1	Receptor For Activated C Kinase 1
PIWI	P-element induced wimpy testis
piRNAs	piwi-interacting RNAs
ncRNAs	non-coding RNAs
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
PAZ/AGO	Piwi Argonaut and Zwille/ argonaut
siRNA	small-interfering RNA
miRNA	microRNA
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
DICER1	Endoribonuclease Dicer/Ribonuclease III
dFMRP	Fragile X mental retardation protein
mRNA	messenger RNA
RNAi	RNA interference
PEV	Position effect variegation
HEN1	Small RNA 2'-O-methyltransferase
AUB	Aubergine protein
Yb	Y-box binding protein
Rhi	Rhino protein

H3K9me3	Histone 3 lysine 9 trimethylation
UAP56 (Hel25E)	Helicase at 25E
MIWI	Mouse PIWI
MILI	MIWI-like gene
A-MYB (MYBL1)	Myoblastosis/ MYBL1 (MYB Proto-Oncogene Like 1
GPAT2	Glycerol-3-phosphate-acetyltransferase
Ct-piwi1	*Capitella teleta*-piwi1
Ct-piwi2	*Capitella teleta*-piwi2
MPC	Multipotent progenitor cells
PGZ	Posterior growth zone
cDNA	copyDNA
PCR	Polymerase Chain Reaction
ddNTPs	Di-deoxynucleotides
GO	Gene Ontology
DAG	Directed acyclic graph
FSW	Filtered fresh sea water
BAM	Binary alignment map
SAM	Sequence alignment map
FASTA	Fast-all
SINE	Short Interspaced Nuclear Element
LINE	Long Interspaced Nuclear Element
LTR	Long Terminal Repeat

CHAPTER I: INTRODUCTION

Regeneration capabilities can be observed virtually in all animals suggesting a common evolutionary history for such pathways. However, the question of how these events are correlated historically is challenging, given that most regeneration occurs in limbs or appendages that do not have direct evolutionary counterparts in species from different phyla (Zattara et al., 2018). The animals with the most famous regenerative abilities are worms from the classes Annelida, Nemertean, and Turbellaria (Saló & Baguñà, 2002). These animals appear to derive their regenerative abilities from the maintenance of a population of pluripotent stem cells in their body called neoblasts (Reddien & Alvarado, 2004). The cells in neoblast migrate and become virtually any necessary cell in the damaged body of the worm (Saló & Baguñà, 2002; Reddien & Alvarado, 2004).

Recent studies of stem cell maintenance showed an array of genes that are expressed in both embryonic stem cells and neoblasts, implicating a conserved pathway that maintains the pluripotency of these cell population (Labbé et al., 2012). Among these genes are the human orthologs *TBL3*, *PSD12*, *TTC27*, and *RACK1*. Another gene type that has homologs in many eukaryotic species is the subfamily of the Argonaute proteins, the P-element induced wimpy testis or PIWI proteins (Xu & Sun, 2020). These proteins have been shown to be important in gametogenesis, spermiogenesis, and germline development by helping to determine cell fate (Thomson & Lin, 2009). The PIWI proteins interact with short sequences of non-coding RNAs (ncRNAs) referred to as piRNAs or PIWI-interacting RNAs. Complexes are formed by the interaction of these two components and result in transposon silencing (Ross et al., 2014) through DNA methylation, thus causing epigenic alterations in the genome (Perera et al., 2019).

In somatic tissue, mutations in PIWI proteins have been linked to infertility in *Drosophila melanogaster* studies and loss of regenerative abilities in Planarians (Thomson & Lin, 2009). Moreover, high expression of PIWI proteins have been linked to cancer growth (Liu et al., 2006). The piRNAs initially were believed to be created and present only in germline tissue, thus being a unique feature of this cell population (Perera et al., 2019). In Planarians, PIWI is also found in pluripotent stem cells (Palakodeti et al., 2008). This suggests that piRNA might have a function in somatic cells, in addition to the germline. However, few studies have been conducted to elucidate the role of piRNAs in somatic tissues.

This study conducted small RNA sequencing from regenerating tissue of the segmented marine worm *C. teleta*. This annelid is known for its ability to extensively regenerate body parts following posterior amputation. The presence of PIWI proteins in the gametes and in regenerating tissue of this annelid has been previously documented (Giani et al., 2011). However, to this day, the identity of piRNAs in the somatic regenerating tissues of *C. telata* has not yet been characterized. The aims of this study were to investigate the function of piRNAs in the regenerating tissue of *C. telata* and to determine their interacting genes. Such studies help to further elucidate the novel role of PIWI and piRNA in somatic tissue, are of interest to the field of regenerative medicine, and provide a better understand epigenetic editing of genes through DNA methylation (Perera et al., 2019).

CHAPTER II: LITERATURE REVIEW

Background

Among the animal kingdom, the most well-described model organisms capable of regenerating body parts are the Planarians. These flat worms are unsegmented organisms lacking circulatory, respiratory, and skeletal structures, and have no coelom and yet are capable of extensive body regeneration (Saló & Baguñà, 2002). The mechanisms behind their capacity for regeneration appears to lie in the formation of a blastema and through morphallaxis (Reddien & Alvarado, 2004). The latter is derived from neoblast cells and the former is defined as the remodeling ability based on already existing tissue to restore the organism to its original form and shape (Reddien & Alvarado, 2004). Studies involving molecular tagging of neoblast cells show that hours after amputation, some cells have already migrated from the parenchyma to the site of injury and formed an "organ primordia." A few days later, cells have already began the process of differentiation to reform the lost body part (Saló & Baguñà, 2002). About 30 minutes after an injury has occurred, it is possible to observe a thin layer of epithelium forming around the wound resulting from cell migration and not from cell proliferation (Reddien & Alvarado, 2004). The epithelial to mesenchyme transition occurs from cells localized in the dorsal and ventral portions of the animal, near the site of injury (Reddien & Alvarado, 2004).

In Nemertean, one extensive study done by Zattara et al. shows regenerative abilities in 35 species along 10 different families, in which posterior and anterior growth appear to have distinct pathways. In the posterior end, neoblast cells are responsible for forming a thin layer of epithelial cells over the wound which help give rise to the regenerating body part (Xu & Sun, 2020). In the anterior end, the formation of a thin layer

of cells is accomplished by undifferentiated migratory cells, which give rise to a projecting regenerative bud and subsequently to the new structures of the head (Xu & Sun, 2020). One organism found within this phylum, *Lineus sanguineas*, demonstrates an extremely high regeneration capability (Zattara et al, 2018). However, other organisms within the same phylum demonstrate little to no regenerative ability, especially of the anterior end. This only further demonstrates the difficulty in comparative efforts to identify developmental and evolutionary causes of regeneration across the animal kingdom.

Capitella teleta

The marine worm *C. teleta* was first described as its own taxological species by Blake at al. in 2009. Previously it was known as *Capitella* sp. I., which was a provisional name given to *C. teleta* and other 5 polychaetes (*Capitella* sp. Ia, II, IIa, III, IIIa) from the *Capitella* genus (Grassle, 1980). In his studies, Grassle was able to identify that although morphologically similar, these animals had differences in their genomes and in their methods of reproduction. One experiment demonstrated that no common alleles were found in eight allozyme loci analyzed, thus proving the existence of different species within the *Capitella* genus (Grassle, 1980). They were classified as members of the superclade Spiralia, and part of the phylum Annelida. These opportunistic segmented polychaetae are well known markers of stressed and disturbed environments (Seaver, 2016) and have been studied as a model organism for the past 30 years (Giani et al., 2011). *C. teleta* was the first marine polychaeta to have its genome sequenced (Blake et al., 2009). It possess 9 thoracic segments and many more abdominal segments, along with an anterior pre-segmented head (Seaver, 2016).

C. teleta are capable of extensive posterior regeneration following amputation and also hold the ability of adding abdominal segments to existing body parts throughout their life span (Seaver, 2016). Animals in this species are morphologically distinct with male, female and hermaphrodite individuals (Giani et al., 2011). Upon fertilization and maturation of the eggs, formation of a "brooding" tube can be observed around the body of the female *C. teleta* (Seaver, 2016). After a period of 9 days, the larvae exit the brooding tube and become a swimming larvae which will then become burrowing juveniles (Seaver et al., 2005). Adulthood is achieved after 6-8 weeks post-fertilization and the mature worm can have up to 65 body segments (Seaver, 2016).

PIWI Proteins

Argonaute (AGO) proteins are an important conserved family of proteins across the Eukarya (Thomson & Lin, 2009). This family has undergone substantial gene duplication in some species and deletion in others, rendering varying number of genes depending on the organism (Hutvagner & Simard, 2008). Variation in the presence of AGO genes was analyzed in a number of different species (Hutvanger & Simard, 2008). It was found that *C. elegans* has 26 AGO genes, of which 5 are AGO-like, 3 are PIWI-like, and 18 are of the group 3 AGOs, making *C. elegans* the organism with most gene duplications (Hutvanger & Simard, 2008). In comparison, *Schizosaccharomyces pombe* has only one AGO-like gene, having lost the piwi-like protein paralogues along the course of evolutionary history (Hutvanger & Simard, 2008). AGO proteins are divided into four domains, two of which are RNA interacting proteins known as PAZ domain or AGO and the PIWI domain. The other two domains are the N-terminus and the Mid domain (Hutvagner & Simard, 2008).

The AGO subfamily of proteins bind to small-interfering RNAs (siRNA) and microRNAs (miRNAs) (Cheloufi et al., 2010) and also have a role in gene silencing, transposon silencing, and RNA stability (Hutvagner & Simard, 2008). In Hutvagner and Simard's study, human AGO1 and AGO2 complexes have been shown to interact with protein complexes that bind RNA and are responsible for RNA maturation, processing, and translation. These RNA biding complexes have also been implicated in the process of producing small, non-coding RNAs such as small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) (Hutvagner & Simard, 2008).

PIWI expression in germline tissue

Studies on PIWI proteins have demonstrated their expression in germline and somatic cells (Thomson & Lin, 2009). In the germline tissue, PIWI proteins are responsible for germ cell formation, maintenance of germline stem cells, spermatogenesis, oogenesis, and meiosis (Thomson & Lin, 2009). In the more widely studied in *Drosophila* germline, PIWI proteins have been found in association with Vasa proteins (Megosh et al., 2006), in electron rich regions called polar granules, and in comparable structures in *Xenopus*, *C. elegans*, and mammals (Thomson & Lin, 2009). These granule structures are involved in the process of gametogenesis (Thomson & Lin, 2009). PIWI proteins in *Drosophila* associate with endoribonuclease dicer (DICER1) and form a complex with the fragile X mental retardation protein (dFMRP), thus suggesting a role in miRNA synthesis and gene regulation. Moreover, PIWI proteins and their interaction with PIWI-interacting RNAs (piRNAs) are responsible for transposon silencing (Brennecke et al., 2007). The silencing pathway appears to occur through the interaction of piRNA loci in heterochromatic regions

of the genome with active transposon sites in euchromatic regions (Brennecke et al., 2007). The loss of PIWI proteins in germline cells has been demonstrated to cause infertility (Ross et al., 2014) and to increase the number of transposon mRNA population in cells (Brennecke et al., 2007).

PIWI expression in somatic tissue

In somatic cell lines, the role of PIWI proteins has been described based on its expression and function in model organisms such as *D. melanogaster, Tetrahymena thermophila* and Planarians (Thomson & Lin, 2009) and appear to play a role in epigenetic tagging of biding sites and in heterochromatin slicing through RNAi pathways (Pal-Bhadra et al., 2004). PIWI proteins also appear to be expressed in the brain and eyes of *Drosophila*, where mutations lead to position effect variegation (PEV) of the gene White, responsible for eye color (Ross et al., 2014). *D. melanogaster* offspring that lack maternal PIWI have considerable birth defects caused by mitosis and chromosome errors, thus confirming the role of this protein in embryo development. (Ross et al., 2014). PIWI proteins have also been observed to affect viability of progeny in *T. thermophila* due to loss of the DNA elimination pathway (Mochizuki et al., 2002) and are over expressed in cancer cells, which could be related to their ability to maintain stem cells in a self-renewal state (Thomson & Lin, 2009). Most of the roles of PIWI in somatic tissue, however, are linked to stem cell self-renewal and the maintenance of a pluripotent state (Perera et al., 2019). For example, in Planarians and other worms, PIWI expression is associated with neoblast cells (Ross et al., 2014). Neoblasts are populations of pluripotent stem cells in the body of Planarians such as *Dugesia japonica* and *Girardia tigrina* that can virtually become any tissue in the

worm's body that have been lost or suffered damage (Thomson & Lin, 2009). In *L. sanguineus*, RNA interference of PIWI causes loss of regenerative abilities and maintenance of the newly formed tissue (Zattara et al., 2018).

Biogenesis of piRNA

There are few studies elucidating the expression of piRNAs in somatic tissues. Until recently, it was believed that piRNAs were generated and expressed only in germline tissue. In mice, one study was capable of demonstrating the expression of piRNA in all three germ layers, although level of expression was noticeable lower than in germline tissue (Perera et al., 2019). The germline piRNAs are about 24-32 nucleotides (nt) in length with some species such as *C. elegans* having shorter, 21 nt long piRNAs (Weick & Miska, 2014). A conserved feature of most piRNAs across different clades is the strong bias for an uridine in the first position of the 5' end and for a adenosine in position 10 (Perera et al., 2019). About 20% of these RNAs can be mapped back into the genome and will align with transposon regions or their remnants, which explains their RNAi ability (Carmell et al., 2007). piRNAs undergo further processing after maturation by methylation of 2' O at the 3' end, a process that is accomplished by a piRNA methyltransferase known as Hen1 (Perera et al., 2019).

In its genesis in germline tissue, piRNAs are believed to be generated in dual-stranded clusters (both strands of DNA encode for piRNAs) mostly through what it is called the ping-pong cycle or secondary piRNAs biogenesis (illustrated in Figure 2). For transcription of piRNAs proteins such as HP1 homolog, Rhino (Rhi) localize at piRNAs clusters and bind to the H3K9me3 (Weick & Miska, 2014). The binding causes recruitment

of Deadlock and Cutoff, a transcription terminator co-factor (illustrated in Figure 1) (Weick & Miska, 2014). Caping of 5' end by Cutoff prevents piRNA degradation, prevents termination of transcription and splicing of the precursor before time (Weick & Miska, 2014). The protein UAP56 (Hel25E) along with Rhi, appears to transport the precursor piRNA out of the nuclear pore and into the cytoplasm for maturation (illustrated in Figure 1) (Weick & Miska, 2014).

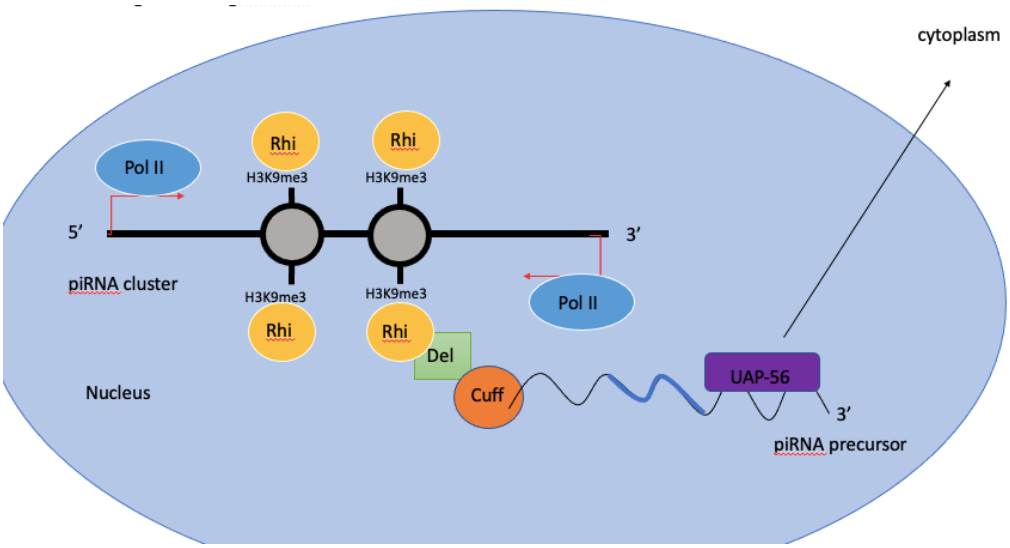


Figure 1:Transcription of piRNA in Germiline. The diagram shows the interactions of Rhi (Rhino) and Del (Deadlock) and Cuff (Cutoff) proteins with histone H3K9me3. The piRNA is a precursor RNA that will undergo further maturation in the cytoplasm

In a previous study, Brennecke et al. aligned piRNAs sequences from *D. melanogaster* to its genome's transposable regions and observed that PIWI and AUB protein complexes appear to have bias for transposons' antisense strand while AGO3 complexes have strong strand bias towards the sense strand of transposons (Figure 1). For about 20% of all piRNAs, mapping of piRNAs into the genome demonstrated a distance of 10 nucleotides between the overhanging 5' end of one piRNA and the next RNA on the

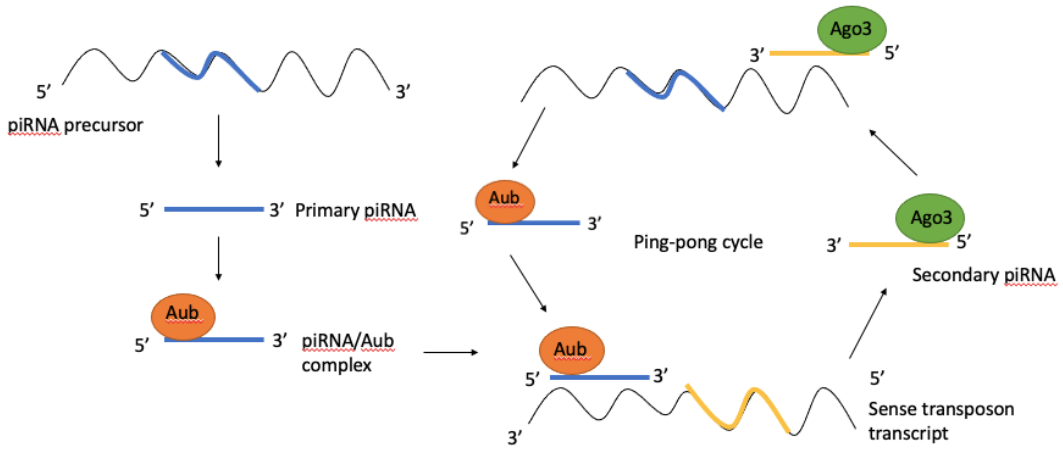


Figure 2:piRNA biogenesis in germline tissue of Drosophila. The diagram shows the interaction of Ago3 and Aub proteins in associations with small RNA sequences and piRNA sequences respectively.

opposite strand (Brennecke et al., 2007). In the case of complementarity of piRNAs for AUB and AGO3, this percentage increases to 48% of the cases. It is also important to note that PIWI and AGO3 interactions have also been found, though to a lesser extent than AUB/AGO3 interactions (Brennecke et al., 2007). It is proposed that piRNAs bound to either PIWI or AUB complexes recognize the sense strand of active transposons and cleave it (Ross et al., 2014). The cleavage occurs on the opposite strand, about 10 nucleotides away from the antisense piRNA site (Brennecke et al., 2007). The new piRNA is then processed and its 3' end cleaved. Due to this strand bias (illustrated in Figure 1), the newly formed piRNA would most likely be incorporated into an AGO3 complex that will recognize and bind to a sense strand on a transposon region to generate an antisense piRNA (Brennecke et al., 2007). The ping pong cycle is found not only in *D. melanogaster* but also in zebrafish and in mice germ cells (Aravin et al., 2007; Houwing et al., 2007).

The primary biogenesis of piRNAs in somatic cells is based on a study conducted on the ovarian follicles of *D. melanogaster* (Malone et al., 2009) and has been shown to

originate from piRNA clusters such as *flamenco* and *traffic jam*, and from the 5' UTR of mRNAs (Ross et al., 2014). The pre-piRNAs are transcribed by Polymerase II as long piRNA reads from uni-stranded clusters (piRNAs are encoded on only one strand) (Weick & Miska, 2014). Once exported to the cytoplasm, the precursor is cleaved by an endonuclease, thus generating its 5' ends and further processed to generate the piRNA/Piwi complex (Figure 3) which will then be localized into the nucleus where it will serve mainly as a transposon repressor, avoiding induced activation of transposon regions in neighboring germ cells in the Ovarian (Ross et al., 2014).

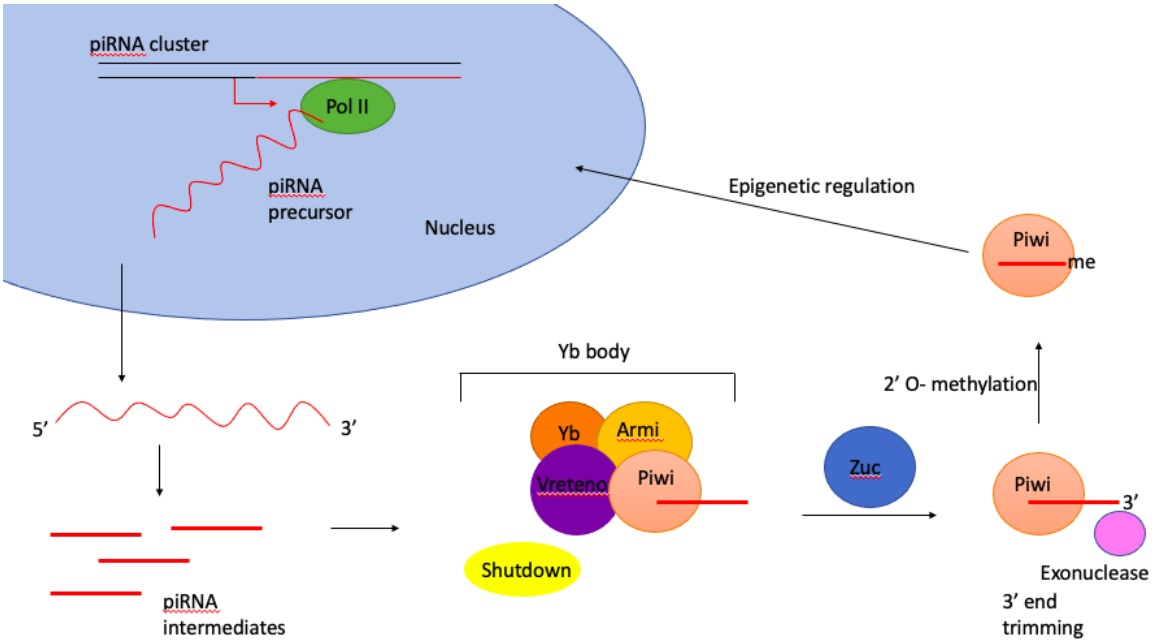


Figure 1: Biogenesis of piRNA in somatic tissue of D. melanogaster. The precursor is exported to the cytoplasm where it's it is cleaved by an endonuclease. The piRNAs are coupled to the Yb body with the help of Shutdown chaperone protein. The piRNA is further processed with the help of Zucchini where it will be trimmed and methylated.

In contrast to *D. melanogaster*, mice piRNA biogenesis occur through a slightly different mechanism. In addition, mutations in MIWI, MILI and MIWI2 cause infertility only in males (Weick & Miska, 2014). For example, the generation of piRNAs that interact with MIWI and MILI proteins, homologs of PIWI proteins in mice, appear to be under the

influence of a master transcriptional regulator A-MYB (MYBL1) (Weick & Miska, 2014). RNA Polymerase II is induced by A-MYB, which triggers the concerted transcription of long piRNA precursors (100 nucleotides or more) and of mRNAs of MIWI and MITOLPD. The latter is the mouse homolog of the *D. melanogaster's* mitochondrial-associated Zucchini protein (Figure 3) (Weick & Miska, 2014). This shows that although some features of the biogenesis of piRNAs might vary from species to species, other features remain conserved. This is yet another indication of the evolutionary linkage between PIWI proteins, piRNAs and the functions they serve in cells.

Transposable Elements

Transposons are short mobile DNA sequences capable of excision, insertion, and multiplication within an organism's genome (Bao et al., 2009). They are sometimes referred to as genetic parasites (Kapitonov & Jurka, 2010), and usually fall under two categories: DNA transposons and retrotransposons (Bao et al., 2009). Among the DNA transposons, two distinct classes are characterized based on their transposition mechanism (Han et al., 2015). Class 1 transposons utilize a copy-and-paste mechanism while class 2 utilize a cut-and-paste mechanism (Han et al., 2015). In retrotransposons, a copy-and-paste mechanism is also used but the intermediate is not DNA but rather RNA (Bao et al., 2009). The RNA intermediate is then retrotranscribed into DNA and reinserted into the genome (Bao et al., 2009). Some important categories of retroelements are the SINEs, LINEs, and LTRs. In most cases, transposons are deactivated by mutations through selective pressure as means to reduce damage to the host genome (Munoz-Lopez & Garcia-Perez, 2010). It has been shown that transposable elements constitute large portions of some organism's

genomes, increasing genome complexity and serving as an indicator of genomic evolution (Han et al., 2015).

PIWI Expression in *C. teleta*

In *C. teleta*, two homologues of PIWI proteins are found named *C. teleta*-Piwi1 (Ct-Piwi1) and *C. teleta*-Piwi2 (Ct-Piwi2) (Giani et al., 2011). These two proteins are expressed throughout the life cycle of an individual worm in both germline and somatic cells (Giani et al., 2011). In contrast, mice have differently expressed PIWI homologs depending on the developmental stage of the embryo (Weick & Miska, 2014). Humans maintain human piwi homolog (Hiwi) expression in some somatic stem cell subcultures, such as hematopoietic stem cells, but not in differentiated cells (Juliano et al., 2011). A similar feature to *C. teleta* which maintain Ct-Piwi1 and Ct-Piwi2 expression in somatic tissue (Giani et al., 2011). As the embryo of *C. teleta* grows and moves from larvae to adult stages, Ct-Piwi1 and Ct-Piwi2 become gradually more restricted to multipotent progenitor cells (MPCs) and to the posterior growth zone (PGZ), where regeneration takes place (Giani et al., 2011). The homolog Ct-Piwi1 is found to be expressed not only in the PGZ but also in the gonads of sexually mature adults, which functions as a transposon silencer as described above for other PIWI homologs (Giani et al., 2011). After the advancement of molecular tagging techniques in the past few years, it is now believed that the niche of cells in the MCPs and PGZ are not pluripotent stem cells, but more likely to be a heterogeneous group, composed of subsets of precursor somatic and germ cells that are differentiated into the required pieces of the regenerating tissue (de Jong & Seaver, 2018). This theory comes from the observation that only a small subset portion of the MCP

expresses the cell marker *Ct-myc* (Dannenberg & Seaver, 2018). In addition to the MCP and PGZ markers, the stem cell markers *nanos* and *vasa* are also present (Seaver, 2016). The expression of these genes is posited to be related to the maintenance of an undifferentiated state of MCP cells and of cells in the PGZ during regeneration (Dannenberg & Seaver, 2018). In a recent study, the origin of *C. teleta* blastema cells were elucidated. By using a technique termed EdU pulse-amputate-wait-BrdU pulse, researchers were able to show in living tissue that cells in the blastema are descendant from pre-existing segments (de Jong & Seaver, 2018). They migrate into the site of injury and begin the process of differentiation following environmental cues and cell-to-cell signaling, after 3 days post amputation, proliferating cells can be observed in all germ layers (de Jong & Seaver, 2018).

RNA sequencing

RNA sequencing (RNAseq) is a high-throughput sequencing technique developed over a decade ago that employs similar techniques to DNA sequencing. The development of this technology facilitates a better understanding and characterization of transcriptomes in cells, levels of RNA expression, post-transcriptional modification, and evolutionary ontology of all species of RNA transcripts (Wang et al., 2010). Even though some techniques have changed throughout the years, the basis of RNAseq remains the same. The process begins with RNA extraction; further purification protocols can be followed to specify a subset of RNAs such as mRNAs or small RNAs. Then, a complementary DNA (cDNA) library is made through PCR and ligated to adaptor sequences. The prepared

cDNA library is then deep-sequenced, generating 10 – 30 million reads per sample on a sequencing platform such as Illumina (Stark et al., 2019).

Illumina sequencing is a highly modernized version of the first generation of genome sequencing, referred to as Sanger sequencing. Sanger sequencing relies on the use of di-deoxynucleotides (ddNTPs) that, once incorporated into the growing RNA chain, stop elongation due to the lack of a hydroxyl group at the 3' end (Heather & Chain, 2016). Fluorescently labeled sequence results are subsequently analyzed by capillary gel electrophoresis (Stark et al., 2019). Sanger sequencing is still commonly used for short DNA/RNA reads, but is limited in its ability to determine the sequence of reads longer than 1kb. To overcome this limitation, Shotgun sequencing allows the assembly of shorter DNA reads containing overlaps with one another to be assembled into one, long sequence of DNA (Stark et al., 2019).

The leader of next generation sequencing is the Illumina technology. It relies on both transcript amplification trough PCR and fluorescent labeled nucleotides to generate the DNA sequence (Figure 4). The data collected from the sequencing is the analyzed through computation. As part of the process to understand what each read in the flow cell means, the sequence is normalized, aligned and assembled to form transcriptomes or even whole genomes depending on the end goal. With the collected information, one can identify changes in gene expression or transcript expression and compare expression of target genes through different developmental stages to mention a few examples (Stark et al., 2019).

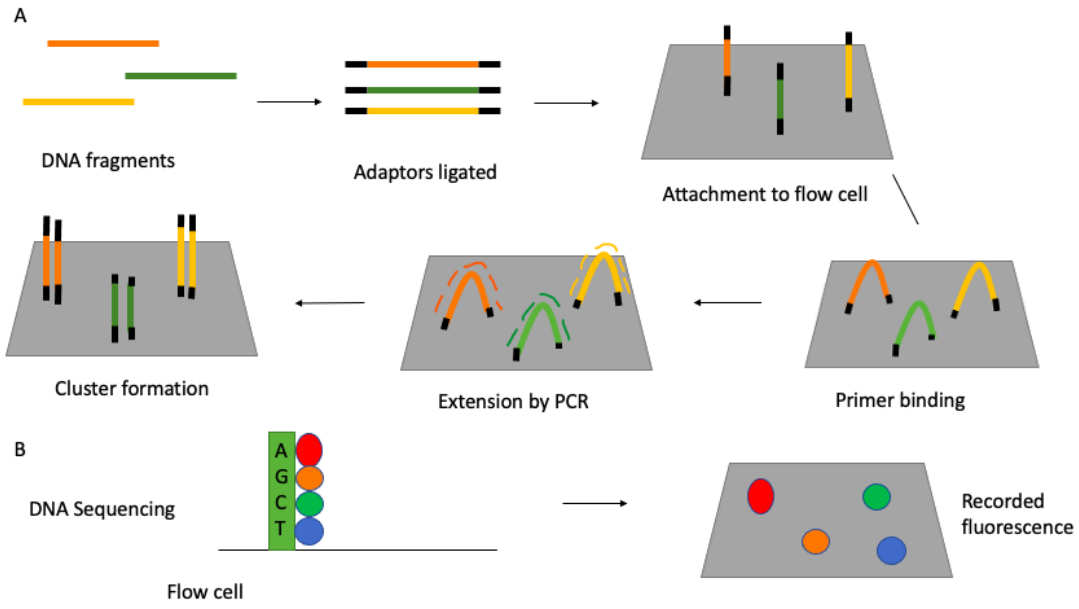


Figure 4: Illumina Sequencing Schematic. The library is prepared by ligating adaptors to the end of the sequences, the adaptors are complementary to the oligonucleotides present in the glass flow cell. Through several rounds of PCR the library is copied and amplified. The system utilizes reversible terminators and modified polymerases to extend the growing chains. With the addition of each new base by the polymerase, a different fluorescent color is emitted based on the profile of the base added and the DNA sequence is annotated.

Illumina sequence allows for thousands of oligonucleotides of 500bp or more to be sequenced simultaneously. Each Illumina sequencing has 300 or more rounds generating a deep-sequencing library with the number of reads needed for proper genome or transcriptome analysis (Slatko et al., 2018).

Gene Ontology Analysis

Gene Ontology (GO) databases started to arise with the advances of high-throughput sequencing technologies. It is a conjoint effort to annotate genes and proteins of eukaryotic organism (Finke et al., 2004). The libraires are divided in three domains: cellular component, biological process, and molecular function (Grossmann et al., 2007). The output of the GO analysis is a directed acyclic graph (DAG) that starts off with one of

the general domain process as the parent bubble and each subsequent daughter bubbles becomes a more specific term for the function of the input DNA sequences (Grossmann et al., 2007). DAGs provide a visual representation of the characteristics of the genes or proteins of interest with increasing significance based on *p*-values. Through GO analysis, researchers have been able to show that eukaryotes possess conserved core biological functions based on data input into the annotated library (Ashburner et al., 2000).

CHAPTER III: MATERIAL AND METHODS

C. teleta animal model

C. teleta used in this study were acquired from Dr. Elaine Seaver at the Whitney Laboratory for Marine Bioscience (University of Florida). Male and female worms were grown together in 500 mL plastic lid containers with mud collected from the Mississippi Gulf coast and sea water. A 0.5" of mud and 1-1.5" of FSW were added to each container. The worms were separated in different containers by life stage and collection date: embryos, larvae, juvenile, adults, and old. They were fed once a week by adding a teaspoon of new mud to the containers. Containers were stored at 20˚C until needed for experiments. Adult containers were examined frequently for the formation of brooding tubes or the presence of embryos. Once found, the brooding tubes and the embryos were transferred to new containers with mud and FSW, and the worms were allowed to grow.

Amputations

Amputations were carried out only in adult species of *C. teleta* based on size and the presence of mature reproductive organs. The worms were selected from containers through examination under a microscope. The specimens were amputated using forceps at 12^{th} body segments. Newly amputated worms were then placed in petri dishes containing 20mL of filtered fresh sea water (FSW) and stored at 20˚C. After 24-48 hrs, the amputated *C. teleta* were transferred to regular containers with mud and FSW. The worms were screened for growth every other day and in case of death, the worms were removed from the container and discarded. The surviving worms were allowed to regenerate for a period of 2-3 weeks prior to RNA extraction.

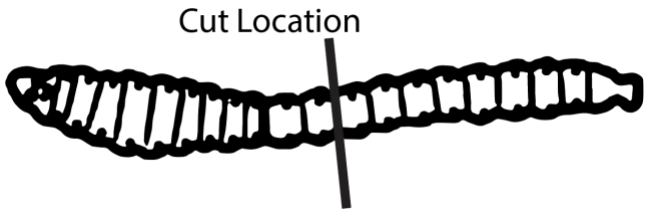


Figure 5: Transverse amputation of C. teleta. The amputations were performed using forceps to stabilize the worm and cut the tissue at the 12th body segment.

RNA extraction

Total RNA extractions using a ThermoFisher TRIzol™ LS Reagent kit according to the manufacture's protocol. For combined male and female data, tissue from three worms of each sex were used for RNA extraction. For the regeneration data, small RNA of regenerating tissue from 5 female worms were used. Briefly, the worms were removed from the containers and placed in a petri dish with filtered-FSW for approximately 24 hrs prior to RNA extraction to remove body waste from their system. The specimens were then placed under the microscope and the regenerating portions of their body segments were amputated using forceps. The body parts were placed in an Eppendorf tube containing 0.25 mL of TRIzol reagent and a tissue grinder was used to homogenize the mixture prior to extracting RNA. Small RNAs were extracted from total RNA using the mirVana™ miRNA Isolation Kit according to the manufacturer's protocol (procedure IV.A). RNA libraries were then sent to be sequenced utilizing a Illumina Next Seq500.

Computational Data Analysis

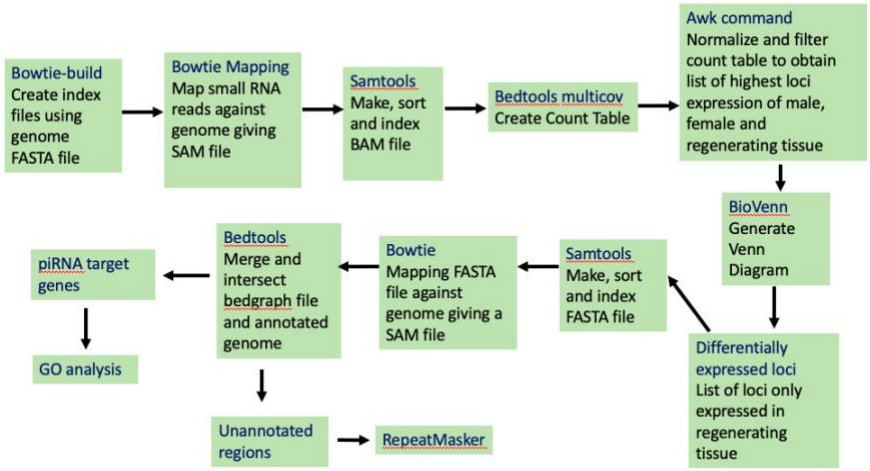


Figure 6: Flow Chart of Bioinformatics for analysis of small RNA in regenerating tissue of C. teleta. The chart demonstrates each step in the computational analysis data. The resulting files from sequencing were indexed and annotated using the programs such as Bowtie, Samtools, Bedtools, Awk and public databases for each of the three samples (male, female and regeneration). From the annotated files, a list of differentially expressed loci were obtained and used to compare and contrast their expression among control (male and female) and regenerating tissue. Uniquely expressed loci in regenerating tissue were further analyzed with RepeatMasker and GO analysis.

CHAPTER IV: RESULTS

274 loci were differently expressed in regenerating tissue

Small RNA sequencing results were combined to generate three *C. teleta* samples: combined male body, combined female body, and regenerating tissue. The combined male body was composed of three male samples. The combined female body was composed of three female samples. The regeneration tissue was acquired from the posterior end of 5 female worms. These samples were then sorted to create three distinct files with the top 1000 highest expressed loci in each of the tissues (Figure 6, Appendix A). The loci names were imported into the BioVenn generator website (Figure 7). The diagram shows similarity for most of the highly expressed loci as observed by the large overlapping circles in the middle of the diagram. Regenerating tissue overlaps almost identically with both male and female samples, which can indicate conserved expression of small RNAs across *C. teleta* species that regulate cellular function. Male and female samples also have uniquely expressed loci (Figure 7) that could be derived from specific germline small RNAs that are differently expressed in each of the sexes. However, regenerating tissue was collected from the most posterior end of the animal, therefore it does not contain traces of germline cells and does do not overlap with neither male and female in that category. From the BioVenn application, a list of 274 differently expressed loci were obtained (Appendix A) and used to identify possible piRNAs targeting specific genes during the regeneration process.

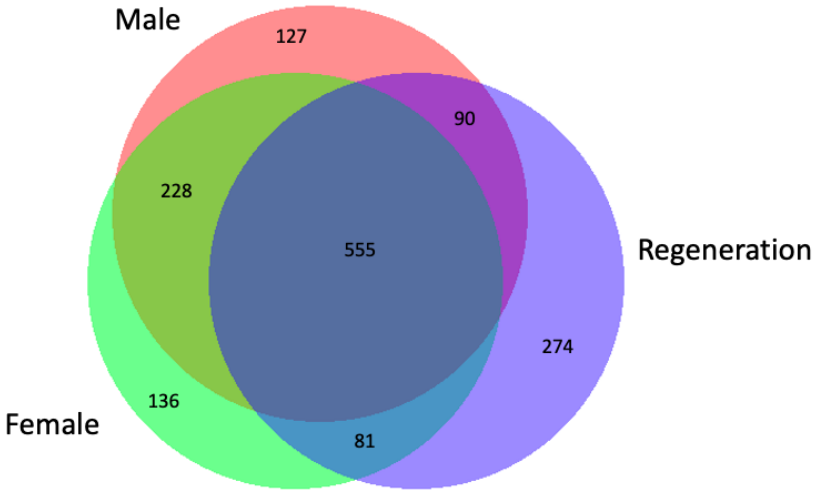


Figure 7: BioVenn Diagram of Highly Expressed Loci. The figure shows overlapping and unique loci for three samples under analysis: combined male (pink), combined female (green) and regenerating tissue (purple) of C. teleta. Numbers represent the number of loci in each of the intersections and of uniquely expressed in each of the tissues.

piRNAs targets are involved in apoptotic signaling

GO analysis of the differently expressed genes targeted (Appendix A) by piRNAs indicated that piRNAs appear to be strongly involved in silencing genes involved in the regulation of apoptosis, programmed cell death, and cell death elements (Figure 8). We also observe possible interactions between piRNAs and the cellular component chromatin in the nucleus during the regeneration process (Figure 9). piRNA activity in the nucleus and interaction with DNA is further supported by the interactions with DNA binding genes (Figure 10). By analyzing the flow chart (Figure 10) step by step, one can observe that piRNAs are interacting with elements that bind organic and heterocyclic compounds such

as nucleotide sequences, which then interact with elements binding nucleic acid and DNA sequences. The piRNAs also appear to interact strongly with intracellular organelles and their lumen, which could simply mean the nucleus of the cell (Figure 9). Overall, piRNAs appear to be localized and active in the nuclei of regenerating tissue, interacting with chromatin to suppress the expression of apoptotic genes. The genes were annotated using public data base and statistics were run using TopGO which runs a fisher exact test for p-value. The p-value is used to assess the significance of a term within that group of genes. In each of the charts (Figure 8, 9 and 10), the p-value, under the description of the function, decreases as the terms become more specific, asserting the significance and the precision of the cellular function or structure being described. The arrows follows a pattern in which each subsequent bubble becomes a more precise function for the set of genes under investigation.

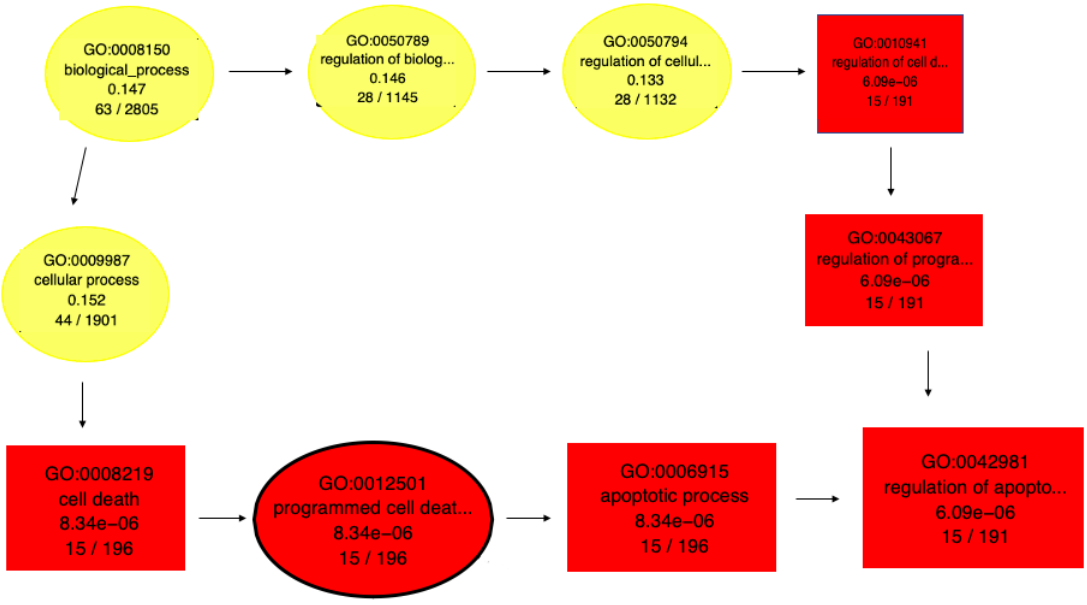


Figure 8: GO analysis using R studio of piRNA target genes during the regeneration process: Biological Process. This flow chart demonstrates GO elements interactions. The color is based on p-values, with warmer shades expressing higher significance (red) and lighter shades expressing lower significance (yellow). p-value < 0.05. Each bubble contains a GO term that is related to the GO database and indicates a cellular function or structure, accompanied by a description of the GO term.

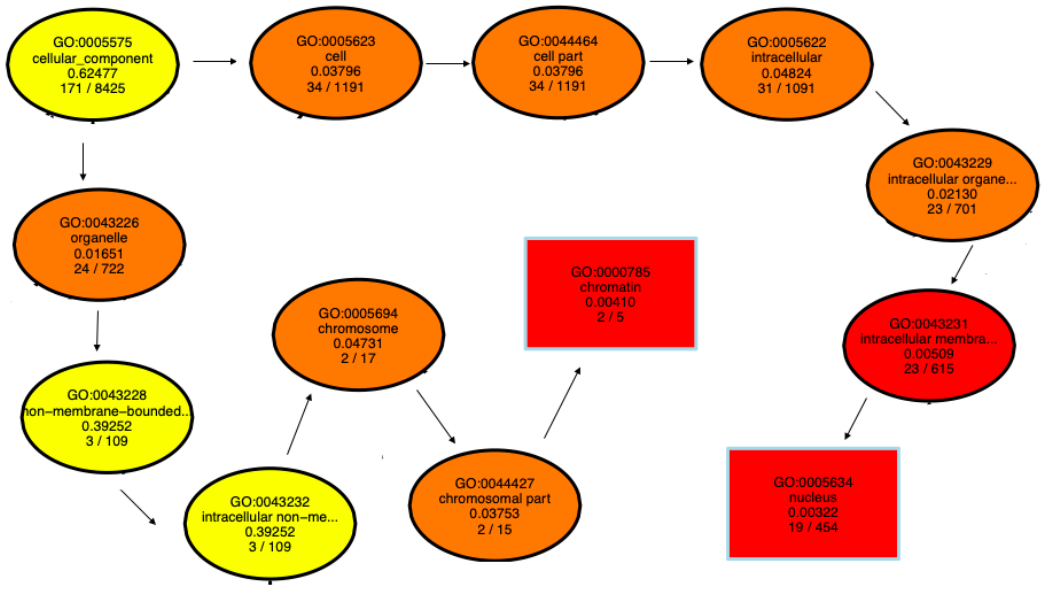


Figure 9:GO analysis using R studio of piRNA target genes during the regeneration process: Cellular component. This flow chart demonstrates GO elements interactions. The color is based on p-values, with warmer shades expressing higher significance (red) and lighter shades expressing lower significance (yellow). p-value < 0.05. Each bubble contains a GO term that is related to the GO database and indicates a cellular function or structure, accompanied by a description of the GO term.

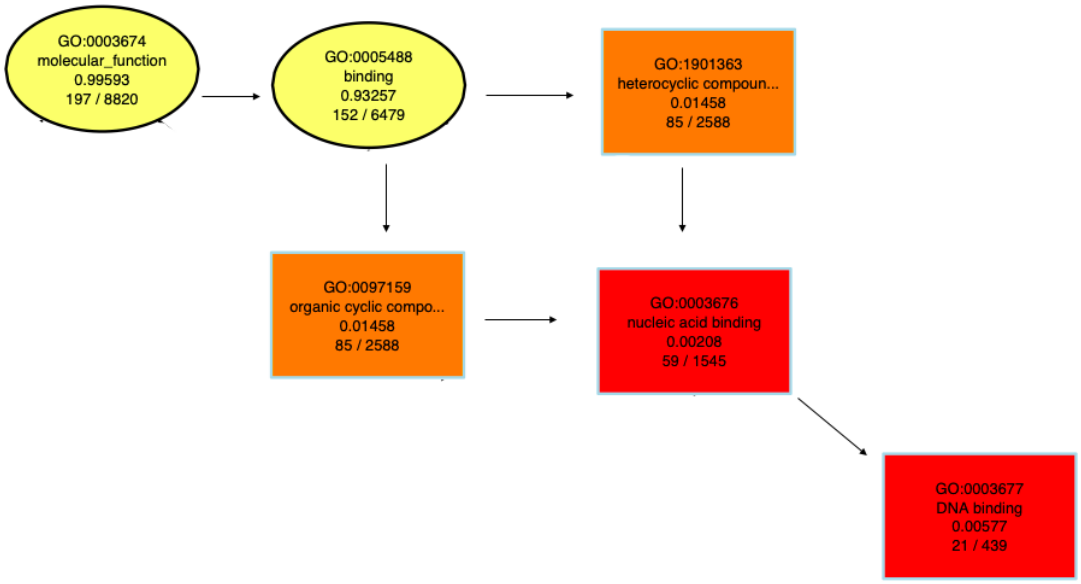


Figure 10: GO analysis using R studio of piRNA target genes during the regeneration process: molecular function. p-value < 0.05. Flow chart demonstrate GO elements interactions. Color is based on p-value with warmer shaded expressing higher significance (red) and lighter shades expressing lower significance (yellow). Each bubble contains a GO term that is related to the GO database and indicates a cellular function or structure, accompanied by a description of the GO term.

piRNAs also target transposable elements

Mapping of the differently expressed loci (Appendix A) using RepeatMasker demonstrated that piRNAs are targeting transposable elements in the genome during regeneration. For that reason, the strong interaction between piRNAs, the nucleus, and some chromosomal elements, more specifically of chromatin, is not abnormal. After mapping unannotated sequences onto the RepeatMasker library, we found 9 retroelements:

- 2 SINES with tRNA fragment heads, thought to be originated from $tRNA^{Lys}$, (Nishihara et al., 2006) a core of the conserved Deu-domain superfamily of SINE

and fragments of CR1 LINE element (Nishihara et al., 2006; Vassetzky & Kramerov, 2013)

- 6 LINES retrotransposon elements, of which 5 elements aligned with L2/CR1/Rex and 1 aligned with R1/L0A/Jockey
- 1 Gypsy/DIRS1 LTR element
- 5 DNA transposons, of which two were from the PHIS transposon superfamily, PIF-ISL2EU (Han et al., 2015), two transposable elements from the Kolobok superfamily present in two distinct loci (Kapitonov & Jurka, 2010) and one Sola1 transposon from the Sola superfamily (Bao et al., 2009).

Most of the repeats in the unannotated file were simple repeats or repeats of low complexity that did not characterized an individual element. The data also show novel targeting sites of piRNAs during regeneration such as the transposon sequences shown above. As elucidated in the Literature Review, piRNAs target active transposon sites, thus one may infer that here, piRNAs are silencing the transposable activity of such elements during the regenerative process to avoid mutation and further damage to the growing tissue.

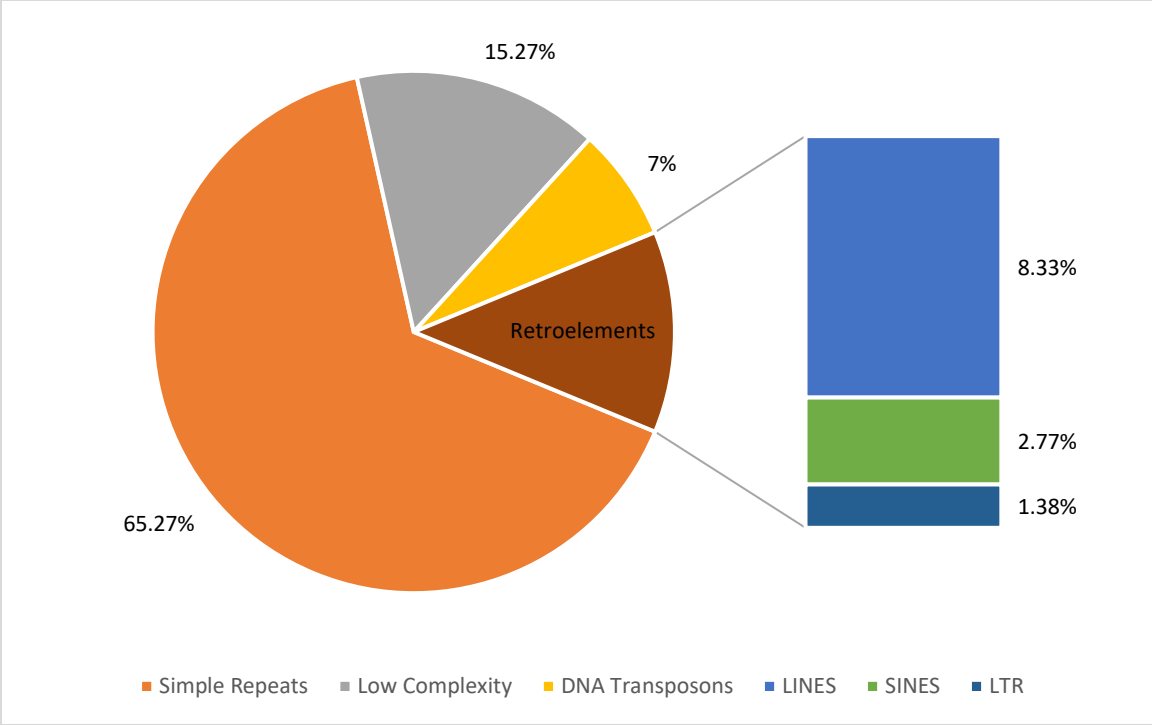


Figure 7. Transposon Alignments Targeted by piRNAs in Regenerating Tissue of C. teleta. The pie graph shows the percentage of elements target by piRNA that were not genes in differently expressed loci. Retroelements are further divided into long interspersed retrotransposable elements, (LINES,) short interspersed retrotransposable elements, (SINES) and long terminal repeats (LTRs).

CHAPTER V: DISCUSSION

The presence and expression of piRNAs in *C. teleta* has been hitherto poorly characterized. The experiments conducted here describe piRNA expression in the regenerating tissue of a marine annelid. Furthermore, analysis of the genome regions targeted by these small RNAs, which have been previously unannotated, demonstrates interactions with both types of transposable elements: retrotransposons and DNA transposons. Regarding retrotransposons, the SINE element (SINE3_IP) from the recently discovered Due-domain superfamily was targeted. These superfamily elements share a common core and are found across species of vertebrate, amphioxus, sea urchins, and other relatives (Nishihara et al., 2006). This shared sequence as the central domain of the SINE may indicate a single evolutionary ancestor prior to the phyla split.

The piRNAs were also shown to interact and target the gypsy retrotransposon. In *D. melanogaster,* gypsy is active in the follicles of the ovaries, if the female possess the *flamenco* gene. Activation of gypsy and its interaction with *flamenco,* produces retroviral particles and leads to subsequent infection of oocytes generating mutant offspring (Tsai et al., 1997). The ability to infect other neighboring cells makes gypsy a target for transposon silencing as seen with the piRNA repression of the retrotransposon in the *C. teleta* tissue. Much like *D. melanogaster* represses expression of transposons in germline cells to preserve the integrity of the offspring genome, *C. teleta* appears to strongly repress this transposon expression during regeneration events to preserve the integrity of the genome of the newly synthesized body segment. Repression of the other transposable elements such as LINES, LTR, and DNA transposons also appears to follow this logic, as they are

strongly targeted in regenerating tissue but not in the non-damaged somatic tissue of both male and female samples.

The GO analysis of target regions by piRNA show three distinct categories of interaction. In the biological processes, we can observe strong interactions between piRNAs and the cellular process of cell apoptosis, programmed cell death, and cell death elements. piRNAs are known, as described above, to repress and silence genomic regions, thus one can make the assumption that in the process of regeneration, piRNAs are actively suppressing the natural cellular response to extensive cell damage, programmed cell death, or apoptosis. The discovery of piRNAs targeting such biological process helps elucidate the mechanisms behind the regenerative mechanism, since extensive cell death would be incompatible with resynthesizing the lost tissue. The DAG for cellular components and molecular function are cohesive with the findings of the RepeatMasker run and the biological processes interacting with piRNAs. Figures 5 and 6 show significant interactions based on *p*-value ($p < 0.05$), between piRNAs, chromatin, and DNA structure and genes that bind nucleic acids and more specifically DNA.

The interactions shown by the data presented in this study support the characterization of piRNA expression in the nucleus of regenerating tissue of *C. teleta* as a silencer of a diverse array of genomic components such as genes and transposable elements. The interaction of piRNAs and these genomic loci are uniquely expressed in the tissue of the regenerating worm (Figure 7). Not only are piRNAs differentially expressed, they are also abundant, showing that the cell is making a strong effort to suppress the expression of these elements. The data uncovered here supports the theory that piRNAs play an important role in the regeneration process in *C. teleta* and possibly in other

organisms as well. However, further experiments are needed to elucidate the exact mechanism behind the silencing of cell apoptosis and if there are more transposable elements also silenced by piRNAs.

APPENDIX A: DATA FROM GENOMIC LOCI EXPRESSION

Data gathered from the small RNA sequencing was computed utilizing the system at display in figure 2. Below are the data sets from the computational results of the top 1000 highly expressed loci in male, female and regenerating tissue of *C. teleta* with the expression level demonstrated to the right of the Genomic Loci (1). The 274 differentially expressed loci in regenerating tissue obtained from the BioVenn website diagram is also listed (2). The data was uploaded to an ePortfolio on Canvas.

1. Honors Thesis Primary Data

REFERENCES

Ahmadian, A., Ehn, M., & Hober, S. (2006). Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta*, *363*(1), 83–94. https://doi.org/https://doi.org/10.1016/j.cccn.2005.04.038

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, *316*(5825), 744–747. https://doi.org/10.1126/science.1142612

Bao, W., Jurka, M. G., Kapitonov, V. V., & Jurka, J. (2009). New superfamilies of eukaryotic DNA tyransposons and their internal divisions. *Molecular Biology and Evolution*, *26*(5), 983–993. https://doi.org/10.1093/molbev/msp013

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/10.1016/j.cell.2007.01.043

Carmell, M. A., Girard, A., van de Kant, H. J. G., Bourc'his, D., Bestor, T. H., de Rooij, D. G., & Hannon, G. J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell*, *12*(4), 503–514. https://doi.org/10.1016/j.devcel.2007.03.001

Cheloufi, S., Dos Santos, C. O., Chong, M. M. W., & Hannon, G. J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, *465*(7298), 584–589. https://doi.org/10.1038/nature09092

Dannenberg, L. C., & Seaver, E. C. (2018). Regeneration of the germline in the annelid Capitella teleta. *Developmental Biology*, *440*(2), 74–87.

https://doi.org/10.1016/j.ydbio.2018.05.004

de Jong, D. M., & Seaver, E. C. (2018). Investigation into the cellular origins of posterior regeneration in the annelid Capitella teleta . *Regeneration*, *5*(1), 61–77. https://doi.org/10.1002/reg2.94

Finke, W., Rachimow, C., & Pfützner, B. (2004). Untersuchungen zu Wasserdargebot und Wasserverfügbarkeit im Ballungsraum Berlin im Rahmen des Verbundprojektes GLOWA Elbe. *Hydrologie Und Wasserbewirtschaftung*, *48*(1), 2–11. https://doi.org/10.1038/75556.Gene

Giani, V. C., Yamaguchi, E., Boyle, M. J., & Seaver, E. C. (2011). Somatic and germline expression of piwi during development and regeneration in the marine polychaete annelid Capitella teleta. *EvoDevo*, *2*(1), 10. https://doi.org/10.1186/2041-9139-2-10

Grassle, J. (1980). *Polychaete Sibling Species BT - Aquatic Oligochaete Biology* (R. O. Brinkhurst & D. G. Cook (eds.); pp. 25–32). Springer US. https://doi.org/10.1007/978-1-4613-3048-6_3

Grossmann, S., Bauer, S., Robinson, P. N., & Vingron, M. (2007). Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis. *Bioinformatics*, *23*(22), 3024–3031. https://doi.org/10.1093/bioinformatics/btm440

Han, M. J., Xiong, C. L., Zhang, H. B., Zhang, M. Q., Zhang, H. H., & Zhang, Z. (2015). The diversification of PHIS transposon superfamily in eukaryotes. *Mobile DNA*, *6*(1), 1–11. https://doi.org/10.1186/s13100-015-0043-7

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H. A., Hannon, G. J., Draper, B. W., & Ketting, R. F. (2007). A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell*, *129*(1), 69–82. https://doi.org/10.1016/j.cell.2007.03.026

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, *9*(1), 22–32. https://doi.org/10.1038/nrm2321

Juliano, C., Wang, J., & Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annual Review of Genetics*, *45*, 447–469. https://doi.org/10.1146/annurev-genet-110410-132541

Kapitonov, V., & Jurka, J. (2010). Kolobok, a novel superfamily of eukaryotic DNA transposons. *Repbase Rep*, *7*, 111–122.

Labbé, R. M., Irimia, M., Currie, K. W., Lin, A., Zhu, S. J., Brown, D. D. R., Ross, E. J., Voisin, V., Bader, G. D., Blencowe, B. J., & Pearson, B. J. (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells (Dayton, Ohio)*, *30*(8), 1734—1745. https://doi.org/10.1002/stem.1144

Liu, X., Sun, Y., Guo, J., Ma, H., Li, J., Dong, B., Jin, G., Zhang, J., Wu, J., Meng, L., & Shou, C. (2006). Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. *International Journal of Cancer*, *118*(8), 1922–1929. https://doi.org/10.1002/ijc.21575

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R.,

& Hannon, G. J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell*, *137*(3), 522–535. https://doi.org/10.1016/j.cell.2009.03.040

Megosh, H. B., Cox, D. N., Campbell, C., & Lin, H. (2006). The Role of PIWI and the miRNA Machinery in Drosophila Germline Determination. *Current Biology*, *16*(19), 1884–1894. https://doi.org/10.1016/j.cub.2006.08.051

Mochizuki, K., Fine, N. A., Fujisawa, T., & Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. *Cell*, *110*(6), 689–699. https://doi.org/10.1016/S0092-8674(02)00909-1

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/10.1016/j.cell.2014.04.031

Munoz-Lopez, M., & Garcia-Perez, J. (2010). DNA Transposons: Nature and Applications in Genomics. *Current Genomics*, *11*(2), 115–128. https://doi.org/10.2174/138920210790886871

Nishihara, H., Smit, A. F. A., & Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Research*, *16*(7), 864–874. https://doi.org/10.1101/gr.5255506

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A., & Elgin, S. C. R. (2004). Heterochromatic Silencing and HP1 Localization in Drosophila Are Dependent on the RNAi Machinery. *Science*, *303*(5658), 669 LP – 672. https://doi.org/10.1126/science.1092653

Palakodeti, D., Smielewska, M., Lu, Y. C., Yeo, G. W., & Graveley, B. R. (2008). The

PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *Rna*, *14*(6), 1174–1186. https://doi.org/10.1261/rna.1085008

Perera, B. P. U., Tsai, Z. T.-Y., Colwell, M. L., Jones, T. R., Goodrich, J. M., Wang, K., Sartor, M. A., Faulk, C., & Dolinoy, D. C. (2019). Somatic expression of piRNA and associated machinery in the mouse identifies short, tissue-specific piRNA. *Epigenetics*, *14*(5), 504–521. https://doi.org/10.1080/15592294.2019.1600389

Reddien, P. W., & Alvarado, A. S. (2004). FUNDAMENTALS OF PLANARIAN REGENERATION. *Annual Review of Cell and Developmental Biology*, *20*(1), 725–757. https://doi.org/10.1146/annurev.cellbio.20.010403.095114

Ross, R. J., Weiner, M. M., & Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*, *505*(7483), 353–359. https://doi.org/10.1038/nature12987

Saló, E., & Baguñà, J. (2002). Regeneration in planarians and other worms: New findings, new tools, and new perspectives. *Journal of Experimental Zoology*, *292*(6), 528–539. https://doi.org/10.1002/jez.90001

Seaver, E. C. (2016). Annelid models I: Capitella teleta. *Current Opinion in Genetics and Development*, *39*, 35–41. https://doi.org/10.1016/j.gde.2016.05.025

Seaver, E. C., Thamm, K., & Hill, S. D. (2005). Growth patterns during segmentation in the two polychaete annelids, Capitella sp. I and Hydroides elegans: Comparisons at distinct life history stages. *Evolution and Development*, *7*(4), 312–326. https://doi.org/10.1111/j.1525-142X.2005.05037.x

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing technologies (and bioinformatics) in cancer. *Molecular Biology*, *122*(1),

1–15. https://doi.org/10.1002/cpmb.59.Overview

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Thomson, T., & Lin, H. (2009). and Prospect. *Cell*, *2006*, 355–376. https://doi.org/10.1146/annurev.cellbio.24.110707.175327.The

Tsai, S.-F., Jang, C.-C., Prikhod'ko, G. G., Bessarab, D. A., Tang, C.-Y., Pflugfelder, G. O., & Sun, Y. H. (1997). Gypsy retrotransposon as a tool for the in vivo analysis of the regulatory region of the optomotor-blind gene in Drosophila. *Proceedings of the National Academy of Sciences*, *94*(8), 3837 LP – 3841. https://doi.org/10.1073/pnas.94.8.3837

Vassetzky, N. S., & Kramerov, D. A. (2013). SINEBase: a database and tool for SINE analysis. *Nucleic Acids Research*, *41*(D1), D83–D89. https://doi.org/10.1093/nar/gks1263

Wang, Z., Gerstein, M., & Snyder, M. (2010). *RNA-Seq: a revolutionary tool for transcriptomics*. *10*(1), 57–63. https://doi.org/10.1038/nrg2484.RNA-Seq

Weick, E.-M., & Miska, E. A. (2014). piRNAs: from biogenesis to function. *Development*, *141*(18), 3458 LP – 3471. https://doi.org/10.1242/dev.094037

Xu, C.-M., & Sun, S.-C. (2020). Expression of Piwi Genes during the Regeneration of Lineus sanguineus (Nemertea, Pilidiophora, Heteronemertea). In *Genes* (Vol. 11, Issue 12). https://doi.org/10.3390/genes11121484

Zattara, E. E., Fernández-Álvarez, F. A., Hiebert, T. C., Bely, A. E., & Norenburg, J. L. (2018). A phylum-wide survey reveals multiple independent gains of head regeneration ability in Nemertea. *BioRxiv*. https://doi.org/10.1101/439497