BIOINFORMATIC IDENTIFICATION OF SMALL RNA (AND THEIR TARGETS) IN THE TESTIS OF CLAWED FROGS (*XENOPUS* AND *SILURANA*): IMPLICATIONS FOR MALE FERTILITY

by

M.J. MADISON-VILLAR

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2011

ACKNOWLEDGEMENTS

I would like to thank all the frogs who died for this study. I would also like to thank Dr. Pawel Michalak. If it weren't for him, I may have lived and died without ever once having set foot in the grate state of Texas. I also want to thank Dr. Paul Chippindale for adopting me in lieu of Dr. Michalak's departure from UTA, as well as my other committee members, Drs. Esthér Betran, Jeff Demuth, and Shawn Christensen. You have all been great people. Further thanks to Drs. Daina Ma and Alie Koroma for being my soundboard and support system; Eldon Prince for assistance with the Ruby scripts; Becky Robinson and Mike Natishyn for being awesome study buddies and having all the stats answers I could ask for; and Dr. Mandy Keogh for just listening and being there when I needed someone to talk to. Last, but certainly not least, I want to thank (insert name here).

November 11, 2011

ABSTRACT

BIOINFORMATIC IDENTIFICATION OF SMALL RNA (AND THEIR TARGETS)
IN THE TESTIS OF CLAWED FROGS (*XENOPUS* AND *SILURANA*):
IMPLICATIONS FOR MALE FERTILITY

M.J. Madison-Villar, PhD

The University of Texas at Arlington, 2011

Supervising Professor: Supervising Professors Paul Chippindale and Pawel Michalak (VBI)

Chapter one sets the stage for this project, which lies somewhere near the intersection of speciation, genome, and reproductive biology.

Chapter two identifies miRNA with shared expression profiles (*Silurana* and *Xenopus*). These miRNA are pulled from a subset that has retained testis expression despite a hybridization induced WGD event early in the *Xenopus* lineage (~21-41 MYA). Allopolyploidization is a revolutionary mechanism of genome evolution, and this event should have afforded an opportunity for diversification and changes in the expression of what are generally highly conserved transcript regulators. It is inferred that the miRNA identified in this study retained testis expression due to strict selection, and that their targets should participate in key pathways, such as those that would ensure reproductive potential. This chapter moves on to identify the most likely targets of these miRNA, and then investigates their fates following WGD. Given the expectation of widespread genome disruption following allopolyploidization, it is hypothesizes that the miRNA targets (*kdm2a, map3k9, kif23, smarca4, clasp1, eed, nob1,* and *abcb1*) will, if retained, show evidence of sub/neofunctionalization. Most targets are clearly

associated with fertility, and several are associated with early embryogenesis. This latter association suggests a possible role for male mediated miRNA outside of the male germline, where they may help to facilitate fertilization and/or help regulate maternally deposited mRNA.

Chapter three identifies a subset of testis expressed miRNA that are misexpressed in the sterile hybrids, and associates these with miRNA known from the testis of mammals. This chapter proposes a subset of miRNA that may have a conserved role in ensuring male reproductive potential in all of vertebrates; it is currently in press with the Journal of Molecular Evolution.

Chapter four investigates pachytene expressed pilRNA (PIWI-like RNA), and how the repeat and non-repeat subpopulations respond to hybridization. Part one describes the repeat associated population (rapilRNA): relative abundance, which transposable elements (TEs) they are associated with, how they map to both Class I and Class II elements, and which DNA strand they are derived from. Part two investigates the response of rapilRNA to hybridization: fold change between hybrids and parental taxa is quantified, and ping-pong amplification is considered, but rejected. Part three identifies misexpressed pilRNA from those having homology to previously identified piRNA, possible targets are identified, and the roles of these targets in male reproduction are investigated. Many of these recover association with SUMOlation, a post-translational modification previously linked to the same cellular locations as the protein (MAEL, MIWI) partners of piRNA (chromatoid bodies).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

OVERVIEW AND BACKGROUND INFORMATION

This dissertation is centered on populations of small RNA from the testis of clawed frogs (family Pipidae, subfamily Xenopodinae, genera *Xenopus* and *Silurana*), particularly those populations which are found in *X. laevis, X. muelleri,* and their sterile interspecific hybrid. Methodologically, the studies contained herein employ high-throughput massive-parallel sequencing, bioinformatic investigation, and a summarization of biological implications. The subject matter lies somewhere near the intersection of speciation, genomics, and reproductive biology. Given that this dissertation is largely multidisciplinary in nature, drawing from multiple subfields within the biological sciences, this overview will be broad; loosely attempting a foundation upon which to rest subsequent chapters.

1.1 Speciation

"When in the offspring of two different animal races one sex is absent, rare, or sterile, that sex is the heterozygous (heterogametic) sex"

(J.B.S. Haldane,1922).

In the study of postzygotic isolation, Haldane's Rule may very well be the only postulate that has held true across a wide array of organisms, including: nematodes, gastropods, insects, fish, amphibians, reptiles, mammals (Schilthuizen et al., 2011), and dioecious plants (Brothers and Delph, 2010). In light of genetics, two competing, though non-mutually exclusive, lines of reasoning have commonly been invoked to explain Haladane's Rule: (1) the dominance hypothesis (Mueller, 1940) and (2) faster male evolution (Wu and Davis, 1993; Wu et al., 1996). However, as with nearly everything in nature, there are no absolutes, and notable exceptions to Haldane's Rule include *Drosophila* (class Insecta, family Drosophilidae; reviewed in Orr, 1997)

and *Xenopus* (reviewed in Orr and Presgraves, 2001; there are no published data for frogs of the genus *Silurana*).

In *Xenopus*, interspecies hybrid males (ZZ) are sterile, though females (ZW) are not (Kobel and Pasquier, 1975; Kobel et al., 1981; Picker, 1985; Picker and de Villiers, 1989; Picker et al., 1996; Fischer et al., 2000). Previous work with *X. laevis, X. muelleri,* and their interspecific hybrids suggested that this clade may be so divorced from the norm (as presented by Haldane's Rule) that sex-reversed frogs will continue to follow the aberrant sterility pattern (Malone and Michalak, 2008a). Interestingly, this deviation occurs in spite of there being a greater degree of misexpression between genes in the ovaries of hybrids and parental taxa than between genes in the testis of hybrids and these same parental taxa (Malone and Michalak, 2008b); this was attributed to the observation that of the misexpressed genes, there exists a bias towards those that are involved in male (as opposed to female) reproduction. Consequentially, hybrid sterility in *Xenopus* was attributed to "maleness," thus making *Xenopus* an exemplary system with which to further test the faster male evolution hypothesis. According to the faster male evolution hypothesis, hybrid males are more likely to be subfertile or sterile because of (1) the sensitive nature of spermatogenesis and/or (2) stronger selection on male-specific genes (Wu and Davis, 1993; Wu et al., 1996). Congruent with this hypothesis, sperm from hybrids is less abundant, less likely to be differentiated, and generally misshapen (Malone et al., 2007).

While this dissertation was not approached in such a manner so as to specifically test the faster male evolution hypothesis, the studies contained within are intended to expand upon current understanding of male hybrid sterility in *Xenopus,* and by default, they will contribute to the growing body of literature on male hybrid sterility.

1.2 Small RNA

Small RNA (smRNA), such as micro RNA (miRNA), small-interfering RNA (siRNA), and PIWI-interacting RNA (piRNA), participate in both gene expression and protein production.

Mature, fully processed miRNA measure ~22 bp in length, and in *Xenopus,* as with many other animals, these predominantly originate from within the introns of the genes they regulate, although some are derived from intergenic locations (Tang and Maxwell, 2008). For miRNA, as is also the case for other species of smRNA, the first step in initiating function is the formation of a heteroduplex with complementary target mRNA. While all smRNA bind to the target mRNA in a nucleotide-dependent fashion, the location of biding can differ depending upon the species of smRNA. For instance, miRNA preferentially bind to the 3' UTR of their target, although they are capable of binding within exons (reviewed in Meister and Tuschl, 2004), and both siRNA and piRNA bind to coding sequence.

MiRNA predominantly act as translational repressors of the genes from which they reside within (Ambros, 2004; He and Hannon, 2004), and for both miRNA and siRNA, the efficacy and mechanism of repression is contingent up how close the smRNA and its target are in terms of complementary base pairing. In instances of low complementation, mRNA translation is repressed through the recruited involvement of miRNA-duplex-containing ribonucleoprotein particles (miRNP) (Meister and Tuschl, 2004). Alternatively, if the degree of complementation is high, the mRNA may be cleaved via the endonuclease properties recruited Arognaute (AGO) proteins in a process termed RNA-interference (RNAi). RNAi belongs to what is dubbed the "RNA Induced Silencing Complex" (RISC) (in He and Hannon, 2004). A third, and less well characterized pathway dubbed the "RNA Induced Transcriptional Silencing Complex" (RITS), is able to assemble a large protein-miRNA complex that engages in site specific heterochromatin formation, thus preventing further transcription (Verdel et al., 2004). In animals, the dominant mechanism is miRNP (Ambros, 2004). While the dominant role of miRNA is knockdown of target mRNA as has just been described, it must be noted that in some instances, miRNA-mRNA interactions actually result in increased protein production (Elmen et al., 2008; He et al., 2005).

Small regulatory RNA, such as miRNA, are not eukaryotic specific (Gottesman, 2005). In metazoans, miRNA families are incredibly conserved, and their evolution follows the pattern of with younger lineages incorporating new families to an ancestrally retained pool (Wheeler et al., 2009), the result of which is strong conservation of core sequences. The degree to which these are retained is so strong that miRNA have been able to resolve the evolution of previously unresolvable inter-species relationships (Sperling et al., 2009). In metazoans, not only are the sequences of miRNA often found to be highly conserved, but so too are the genes that they regulate (Stark et al., 2005), as well as the tissues in which they are expressed (Kai-How Farh et al., 2005). This is thought to be the result of a significant amount of evolutionary pressure maintaining miRNA-mRNA complementation, thereby helping to ensure the proper execution of gene pathways associated with processes such as development and the homeostatic maintenance of specific tissue types (see Massirer and Pasquinelli, 2006). This additional level of conservation likely explains why, in clawed frogs, are not only are those miRNA shared with mammals nearly identical in sequence, but why they largely reside within syntenic blocks (Tang and Maxwell, 2008).

1.3 SmRNA, Epigenetics, and Male Reproduction

Because smRNA are heritable and confer changes to a given gene's expression, independently of this same gene's primary sequence, smRNA are considered to be epigenetic (Russo et al., 1996) genome regulators. Additional examples of epigenetic genome regulators include pre-mRNA processing (e.g. alternative splicing), the results of which are alternate forms of a given protein, all of which have been obtained from a single primary sequence; DNA and histone modifications, the results of which are modifications to chromatin structure and altered transcriptional output, despite an unchanged underlying sequence; and long non-coding RNAs, the results of which are activated complexes that engage in chromatin activation (CACs) or chromatin repression (CRCs) (reviewed in Mattick et al., 2009).

In male mammals, epigenetic abnormalities have a known association with fertility (Emery and Carrell, 2006), and mice unable to properly produce small RNA have elongated sperm and are sterile (Maatouk et al., 2008). It has also been found that mice that misexpress miRNA in their testis are prone to an absence of spermatozoa and degeneration of testis-specific tissues (Papaioannou et al., 2009). Although the exact mechanisms by which miRNAs influence fertility and cell integrity remain poorly understood, it is clear that their dysfunction (or malfunction) can negatively impact mammalian spermatogenesis (Dadoune, 2009; Grandjean and Rassoulzadegan, 2009; Papaioannou et al., 2009; Papaioannou and Nef, 2010). At present, in vivo studies demonstrating a direct role for miRNAs in male fertility of non-mammalian vertebrates are lacking; however, expression profiles in *Xenopus* have been linked to fertile or sterile phenotypes in hybrids (Chapter 3 of this Dissertation; Michalak and Malone, 2008).

In zygotes and early embryos of placental mammals (Eutheria), developmental processes preceding zygotic genome activation have been traditionally attributed to maternal factors within the oocyte. Maternal genome regulation is thought to prevent a conflict of interest between the two parental genomes in regards to factors such as embryo size and maternal and fetal survival. This asymmetric expression of parental genomes is achieved by imprinting, and thus far has only been recovered in mammals (reviewed in Hore et al., 2007). Recent studies, however, suggest this may be oversimplified, as in addition to the paternal genome, sperm have also been found to transmit various RNA, including miRNA (Kimmins and Sassone-Corsi, 2005; Martins and Krawetz, 2007; Rousseaux et al., 2005, 2008). To date, a role for paternally-transmitted miRNA has yet to be found in either fertilization or early embryogenesis (Amanai et al., 2006). This, however, is likely to change as additional research accumulates. Contrary to eutherians, egg-laying vertebrates need not have a system in place to reduce conflict between parental genomes. As such, it appears to not be necessary for early development to be

dominated by maternal factors, and regulatory elements transmitted via sperm have greater potential to carry an active role during embryogenesis.

Sperm cells are, arguably, the most differentiated cells and spermatogenesis may very well be one of the most complex biological processes undergone by any organism. This is because spermatogenesis not only features the germ-cell specific processes of meiotic division, but also processes necessary for the characteristic properties of sperm: shape (chromatin remodeling) and mobility (flagellum development). Spermatogenesis, which incorporates juxtacrine, paracrine and endocrine factor information, evokes thousands of genes and proteins, many of which are testis specific. Although the processes by which spermatogenesis occurs may vary between taxa, germline gene families, such as the *Daz* (Deleted in Azoospermia) family, are conserved from invertebrates to mammals (Xu et al., 2005).

1.3.1 Overview of Spermatogenesis

The process of spermatogenesis occurs in multiple structures of the male reproductive system, including the epithelial Sertoli cells of the seminiferous tubule, and can be partitioned into three major categories: spermatocytogenesis, spermatidogenesis, and spermiogenesis. In spermatocytogenesis, through mitotic and meiotic I division, haploid spermatocytes are formed; in spermatidogenesis, secondary spermatocytes are formed, which then enter meiosis II to produce haploid spermatids; in spermiogenesis, the flagellum of spermatids grows, mitochondria gather and form an axomeme, and spermatid DNA undergoes repackaging and condensation. Nuclear condensation occurs by replacement of somatic histones with intermediate proteins, which themselves are exchanged for protamines (Green et al., 1994; Kistler et al., 1996; Meistrich, 1989). Protamines are a diverse family of argenine-rich late-state spermatid proteins found in both plants and animals, and are essential for normal sperm function. In mammals, there are two protamines: P1 and P2.

Anuran (lineage of frogs and toads) and amniote testis both contain a mosaic of somatic tissues and germ cells. During spermatogenesis, germ cells develop within somatic

cyst-like structures called Sertoli cells, also called follicle, supporting, sustenacular, cysts, nests or nurse cells (reviewed in Oielska, 2009). In anurans, Sertoli cells are large flat cells with multiple structural processes that are connected by desmosomes and tight junctions. These junctions separate the spermatogenic compartment from the seminiferous tubule that they are housed within, producing a blood-testis barrier. Individual tubules are separated by additional somatic interstitial tissues, including Leydig cells, myoid cells, nerve fibers, blood vessels, fibroblasts, and collagen fibrils. Furthermore, Sertoli cells in anuran are the location of sperm maturation, such that by the time spermatozoa exit the testis, they are fully mature (in contrast to mammals, where sperm maturation occurs in the epididymis). In seasonally reproducing taxa, cyst-like Sertoli cells remain closed until environmental cues trigger chromatids to elongate. Once this happens, Sertoli cells open and release mature spermatozoa into the lumen of a tube via spermiation. After traveling through a series of ducts, sperm are expelled into the cloaca.

1.4 Hybridization, Genomic Shock, and Mobile Elements

McClintock (1984) noted that progeny arising from the merging of disparate *Maize* genomes frequently suffered changes in gene expression patterns and DNA-sequence. This hybridization-induced phenomenon was dubbed ‘genomic shock,’ and it was proposed that these genotypic and phenotypic instabilities were the result of relaxed suppression on mobile elements. Mobile elements (also called transposable elements or mobile repeats) are self-proliferating components of the genome which have no obvious purpose; as such, they are frequently referred to as being "selfish," "junk," or "parasitic." While this may sometimes be a fair assessment, mobile elements have also contributed to genome evolution in more traditional ways, as they have, on numerous occasions, been exapted by host genomes to acquire novel functions (Bejerano et al., 2006; Feschotte, 2008). One such example are the RAG1/2 genes, which have immune function; both RAG paralogs facilitate immunoglobulin V(D)J recombination, and RAG1 is required for mature B and T lymphocytes (Mombaerts et al., 1992).

In hybrids, genomic shock may result in the relaxation of heterochromatin, subsequently allowing for increased mobile element transcription, translation, and possible translocations. While both successful and unsuccessful translocations can induce double-stranded DNA breaks, ectopic recombination, chromosomal aberrations, and cellular apoptosis (Pfeiffer et al., 2000), hybridization does not dictate genomic disaster. Frequently, the effects of hybridization are benign or even beneficial (Comings and MacMurray, 2000), and often hybrids can have greater fitness than either parent (positive heterosis). However, when genomic shock does occur in hybrids, the consequences can be devastating. One such example involves resynthesized F1 hybrids with genetic architecture similar to *Aribodopsis suecica* (this taxon originates from a mispairing between *Arabidopsis thaliana* and *A. arenosa* between ~12,000-300,000 ybp). In hybrids, Josefsson et al., (2006) noted impaired fertility, consequential to high rates of seed abortion. In this study, ~95% of seed abortion was ultimately attributed to the derepression of a single mobile element. Hybridization induced activation of mobile elements has also been linked to male sterility (Evgen'ev et al., 1997; Petrov et al., 1995; Vieria et al., 1997), embryonic lethality (Comai et al., 2003), and complete reproductive isolation (Ungerer et al., 2006).

In RNAi as it pertains to mobile elements, two classes of smRNAs have been identified as playing a significant role in their suppression: siRNA and, more recently, piRNA (Aravin et al., 2007a,b; Aravin et al., 2008; Houwing et al., 2007b [repeat associated siRNA (rasiRNA), identified earlier in *Drosophila* (Aravin et al, 2003), has now been synonymized piRNA (see Wahid et al., 2009)]. SiRNA mediated silencing relies heavily on the RISC pathway, while piRNA mediated silencing utilizes both the RISC and RITS pathways. At present, it is unstudied how these repeat-associated smRNA respond to hybridization in vertebrates.

CHAPTER 2

IDENTIFYING TESTIS-DERIVED MIRNA RETAINED IN LIGHT OF DYNAMIC GENOME RESTRUCTURING, AND INFERRING THE LONG-TERM EVOLUTIONARY (21-41 MY) TRAJECTORY OF REVOLUTIONARILY ACQUIRED (ALLOPOLYPLOID) MIRNA TARGETS

XENOPUS LAEVIS, XENOUPS MUELLERI, AND *SILURANA TROPICALIS*

Despite female *Xenopus* having a long history as a model organism for reproductive studies, male *Xenopus* have been virtually ignored. As a result, scientific literature regarding anuran spermatogenesis is lacking (a PubMed search recovers 81 hits for "*Xenopus* sperm," while "*Xenopus* oocyte" recovers 1484 hits). In order to provide insight into this area of reproductive biology, miRNA profiles the clawed frogs *Silurana tropicalis*, *Xenopus laevis*, and *Xenopus muelleri* were investigated. These taxa were selected because: (1) *Silurana tropicalis* and *Xenopus laevis* are the only anurans represented in mirBASE; (2) *Silurana tropicalis* is the only anuran with a sequenced genome (Hellsten et al., 2010); (3) these species correspond to the three largest, most ancient clades within Xenopodinae (Figure 2.1), thus providing an evolutionary frame to work within. Additionally, the emergence of *Xenopus laevis,* having been marked by allopolyploidization (Tymowska and Fischberg, 1982; Bürki and Fischberg, 1985; Chain and Evans, 2006) facilitates an unique opportunity to investigate the long-term evolutionary fate of formerly orthologous miRNA targets forced to cohabitate in genomic chaos.

2.1 miRNA-mRNA Heteroduplex Dynamics

In vivo studies of flies (Brennecke et al., 2003; Xu et al., 2003), worms (Slack et al., 2000), and mammals (Zhao et al., 2007) have revealed that protein production can have an

acute response to the slightest alterations in miRNA abundance (see Ambros, 2003) or miRNA-mRNA binding potential, with a single base pair mutation in the miRNA's 5' complementary

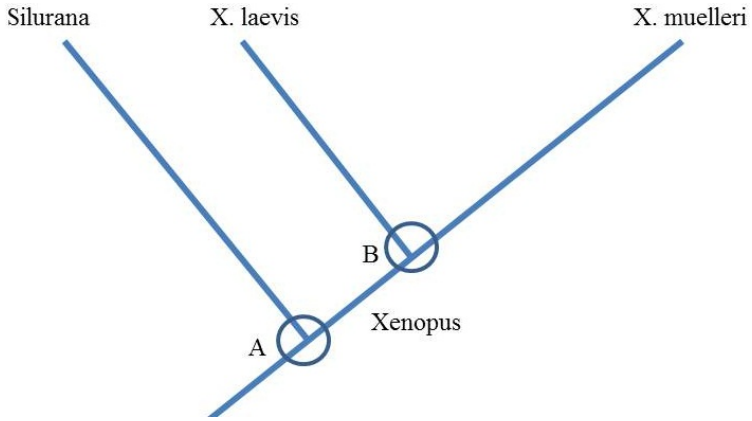

Figure 2.1 Phylogenetic relationships between *S. tropicalis, X. laevis,* and *X. muelleri.* Based on estimates of divergence times, node A is ~53-64 mya and node B is ≤21-41 mya (Evans, 2004; Evans and Chain, 2006).

region resulting in significantly reduced efficiency (Kertesz, 2007). Despite the highly reactive (measured as the strength of the observed knock-down) nature of some of the tens (Lewis et al., 2003) or hundreds (Lewis et al., 2005) of theoretically possible miRNA-mRNA heteroduplexes, the average in vivo pairing only reduce protein production by ≈2-fold (Seitz, 1999). The stark differences in reactivity between miRNA-mRNA heteroduplexes led Seitz (1999) to propose that the bulk of mRNA-miRNA heteroduplexes exist as a mechanism to saturate miRNA so that they are unable to over-repress their true, and highly sensitive, targets. As such, the average low-efficiency observed of lab-induced miRNA-mRNA heteroduplexes is a consequence of "incorrect" pairings.

Given finely orchestrated balance expected of miRNA and their true and pseudo-targets, it can be hypothesized that WGD duplications would be disruptive. Additionally, it can be hypothesized that different origins of said duplications should possess different potentials to disrupt the synergy that has coevolved between genes and their regulatory elements. While in theory everything is "doubled," allopolyploidization brings into a single regulatory network

miRNA of one species and mRNA from another. Despite miRNA and their binding sites being conserved, it can be expected that not all pairings of elements originating from different genomic sources will be as efficient as those between elements originating from the same genomic source. For example, it has been observed in *Xenopus laevis* and *Silurana tropicalis* that miRNA originating from (nearly) identical pre-miRNA are differentially processed, the results of which are taxon (genome) specific (Tang and Maxwell, 2008). As such, it should be expected that miRNA from one taxon may not properly regulate its target in the other. This is consistent with "genomic shock."

Ohno (1970) proposed that gene duplications ought to free one copy of selective constraint, thus promoting its ability to acquire a novel function. However, since advantageous mutations ought to be less common than neutral or deleterious ones, theory (Lynch and Connery, 2000; Lynch and Force, 2000; Lynch et al., 2000) and empirically based models (Lynch and Connery, 2003) all demonstrate that the most likely fate of a duplicated gene is pseudogenization and, for polyploids, the result is expected to be the subsequent purging of one copy from the genome through diploidization (see Wolf, 2001). In terms of whole genome duplications (WGD), this is because one of the consequences of polyploidization is that gene expression increases on a per cell basis (Guo et al., 1996), and as such, changes can result in phenotypic alterations for genes which are sensitive to the effects of dosage (see Chen, 2007; Gaet et al., 2007; Osborne, 2003). Therefore, it should be expected that duplicates of these sorts of genes can only persist under one of two conditions: (1) they confer a beneficial phenotype to the organism, whether that be in the original manner of the gene [theoretically unlikely, though it has been observed in yeast (Dean et al, 2008)] or in the acquisition of a novel function for the gene (neofunctionalization), or (2) the duplicates are expressed in such a manner that the overall performance of the homeologs remains stable (subfunctionalization), such as through the acquisition of complementary mutations that render each copy partially functional. It has been shown that miRNA respond to WGD in a fashion similar to genes, with

many copies being lost (Zhang et al., 2009) and diversification via sub/neofunctionalization in those that are retained (Maher et al., 2006).

The emergence of *Xenopus laevis* some ~21-41 mya was triggered by allopolyploidization (Tymowska and Fischberg, 1982; Bürki and Fischberg, 1985; Chain and Evans, 2006), hybridization-initiated whole genome duplication via the merging differentially regulated genomes. Consequentially, descendants of these ancient hybrids carry the signature of their unique history in the form of previously orthologous genes pairs, dubbed "homeologs" (not to be confused with "homologs"). In response to the widespread misregulation of genes that must have occurred in these hybrids from the combined effects of altered dosage and novel genomic interactions [supported by the observation that large numbers genes (Malone et al., 2007) and miRNA (Malone and Michalak, 2008; Michalak and Malone, 2008; Chapter 3 of this dissertation) are misexpressed in the hybrids of *X. laevis* x *X. muelleri* (Malone et al., 2007), as well as the observation that reciprocal crosses of these taxa can result in deformed offspring, while crosses of both directions frequently result in intersex offspring (unpublished data)], it should be expected that measures to alleviate the deleterious consequences of genomic shock would have been (nearly) immediately employed. One mechanism by which this need could initially be met would be differential regulation of miRNA originating from one, or the other, parental taxa, the result of which would be the quieting of one of the sources of genomic conflict. This response has been observed in F1 allopolyploid *Arabidopsis* hybrids (Ha et al., 2009) and F1 *Xenopus* hybrids (Malone and Michalak, 2008).

In time, theory (Lynch and Force, 2000; Walsh, 1995) and empirical data (Blanc et al., 2003; Blanc and Wolffe, 2004; Bowers et al., 2003; Brunet et al., 2006; Kellis et al., 2004; Lynch and Connery, 2003; Paterson et al., 2004; Seioghe and Wolfe, 1998; Wolfe et al., 2001) tell us that most duplicates would have been lost, and those that are retained should display evidence of sub- or neofunctionalization (Force et al., 1999). However, given the finely orchestrated balance of "true" miRNA-mRNA targeting and the highly reactive nature characteristic of these

heteroduplexes (see Chapter 1), it is reasonable to expect that the timeline for restoring balance would have been accelerated for genes that are regulated by miRNA. Therefore, relatively soon after duplication, one of the homeologs ought to be well on its way to loss or sub/neofunctionalization.

This study has two aims. The first of these is to identify miRNA that have retained tissue specific expression (Ha et al., 2008; Massirer and Pasquinelli, 2006) in *Silurana* and *Xenopus* in spite of an opportunity for diversification and/or change afforded by WGD (Maher et al., 2006). It is expected that the evolution of these miRNA is more restricted than that observed for differentially processed or expressed miRNA (Tang and Maxwell, 2008), indicative of these having association with the most critical of processes, such as ensuring reproductive potential. The second objective is to infer the fate met by their duplicated targets in the 21-41 MY since allopolyploidization.

2.2 Materials and Methods

2.2.1 cDNA Acquisition, Expression Profiling, and Cluster Analysis

Total RNA was obtained from the ground testes of freshly sacrificed *X. laevis*, *X. muelleri* and *S. tropicalis* using Ambion RNA extraction kits (IACUC protocol number A08.002). To obtain <40 nt RNAs, the samples were fractionated using an Ambion flashPAGE fractionator. Libraries of small RNA cDNAs were constructed and sequenced using ABI's SOLiD sequencing next generation technology (outsourced at the University of Oklahoma Medical Sciences core facility). All samples were run on a 35 nt array.

MiRNA expression was quantified employing a GeneSifter (Perkin Elmer, 2011) pipeline. This was accomplished by identifying the miRNA components of each individual's small RNA library (employing miRBase v. 15; Griffiths-Jones, 2006, 2010; Griffiths-Jones et al., 2006, 2008; Kozoma and Griffiths-Jones, 2011), tallying the number of identified reads, normalizing these against the total number of reads per library, and producing clusters of miRNA with similar expression profiles.

In k-medoids clustering, the objective is to find k groups with a minimal distance between all members of the group and those members with the lowest average dissimilarity (medoid) using the equation:

$$\sum_{j \in Ci} d(i,j)$$

where d(i,j) is the distance between medoid i and object j in cluster Ci. This iterative process can be summarized by (a) choosing k objects at random to be the initial medoids, (b) assigning each object to the cluster having the closest medoid, (c) recalculating the position of the medoid, and (d) repeating b and c until no further change occurs.

PAM (Partitioning Around Medoids) clustering is a more robust method of unsupervised cluster analysis that builds upon the basic principal of k-medoids clustering, and has the additional feature of the silhouette plot. This enables users to select the most optimal number of clusters or to select individual clusters that are well supported by the analysis. The mean silhouette width is a measure of how well all items fit into their assigned cluster. General interpretation of silhouette width dictates that negative values are incorrectly clustered; 0.0-0.25 indicates the clusters are not structured; 0.25-0.5 indicates they are weakly structured; 0.5-0.75 moderately structured; and 0.76-1.0 strongly structured.

For this study, cluster analysis was conducted in order to identify groups of miRNA that had similar expression profiles. To achieve this, a PAM clustering algorithm was employed to produce between two and 25 individual clusters, each having an undefined number of members. For each round of clustering (e.g. two clusters for round one, three clusters for round two, etc.) distance was measured in Manhattan units with complete linkage (maximum distance between any two objects from different clusters), and the row center was set to row mean. Once the initial compilation of clustered miRNA having similar expression profiles was completed, their corresponding dendrograms, which depicted the expression profiles for all

members of a single cluster, were collapsed by similarity of membership into the fewest number possible. These groupings were then given a unique alphabetic identifier (e.g. A). If a given cluster was particularly large, its dendrogram was used to break the dataset into subgroups, which were then given an additional numeric identifier (e.g. A1).

2.2.2 Identification of Gene Targets

Putative miRNA targets of a single cluster (e.g. "A" or "A1") were identified employing the software programs miRAT! (Dombkowski, 2011) and DAVID Bioinformatics Resources (version 6.7; Huang et al., 2009). miRAT! is an online software application specific to the identification of miRNA targets from mice, rats, and humans, and it is combinatorial in that it employs other software as well as major databases: it directly accesses data from MicroCosm Targets (version 5; EMBL-EBI, 2011), which, in turn, obtains sequences from Ensembl Genome Browser (EMBL-EBI and Welcome Trust Sanger Institute, 2011). In addition to this connection, miRAT! also links to NCBI and Ensembl, as well as allows for the automated transfer of the list of target genes it identifies to DAVID. This web-based software provides functional annotation of ontologies and pathways for genes on said list, as well as provides statistical support for each grouping it proposes. Groups with $p \geq 0.05$ were excluded from further analyses.

Clawed frogs are not among the organisms miRAT! supports, and therefore miRNA were entered with a *Mus* identifier [for instance, xtr-mir-222 was entered as mmu-mir-222; nomenclature for the naming of miRNA is that "new" miRNA from one species are named after previously identified miRNA from a different species, such that xtr-mir-222 and mmu-mir-222 should be identical, or nearly so, in sequence (http://www.mirbase.org/help/nomenclature.shtml)]. While the swapping of species identifiers may be unconventional, it was felt to be justified by the fact that metazoan miRNAs tend to be highly conserved (Wheeler et al., 2009). In these organisms, miRNA sequences are retained due to evolutionary pressure that maintains miRNA-mRNA complementation, thereby helping to ensure the proper execution of gene pathways for key biological processes (see Massirer and

Pasquinelli, 2006). For the sake of clarity, each list of putative target genes that were identified using a specific miRNA cluster were called “Targets (cluster).” For example, if four miRNA clusters were obtained from the PAM clustering, the target genes identified by miRAT!/DAVID for the first cluster, “A,” were called “Targets A” and those for the fourth cluster, “D”, were called “Targets D.”

In order to identify which of the putative miRNA gene targets were most likely to be “true” miRNA targets, and therefore reduce the likelihood of spurious target identification, all “Targets” lists having $p \leq 0.05$ were imported into Microsoft Excel. Here, “Targets” lists were subject to pairwise comparisons to identify which, if any, genes were shared between multiple “Targets” lists, and their associated information was noted.

2.2.3 Identification of Homeologs

Duplicated pairs of genes in *X. laevis* were identified using transcripts obtained for *X. laevis* and *S. tropicalis*. *X. muelleri* was not included in these analyses because, unlike the other two, this taxa is not well represented in either NCBI's GenBank (Benson et al., 2010) nor NCBI’s UniGene (Pontius et al., 2003) database. Currently, UniGene contains the following data for *X. laevis:* 215,184-3' ESTs, 384,357-5' ESTs, 249 unknown ESTs, 29,861 mRNA sequences, and 351 high-throughput cDNA (HTC) sequences; a grand total of 630,002 entries organized into 35,169 gene clusters. Each cluster, in theory, corresponds to a single gene.

In order to discern *X. laevis* homolog pairs from alleles, Genious software (Drummond et al., 2011) was employed. Using this software, DNA sequences for a given gene were aligned and compared using sequence divergence as well as by a maximum likelihood tree. Trees were produced by using the PhyML algorithm (Guindon and Gascuel, 2003), employing the general time reversible (GTR) model of evolution (Yang, 1994). Brach support was estimated by 100 bootstrap replicates. The GTR model of evolution was selected for the fact that it allows for variable substitution rates across the terminal nodes; as such, it is particularly appropriate for estimating rates of evolution among gene copies that are likely to be evolving asynchronously.

Homeologs were defined as gene copies that were positioned separate from one another as well as separate from the ortholog, while alleles were defined as gene copies that clustered together on the tree, but were separate from other clusters as well as separate from the ortholog (Figure 2.2). Furthermore, it was expected that the genetic distance of alleles was much lower than distances between homeologs or between orthologs.

Sub/neofunctionalization was inferred from the following criteria: (1) the identification of complementary mutations that resulted in loss of function for domains, as identified from the mRNA transcript of a given gene; performed using Pfam (version 25; Welcome Trust Sanger Institute, 2011) or (2) clear evidence of differential expression between the members of a homolog pair (e.g. expression localized to different tissues or expression associated with different life stages); performed using NCBI's UniGene Database (Pontius et al., 2003). The additional finding of clear differences in the rates of evolution for homeologs, as identified by (a) calculating branch lengths on the ML tree, measuring the distance between the *S.tropicalis* ortholog and each member of a homolog pair or (b) pairwise genetic distance between each member of a homolog pair and the *S. tropicalis* ortholog, was considered additional support for the identification of sub- or neofunctionalization.

2.2.4 Identification of Sub/Neofunctionalization

It is duly noted that this outlined approach for identifying sub/neofunctionalization may be imprecise; however, when all three criteria are used in combination, they constitute a more rigorous approach to identifying sub/neofunctionalization than was employed by Morin et al. (2005). Their study relied solely on data available in the UniGene database.

2.2.5 Ka/Ks Measures of Selection

In order to identify the mode of selection under with a homolog pair is currently evolving, a Ka/Ks tests was performed using the software program KaKs_Calculator (Zhang et al., 2006). This test infers the mechanism of selection a given pair of genes is presently evolving under by producing a ratio between Ka (numerator) and Ks (denominator). The value

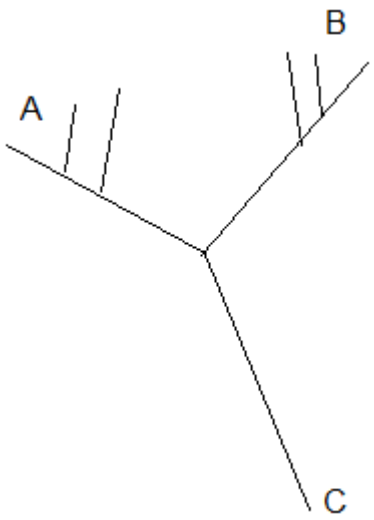

Figure 2.2 Homeologs vs. Alleles. Homeologs were those sequences that did not group together (A and B), nor did they group with the ortholog (C). Alleles for a homeolog were those sequences that clustered near one another on the tree, such as those near A or B, separate from the ortholog C.

Ka is calculated as the number of non-synonymous substitutions per non-synonymous site; Ks is calculated as the number of synonymous substitutions per synonymous site.

The software program KaKs_Calculator calculates values for Ka, Ks, and an associated p-value under a number of statistical models, employing both approximation and maximum likelihood methods. In order to avoid any bias that may have been associated with the selection of one subset of values from all that were returned, the grand mean for each parameter was calculated and reported. The results of this test were interpreted using standard practice:
>1 was used to indicate positive selection, <1 was used to indicate purifying selection, and ~1 was used to indicate neutral/nearly neutral evolution.

2.3 Results

2.3.1 MiRNA and Gene Target Identification

SOLiD sequencing recovered 21,738,343, 18,991,070, and 18,273,999 reads for *X. laevis*, *X. muelleri*, and *S. tropicalis*, respectively (Table 2.1). Only a minority of the reads

mapped to known rRNAs, snRNAs, miRNAs, or to the reference *S. tropicalis* genome. Despite overall low mapping of reads to the genome, the low recovery of rRNA reads clearly demonstrate that the libraries were prepared from high quality small RNA. Due to the sheer abundance of rRNA in total RNA samples, a high number of rRNA hits would indicate that the original sample was degraded.

Table 2.1 Reads recovered from small cDNA libraries for *X. laevis*, *X. muelleri*, and *S. tropicalis*. The low representation of rRNA and snRNA for each library shows that the quality of the reads is high, despite the low overall mapping to the reference (*S. tropicalis*) genome.

	X. laevis	***X. muelleri***	***S. tropicalis***
Total Reads	21,738,343	18,991,070	18,273,999
rRNA & snRNA	9,832	3,750	2,889
miRNA	394,550	197,914	201,082
Map to genome	2,068,742	1,429,957	2,360,902
Remainder unmapped	18,370,543	16,728,211	15,732,721

Cluster analysis of the miRNA identified five clusters, named “A”-“E” (Figure 2.3). Cluster “E” was much larger than the others, and was subsequently divided into subclusters “E1”- “E5.” Across all, miRAT! -DAVID analyses returned 556 “Targets,” (APPENDIX A) whose annotation category had $p \leq 0.05$. Many of these 556 targets belonged to more than one annotated category, making the number of unique targets much smaller. Pairwise comparisons between all 556 possible gene target revealed eight that were “Targets” of more than one miRNA cluster (Table 2.2).

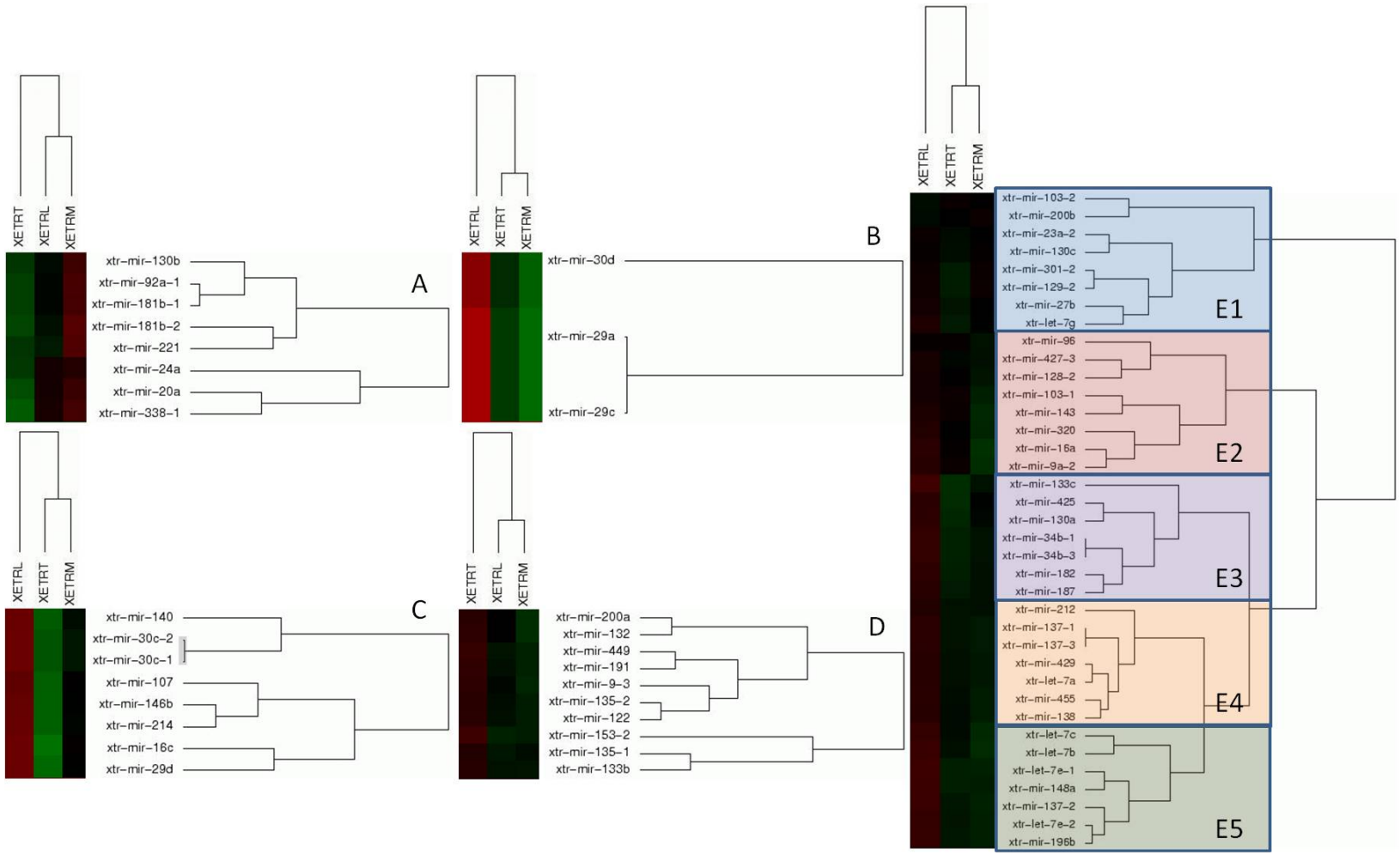

Figure 2.3 Clusters of miRNA obtained from the PAM cluster analyses.

Table 2.2 Eight gene IDs existing in at least two of the clusters A-E5. The Uniprot gene and protein names associated with each is listed, with those in bold representing the most recently accepted name.

Gene ID	***Clusters***	***Gene Name(s)***	***Protein Name(s)***
225876	A,B	**KDM2A**, CXXC8, FBL7, FBXL11, JHDM1A, KIAA1004	**lysine (K)-specific demethylase 2A**; JmjC domain-containing histone demethylation protein 1A
13626	A, E3	EED	embryonic ectoderm development
20586	A, E4	**SMARCA4**, BAF190A, BRG1, SNF2B, SNF2L4	**Transcription activator BRG1**; SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 4
338372	E1, E3	**MAP3K9**, MLK1, PRKE1	Mitogen-activated protein kinase kinase 9
71819	B, E1,E2	**KIF23**, KNSL5, MKLP1	**Kinesin-like protein KIF23**; Kinesin-like protein 5; Mitotic kinesin-like protein 1
76499	E3, E4,E5	**CLASP2,** KIAA0627	**CLIP-associating protein 2**, Cytoplasmic linker-associated protein 2, Protein Orbit homolog 2
67619	A, E4	NOB1	NIN1/RPN12 binding protein 1 homolog (*S. cerevisiae*)
18669	B, E5	ABCB1B	ATP-binding cassette, sub-family B (MDR/TAP), member 1B

2.3.2 Homeologs of kdm2a *and Inferences of Sub/Neofunctionalization*

GenBank contained two *S. tropicalis* and zero *X laevis* transcripts (Table 2.3), though BLAST analysis against the *X. laevis* ESTs database recovered six hits (Table 2.4). One of these was potentially protein coding (54841398), and would code for an >214 amino acid protein. This transcript contained a PHD finger-like zinc binding domain and a nucleoplasmin-like domain; PHD fingers are found in other F-box proteins and nucleoplasmin is similar to the F-box. Due to this transcript having evolved domains not found in the original protein or the homeolog, sub/neofunctionalization was inferred. Ka/Ks was not calculated due to numerous indels in the 1911 bp sequence that made it impossible to have a long stretch of sequence to use for calculations (Figure 2.4a). Strangely, transcripts were randomly and frequently

recombined within, and across, taxa (Figure 2.4b), resulting in a reticulated gene tree not reflective of the species tree (Figure 2.4c).

Table 2.3 BLAST and Pfam results for Xenopodinae *kmd2a.*

Species	**Sequence Length**	**Sequence Name**	**GenBank ID**	**Domain(s)**
S. tropicalis	4491 bp	kdm2a	NM_001011176.1	JmJc, zf-cXXXc, SR-25, Fbox
S. tropicalis	1661 bp	kdm2a-like	XM_002938501.1	Fbox

Table 2.4 ID of *X. laevis kmd2a* EST hits.

Stage	***Tissue***	***EST Name***	***GenBank ID***	***Domain(s), etc.***
Adult	Testis	AGENCOURT_18791244	CK800324.1	
Embryo	Stage 17/19	AGENCOURT_54841398	DR720257.1	PHD finger-like, nucleoplasm-like
Embryo	Stages 56-62	AGENCOURT_73651956	EB479077.1	Mobile Elements
Adult	Spleen	AGENCOURT_72767842	EB482422.1	
Embryo	Dorsal blastopore lip	AGENCOURT_26179322	CO388033.1	Mobile Elements
Adult	Spleen	AGENCOURT_72757474	EB471730.1	

2.3.3 Homeologs of eed *and Inferences of Sub/Neofunctionalization*

GenBank contained one *S. tropicalis* and eight *X. laevis* transcripts (Table 2.5). Across 1281 bp, two groups were evident (Figure 2.5). They differed by ~7% across groups, ≤1% within group, and were ~91.5% similar to *S. tropicalis*. Pfam identified one domain per sequence (Table 2.5). Therefore, complementary mutations could not be noted. Based upon tree topology, homeolog *eed-b* appeared to be evolving under a recent reduction in selection, with all divergence from this branch contained in a star-like terminus; this topology can be

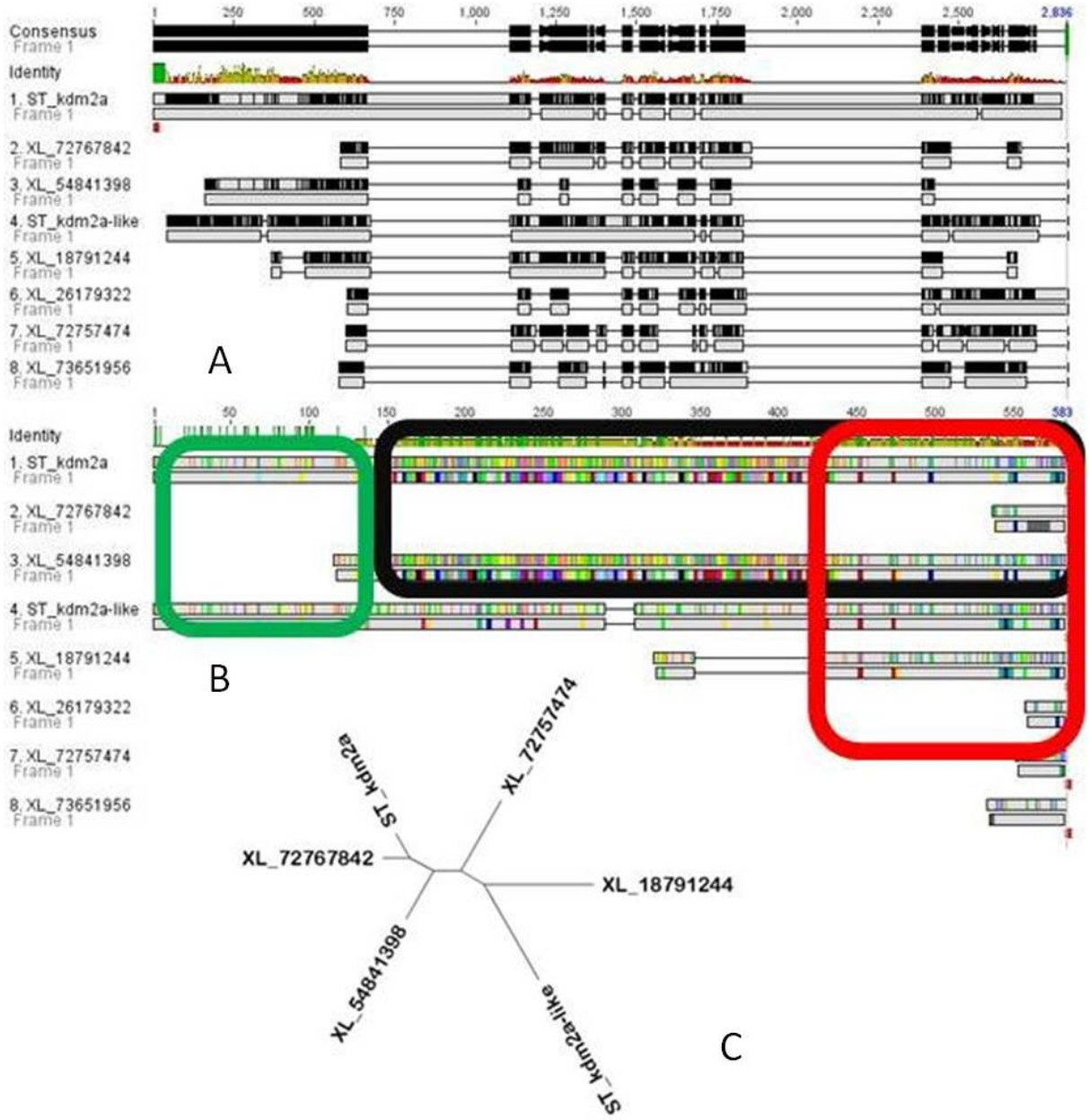

Figure 2.4 (a) Sequence alignments of *S. tropicalis kdm2a* and *X. laevis* EST hits. (b) Across the length of the sequences, consecutive stretches of the same sequence would almost/perfectly match another, regardless of taxon. (c) This resulted in an unrooted maximum likelihood gene tree different from the species tree.

Table 2.5 BLAST and Pfam results for Xenopodinae *eed.*

Species	*Sequence Len*	*Sequence Name*	*GenBank ID*	*Domain(s)*
S. tropicalis	1917 bp	eed	NM_001017325	WD40
X. laevis	1850 bp	eed-a	NM_001096048	WD40
X. laevis	1793 bp	eed-a partial	BC077511.1	WD40
X. laevis	1831 bp	xeed	AJ421945.1	WD40
X. laevis	1311 bp	eed partial	AY034136.1	WD40
X. laevis	1473 bp	eed protein variant 1	AF460180.1	WD40
X. laevis	2418 bp	eed-b	NM_001088885.1	WD40,WD40
X. laevis	1513 bp	eed pseudogene	AF460179.1	WD40,WD40

observed following release from a selective sweep. Using *xeed* and *eed-b* (due to the lack of intra-group divergence for either homeolog, these specific sequences were selected to represent their homeolog because they were full, rather than partial, transcripts). Ka/Ks indicated strong positive selection (Ka=0.076; Ks=0.033; Ka/Ks=2.338; p=0.013), thus supporting the tree topology. Furthermore, UniGene profiling detected differences in timing, tissue, and strength of expression between homeologs (Table 2.6). Given all data, neofunctionalization was inferred.

Table 2.6 UniGene digital northern for Xenopodinae *eed.* Expression is measured as transcripts per million (TPM).

Life Stage	***eed-a* TPM**	***eed-b* TPM**	Tissue	***eed-a* TPM**	***eed-b* TPM**
Oocyte	166	611	**Animal cap**	0	820
Egg	215	573	**Ectoderm**	61	246
Blastula	0	348	**Endomesoderm**	32	65
Gastrula	47	177	**Ovary**	61	123
Neurula	111	0	**Spleen**	54	51
Tail-bud embryo	37	111			
Metamorphosis	0	34			
Adult	9	19			

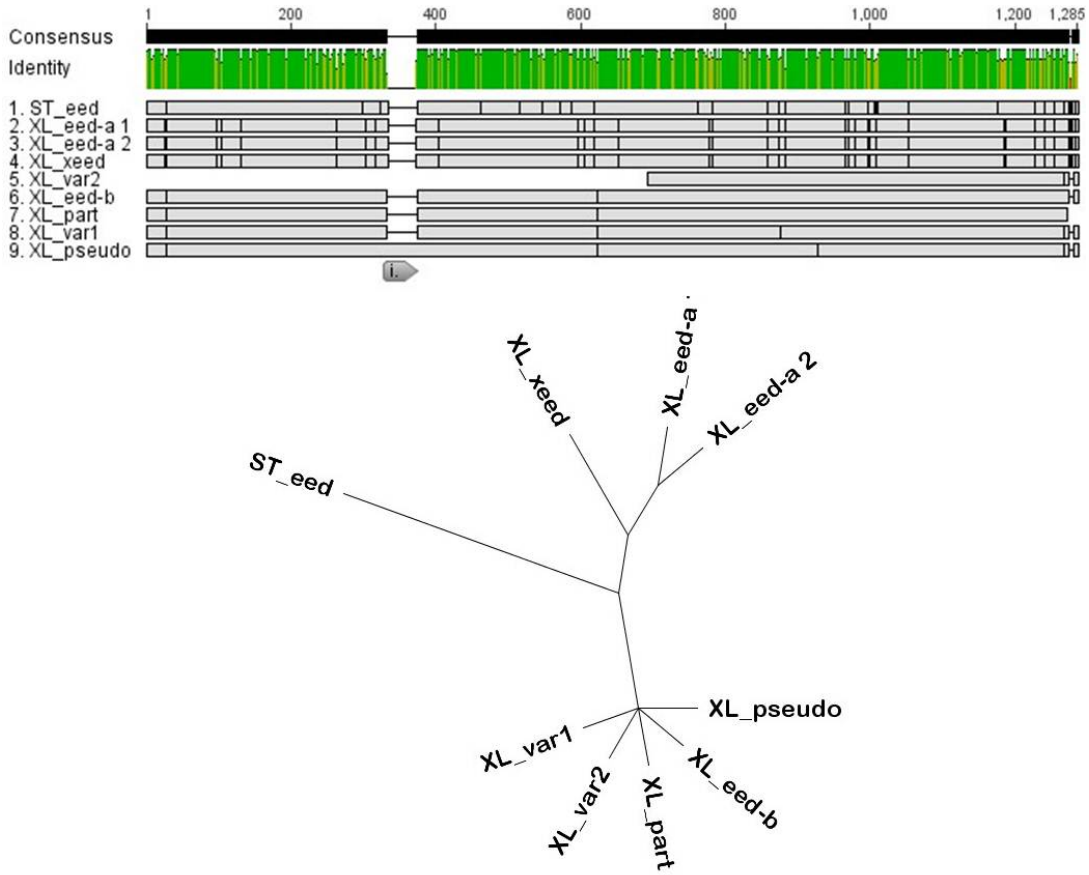

Figure 2.5 Sequence alignment and unrooted maximum likelihood tree of all Xenopodinae *eed* homologs. The extended 3' end of *X. laevis eed-b* has been trimmed; lines on alignments denote transversions.

2.3.4 Homeologs of smarca4 *and Inferences of Sub/Neofunctionalization*

GenBank recovered four *S. tropicalis* and seven *X. laevis* transcripts (Table 2.7). Across 1715 bp, *X. laevis* transcripts were ~99% similar, and ~90-93% similar to *S. tropicalis* (Figure 2.6a). Given the degree of sequence identity, there were no complementary mutations to be detected. Interestingly, the *S. tropicalis* ortholog stopped at about 3200 bp, while the *X. laevis* homeologs contained a gap at this region, and continued on with another 1907 bp of coding sequence (Figure 2.6b). Despite the recovery of 155 ESTs, it was not possible to

Table 2.7 BLAST and Pfam results for Xenopodinae *smarca4*. Note that *brg1* is another name for this same gene.

Species	***Sequence Length***	***Sequence Name***	***NCBI ID***	***Domains***
S. tropicalis	1765 bp	TGas053p20	CR761474.2	QLQ, HSA
S. tropicalis	1844 bp	7030469	BC084531.1	QLQ, HSA
S. tropicalis	1845 bp	7597000	BC135259.1	QLQ, HSA
S. tropicalis	4477 bp	smarca4	NM_001113662.1	QLQ, HAS, BRK, SNF2 N
X. laevis	5487 bp	smarca4	NM_001171953.1	QLQ, HAS, BRK, SNF2 N, Helicase C, Bromodomain
X. laevis	5031 bp	brg1	AY726636.1	QLQ, HAS, BRK, SNF2 N, Helicase C, Bromodomain
X. laevis	1814 bp	5078610	BC126040.1	QLQ, HSA
X. laevis	1801 bp	4202564	BC084056.1	QLQ, HSA
X. laevis	1845 bp	4030790	BC097511.1	QLQ, HSA
X. laevis	358 bp	brg1 part	AF373872.1	SNF2 N

produce an alignment. Ka/Ks was not calculated due to the absence of full length transcripts for homeologs.

The extreme degree of similarity between all sequences was unexpected, and could be explained by three very different scenarios: (1) one copy has been purged, and that which remains is under extreme purifying selection, (2) one copy has been purged, and there is currently, or was until very recently, a prolonged selective sweep on the remaining copy, thus preventing the accumulation of divergent alleles, or (3) sampling error; additional diversity exists, but is, by chance, not included among the sequences deposited in GenBank. Given the lack of a single interpretation, fate could not be discerned.

Figure 2.6 (a) Sequence alignment across all of *smarca4.* (b) As denoted by the orange STP arrows, the 3' end of the *X. laevis* transcripts were 1907 bp longer than their *S. tropicalis* ortholog.

2.3.5 Homeologs of map3k9 *and Inferences of Sub/Neofunctionalization*

GenBank contained one *S. tropicalis* and one *X. laevis* (Table 2.8) transcript. *X. laevis* contained short deletions in the 3' region which resulted in a frameshift and 1030 bp early truncation (Figure 2.7); this truncated protein contained 63 fewer active sites than its ortholog. There were too few transcripts to perform a Ka/Ks test. Given the retention of function in this transcript, subfunctionalization may be at play; however, the present inability to identify complementation makes a confirmation impossible, and fate is not possible to infer.

Table 2.8 BLAST and Pfam results for Xenopodinae *map3k9.*

Species	*Sequence Lenght*	*Sequence Name*	*NCBI ID*	*Domains*	*Active Sites*
S. tropicalis	3306 bp	mapk9	XM_002936894.1	SH3, Pkinase Tyr	309
X. laevis	3768 bp	5542423	BC045104.1	SH3, Pkinase Tyr	246

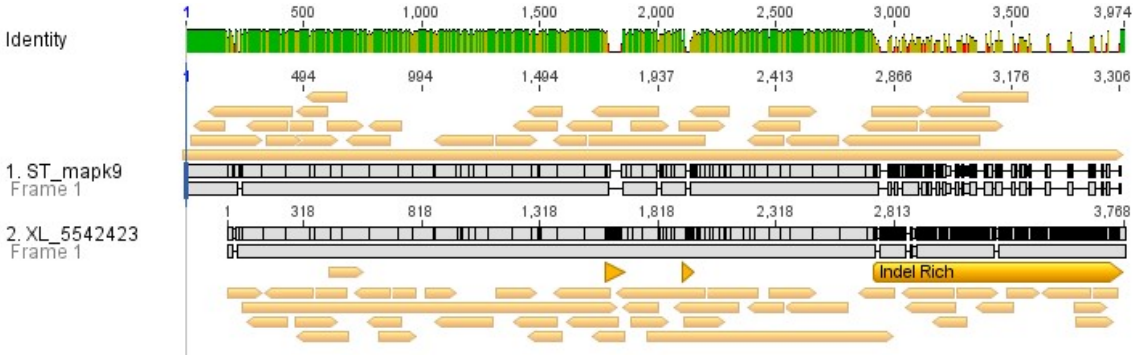

Figure 2.7 Alignment of orthologous *map3k9* transcripts. The *X. laevis* transcript contains a number of indels, including within a conserved region, despite purifying selection.

2.3.6 Homeologs of kif23 *and Inferences of Sub/Neofunctionalization*

GenBank contained one *S. tropicalis* and three *X. laevis kif23* (Table 2.9). Two *X. laevis* groups were evident (Figure 2.8a); these were ~9.5% divergent from one another and ~90.8% similar to *S. tropicalis*. Pfam identified only a single kinesin domain in the full length transcripts; therefore, complementary mutations could not be identified. The ML tree placed *S. tropicalis* and 5542010 together, separate from the *kif23* homolog (Figure 2.8b); this was corroborated by the alignment, as these two sequences shared an 11 bp indel at ~ 2060, as well as a 269 bp indel at ~1170; despite these shared regions, there were regions that were exclusive to the *X. laevis* sequences, so it was evident that this particular sequence was intermediate the two taxa. Pfam did not identify any domains in the region of the 269 bp indel. The addition of five ESTs (Table 2.9) recovered that the *kif23* branch had fewer branching events than 5542010, and that these appeared to be of somewhat more recent origin (Figure

2.8c). Despite the recurrent diversification of the homeologs along both branches, the Ka/Ks test recovered evidence of purifying selection (Ka=0.032, Ks=0.249, Ka/Ks=0.130, p≤0.001), and identification of sub/neofunctionalization from these data was ambiguous. However, one lineage does appear to have a lower rate of evolution, indicating that one homeolog is less constrained than the other.

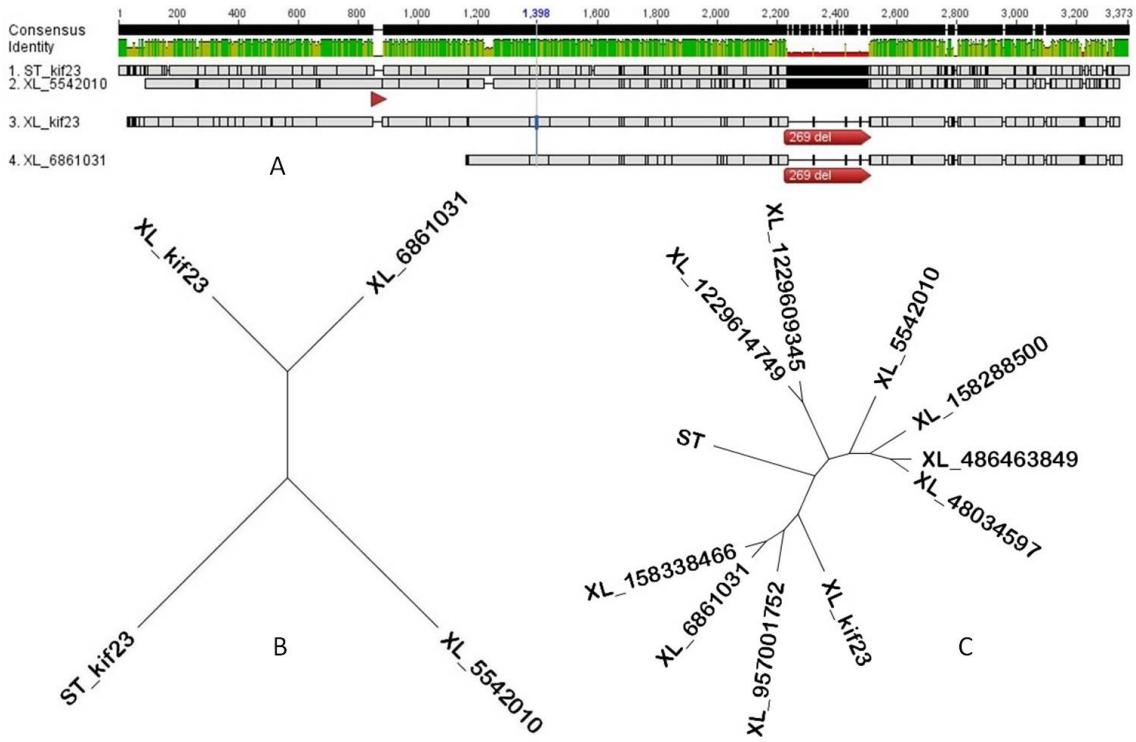

Figure 2.8 (a) Sequence alignment across the three sequences available in GenBank. Lines on individual alignments denote positions where transversions were located. (b) Unrooted maximum likelihood tree of all Xenopodinae *kif23* homologs. (c) Tree from all orthologs plus five ESTs.

Table 2.9 BLAST, UniGene, and Pfam results for Xenopodinae *kif23.*

Species	***Sequence Length***	***Sequence Name***	***NCBI ID***	***Domains***
S. tropicalis	3304 bp	kif23	NM_001011104.1	Kinesin
X. laevis	2978 bp	kif23	NM_001095075.1	Kinesin
X. laevis	1887 bp	6861031	BC094157	Kinesin, E2R135

Table 2.9--*Continued*

Species	***Sequence Length***	***Sequence Name***	***NCBI ID***	***Domains***
X. laevis	3183 bp	5542010	BC046735	Kinesin
X. laevis	EST	957001752	DR718092	
X. laevis	EST	6861031	BC094157	
X. laevis	EST	158338466	CA788129	
X. laevis	EST	1229614749	DY565807	
X. laevis	EST	1229609345	DY548457	
X. laevis	EST	486463849	CO554586	
X. laevis	EST	158288500	DY565807	

2.3.7 Homeologs of clasp1 *and Inferences of Sub/Neofunctionalization*

GenBank contained one *S. tropicalis* and five *X. laevis* sequences (Table 2.10); two groups were evident (Figure 2.9). Across 4407 bp, these differed by ~7%, and both were ~88% similar to *S. tropicalis*. Analysis of the domains did recover a HEAT domain in one of the homeologs; this domain is present in mammalian homologs, and as such, was not used to indicate the acquisition of novel function (Table 2.10). The tree indicated that despite slightly greater diversification of one homeolog, the rate of evolution for this lineage was less than that of the other (Figure 2.9). This other branch contained a single bifurcation; the distance between the internal node and the terminal nodes was greater than the distance between this internal node and the *S. tropicalis* ortholog. UniGene expression profiling found differences in timing, tissue, and overall abundance (Table 2.11); therefore, sub/neofunctionalization could be inferred. Interestingly, the Ka/Ks test recovered strict purifying selection (Ka=0.015, Ks=0.230, Ka/Ks=0.065, $p \leq 0.001$). The diversification of this gene and a seeming increase in the rate of evolution in one homeolog in spite of purifying selection suggests relaxed selection following allopolyploidization.

2.3.8 Homeologs of nob1 *and Inferences of Sub/Neofunctionalization*

GenBank contained two *S. tropicalis* and two *X. laevis t*ranscripts (Table 2.12). *X. laevis* transcripts were 98.9% similar. Given their degree of similarity, it was not possible to

Table 2.10 BLAST and Pfam results for Xenopodinae *clap1*

Species	***Sequence Length***	***Sequence Name***	***NCBI ID***	***Domains***
S. tropicalis	4947 bp	clasp1	NM_001097239.1	CLASP N, Ipi1 N, TAN, Cdn1
X. laevis	7426 bp	clasp1-a	NM_001094646.1	CLASP N, Ipi1 N, TAN, Cdn1, HEAT
X. laevis	3089 bp	4679959	BC077460.1	Ipi1 N, Cdn1
X. laevis	5796 bp	clasp1	BC133747.1	CLASP N, Ipi1 N, TAN, Cdn1, HEAT
X. laevis	5802 bp	clasp1-b	NM_001135034	CLASP N, Ipi1 N, TAN, Cdn1, HEAT
X. laevis	1809 bp	4680619	BC082871.1	Ipi1 N, Cdn1

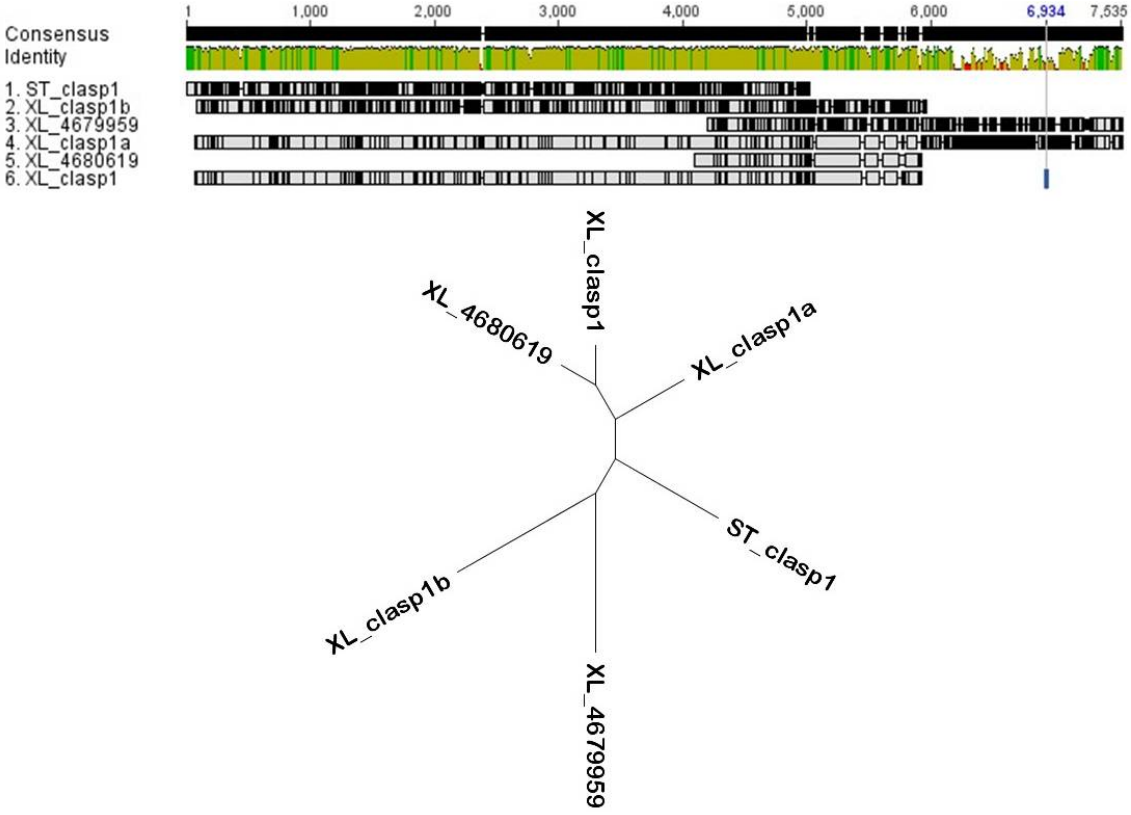

Figure 2.9 Sequence alignment and unrooted maximum likelihood tree of all Xenopodinae *clasp1* homologs. Lines on individual alignments denote transversions.

identify complementary mutations, nor was it necessary to produce a tree. The UniGene database contained 86 transcripts; nucleotide, amino acid, and combined methods could not produce a continuous alignment. Given these, there was not enough data to make inferences regarding the fate of the gene duplicates.

Table 2.11 UniGene digital northern for *X. laevis clasp1.* Expression is measured as transcripts per million (TPM).

Life Stage	***clasp1a* TPM**	***clasp1b* TPM**	**Tissue**	***clasp1a* TPM**	***clasp1b* TPM**
Oocyte	111	0	**Brain**	71	0
Gastrula	47	6	**Ectoderm**	15	0
Neurula	335	111	**Endomesoderm**	32	0
Tail-bud Embryo	335	0	**Head**	76	0
Adult	47	9	**Heart**	483	0
			Ovary	123	61
			Skin	188	0
			Thymus	0	395

Table 2.12 BLAST and Pfam results for Xenopodinae *nob1.*

Species	**Sequence Length**	**Sequence Name**	**NCBI ID**	**Domains**
S. tropicalis	2170 bp	nob1	NM_001016830.2	Nob1 zn binding
S. tropicalis	1475 bp	nin1 partial	BC091607.1	Nob1 zn binding
X. laevis	1700 bp	nin1 binding protein	BC084069.1	Nob1 zn binding
X. laevis	1700 bp	nin1 binding protein	BC084069.1	Nob1 zn binding
X. laevis	1619 bp	nob1	NM_001089279.1	Nob1 zn binding

2.3.9 Homeologs of abcb1b *and Inferences of Sub/Neofunctionalization*

GenBank contained one *S. tropicalis* and two *X. laevis* transcripts (Table 2.13). The *S. tropicalis* and *X. laevis* transcripts were only 58.7% similar, and *X. laevis* transcripts were 99.5% similar. Given their similarity, it was not possible to identify complementary mutations, nor was it necessary to produce a tree. The UniGene database contained 30 transcripts; however, nucleotide, amino acid, and combined methods could not produce a continuous alignment. Given these, there was not enough data to make inferences regarding the fate of gene duplicates.

Table 2.13 BLAST and Pfam results for Xenopodinae *abcb1b.*

Species	Sequence Length	Sequence Name	NCBI ID	Domains
S. tropicalis	4050 bp	abcb1	XM_002933297.1	ABC membrane, ABC trans, ABC membrane, ABC trans
X. laevis	4390 bp	abcb1	NM_001087925.1	ABC membrane, ABC trans, ABC membrane, ABC trans
X. laevis	4228 bp	8820796	BC167559.1	ABC membrane, ABC trans, ABC membrane, ABC trans

2.4 Discussion

The current study identified five clusters of miRNA retained in the testis/sperm of Xenopodinae frogs, as well and eight possible candidate miRNA targets. The protein products of the eight candidate genes serve an array of functions, many of which are presently known to be critical for male reproduction/fertility as well as embryonic development. Of these eight genes, four are transcriptional regulators, two of which moderate transcription through the modification of chromatin architecture [*kdm2a* (Tsukada et al., 2006) and *smarca4* (reviewed in Magnani and Cabot, 2009; Zheng et al., 2004)]. A third regulator of transcription [*map3k9* (reviewed in Thompson et al., 2001)] phosphorylates the transcription factor c-JUN.

Interestingly, the last of the transcriptional regulators (*eed*) spans the gap, and associates with both chromatin remodeling factors (Morey and Helin, 2010) and general transcription factors (Showell and Cunliffe, 2002).

In addition to participation in the genomic regulation of gene expression, these four genes should also participate in the relay of genic information during both mitosis and meiosis. Furthermore, the chromatin associated genes should be expected to serve additional functions related to genome packaging during the later stages of spermatogenesis. Not surprisingly, misexpressed *eed* has been attributed to arrested sperm maturation (Steilmann et al., 2010), and *kdm2a*, which is thought to activate the testis specific genes *tnp2* and *pro1,* is clearly linked to defective chromatin condensation as well as lower sperm count and immotility; the result of which are male sterility (Okada et al., 2007). Interestingly, this phenotype matches what is observed in sterile *X. laevis* x *X. muelleri* males (Malone et al, 2007), suggesting a role for this gene in hybrid sterility. Expression profiles for this gene were not performed as part of this study; as such, this gene should (1) be sequenced from *X. laevis* using the *S. tropicalis* ortholog and (2) qRT-PCR should be conducted on hybrids and pure species, so as to determine if it is, in fact, misexpressed in the hybrids. Should evidence of misexpression be found, knock-down studies, such as that by Young (2011), should follow.

In addition to roles in chromatin structure and transcription regulation, genes identified in this study as top candidates for miRNA regulation participate in homeostatic maintenance, structure, and cytoskeletal integrity during fertilization. For instance, *kif23* helps to facilitate the exchange of mRNA, proteins, and organelles between genetically haploid, yet phenotypically diploid, gametes by helping to regulate intercellular bridges (reviewed in Haglund et al., 2011; Hermo et al., 2010), and is key in ensuring chromosome-kinetochore connections during mitosis (UniProt, 2011). Likewise, *clasp1* is associated with sperm motility (Rieder et al., 2000), as well as drawing male and female nuclei together during fertilization (Schatten, 1983). It also aids in providing the newly formed zygote with the cytoskeletal support necessary to enable mitotic

division (Schatten, 1983). *Abcb1,* in mammals, helps to regulate, and remove xenobiotics through, the blood-testis barrier (Bart et al., 2004; Jones and Cyr, 2011; Su et al., 2009), is found in epididymal spermatozoa (Schinkel et al., 1994), and is associated with male infertility (Drozdzik et al., 2009). While lacking characterization in Anurans, *abcb1* is likely to serve an even greater role in the preservation of male fertility. This is because, unlike most vertebrates, environmental toxins are able to come into near direct contact with their testis by means of incredibly permeable abdominal skin.

In frogs, embryonic development occurs outside of the female, making the physiological constraints that have been evoked to argue an absence of male-mediated epigenetic inheritance void. As such, perhaps it should be expected that some of these genes are also associated with early embryonic processes, suggesting that testis-derived miRNA may also convey a role for male mediated epigenetic inheritance. For instance, SMARCA4 is a chromatin remodeling enzyme, which moderates changes in zygotic chromatin structure (Magnani and Cabot, 2008; Zheng et al., 2004) and is critical for genome activation (ZGA; Bultman et al., 2006). Similarity, EED, which is a member of the polycomb group methyltransferases, is thought to aid in the prevention of premature zygotic genome expression (Satijn et al., 2001); furthermore, EED contributes significantly to a number of key developmental pathways (Morey and Helin, 2010). The observation that clawed frog testis are enriched for miRNA that can regulate keystone genes such as these points strongly at the exciting possibility of paternal participation in embryogenesis, as well as provides a mechanism by which male mediated epigenetic inheritance may occur. Contrary to prior understanding, it is now know that mammalian males do transmit chromatin mediated epigenetic information to the next generation (Hammoud et al., 2009; Kimmins and Sassone-Corsi, 2005; Martins and Krawetz, 2007; Pembrey et al., 2006; Rassoulzadegan et al., 2006; Rousseaux et al., 2005, 2008), and roles for sperm-mediated RNA have been considered (Amanai et al., 2006; Boerke et al., 2007; Carrell and Hammoud, 2010), but have yet to be confirmed.

In addition to recovering a strong association between populations of miRNA that are enriched in frog testis and critical reproductive processes, this study also recovered evidence of sub/neofunctionalization in three of the genes investigated (*kdm2a, eed,* and *clasp1*), though two others may also be undergoing these processes (*kif23* and *map3k9*). Given the expected sensitivity of miRNA-mRNA targeting, this observation is entirely unexpected, as is finding that several of the predicted targets' identifiable homolog pairs display evidence of purifying selection. These observations are contrary to theory (Lynch and Force, 2000; Walsh, 1995) and what has been observed in other paleoploids (Blanc et al., 2003; Blanc and Wolffe, 2004; Bowers et al., 2003; Brunet et al., 2006; Kellis et al., 2004; Lynch and Connery, 2003; Paterson et al., 2004; Seioghe and Wolfe, 1998; Wolfe et al., 2001), though they are congruent with what has been previously noted in *X. laevis* (Hughes and Hughes, 1993).

The findings of this study suggests that, while WGD events in animals may be rare (Mallet, 2008), the rewards offered by genome expansion may be greater than the initial disadvantages associated with widespread genomic misregulation. Such thinking is reminiscent of Ohno (1970), who proposed that animal complexity is a function of multiple rounds of polyploidism, and Kimura and Ohta (1974), who concluded that “gene duplication must always precede the evolution of a gene having new function."

CHAPTER 3

MISEXPRESSION OF TESTICULAR MICRORNA IN STERILE *XENOPUS* HYBRIDS POINTS TO TETRAPOD SPECIFRIC MICRORNAS ASSOCIATED WITH MALE FERTILITY

XENOPUS LAEVIS AND *XENOPUS MUELLERI*

Sperm is now known to be responsible not only for the delivery of genetic, but also epigenetic information, to the next generation (Kimmins and Sassone-Corsi, 2005; Martins and Krawetz, 2007; Rousseaux et al., 2005, 2008). Via these modifications, heritable changes in gene expression or cellular phenotype can occur without mutation to the underlying DNA sequence (Golderberg et al. 2007). Epigenetic modifications can take place as a result of histone modifications, DNA methylation, and chromatin remodeling. In plants and animals, small noncoding RNAs have been found to guide epigenetic modifications to underlying DNA via RNAi pathways (reviewed in Tanzer et al., 2010). Recent advances in our understanding of non-coding RNAs have clearly linked them to male mediated transgenerational epigenetic inheritance (reviewed by He et al., 2009; Cuzin and Rassoulzadegan, 2010). MiRNAs (miRNA) are a class of non-coding RNAs, ~22 nt in length, found in plants, animals, viruses, and algae (Papaioannou and Nef, 2009). In clawed frogs, miRNAs are often coded by sequences located in introns of the protein coding gene that they regulate (Tang and Maxwell, 2008), though they can also be transcribed from intergenic regions, in which case their transcription is not directly coupled with the gene(s) they regulate. Although often associated with knock-down, as opposed to knock-out, miRNA can also be involved in the upregulation of the genes they affect (Vasudevan, 2007). Although it is now clear that RNA, including miRNA (Grandjean and Rassoulzadegan, 2009; Dadoune, 2009), are involved in sperm development, many aspects of the fundamental biology have yet to be elucidated. However, it is clear that somehow small

RNAs allow for the silencing of mobile elements, the imprinting of paternal genes, and DNA compaction (reviewed in Zamudio et al., 2008). Epigenetic mediated abnormalities have also been linked to male fertility (Emery and Carrell, 2006), as errors in genomic imprinting, RNA profiles, and abnormal chromatin packaging are now viewed as major contributors to subfertility and/or sterility. One such example is that mice unable to produce small RNA properly produce abnormally elongated spermatids and are consequentially infertile (Maatouk et al., 2008), while misexpression of Dicer, and thereby misexpression of miRNA, in testicular cells also results in infertility (Papaioannou et al., 2009; Papaioannou and Nef, 2010).

3.1 Materials and Methods

Total RNA was obtained from the ground testes of freshly sacrificed *X. laevis, X. muelleri* and *X. laevis:X. muelleri* hybrids using Ambion RNA extraction kits (IACUC protocol number A08.002). To obtain <40 nt RNAs, the samples were fractionated using an Ambion flashPAGE fractionator. Libraries of small RNA cDNAs were constructed and sequenced using ABI's SOLiD sequencing next generation technology (outsourced at the University of Oklahoma Medical Sciences core facility). All samples were run on a 35 nucleotide array and a GeneSifter (Geospiza, 2011) pipeline was used to identify and quantify components of the sample, including microRNAs (employing miRBase v. 15) and piRNA (employing RNAdb; Pang et al., 2005). Additionally, all reads were mapped against the *Silurana tropicalis* reference genome (Hellsten et al., 2010), the only anuran genome available thus far. Once each sample's microRNA expression profile was characterized, pair-wise analyses between *X. laevis:*hybrid and *X. muelleri:*hybrid were conducted to identify the most differentially expressed microRNA. These likelihood ratio tests were conducted on log transformed data that had been normalized against the total number of reads per sample. A list of the top 20 misexpressed microRNA was compiled for each comparison, and the lists were then condensed to reflect microRNA common to both. This new list was cross-referenced to microRNA known from mouse testes *(Mus musculus)* (Mishima et al., 2008).

To determine putative gene targets of microRNA shared between *Xenopus* and *Mus*, the EMBL-EBI Microcosm v5 (Enright, 2011) database was searched using the *Xenopus* microRNA name. If no results were returned, the precursor and mature sequences were retrieved from MirBase v16 (Griffith-Jones, 2008). These sequences were then entered into microRNAminer (Artzi et al., 2008) and *Mus* homologs were identified. The homologs were then entered into Microcosm. If no targets were identified, then the microRNA target was considered "unknown." For those microRNA whose targets were identified, the Universal Protein Resource (UniProt; Jaine et al., 2009; The UniProt Consortium, 2011) database was used to identify proteins known to be expressed in testes.

PiRNA, which are almost entirely gametically expressed were not included in analyses, as piRNA currently available in the publically accessible RNAdb (Pang et al., 2005) are from pachytene populations. Unlike pre-pachytene piRNA, which frequently map to transposable elements (Aravin et al., 2007a) and have been touted as “defenders of the germline”, these later expressed piRNAs are not TE associated, and presently remain of unknown function (reviewed in Aravin et al., 2007b).

3.2 Results

SOLiD sequencing of the small cDNA library from testes of recovered 21,738,343, 18,991,070, and 15,585,086 reads for *X. laevis*, *X. muelleri*, and hybrid samples, respectively. Only a minority of the reads mapped to known rRNAs, snRNAs, miRNAs, piRNA, intron/exon or other parts of the reference genome (Table 3.1). Despite overall low mapping of reads to reference sequences, the low number of rRNA reads clearly demonstrate that the small RNA libraries were prepared from high quality small RNA; due to the sheer abundance of rRNA in total RNA samples, a high number of rRNA hits would indicate that the original sample was degraded. This was not the case.

Global miRNA expression was reduced in the hybrid, and this massive misexpression makes the hybrid cluster outside of the group containing both *X. laevis* and *X. muelleri* in the

expression profile (Figure 3.1). Expression patterns almost universally placed *X. laevis* as having the highest expression, followed by *X. muelleri*, with hybrid expression consistently the lowest. This contrasts with microarray gene expression data in these taxa (Malone et al., 2007) where not only was the bulk of differentially expressed genes highest between *X. muelleri* and the hybrid, but that most expression differences between *X. laevis* and the hybrid revealed overexpression in hybrids. It is, however, consistent with the results of an earlier microarray study of miRNA expression in these taxa (Michalak and Malone, 2008).

Table 3.1 Reads recovered from small cDNA libraries for *X. laevis*, *X. muelleri*, and *S. tropicalis*. The low representation of rRNA and snRNA for each library shows that the quality of the reads is high, despite the low overall mapping to the reference (*S. tropicalis*) genome.

	X. laevis	***X. muelleri***	***S. tropicalis***
Total Reads	21,738,343	18,991,070	18,273,999
rRNA & snRNA	9,832	3,750	2,889
miRNA	394,550	197,914	201,082
Map to genome	2,068,742	1,429,957	2,360,902
Remainder unmapped	18,370,543	16,728,211	15,732,721

which recovered underexpression in hybrids relative to *X. laevis* (*X. muelleri* was not included). Of the top 20 miRNA misexpressed in comparisons between *X. laevis* and the hybrid and *X. muelleri* and the hybrid (Table 3.2), eight were shared between lists. These eight were then cross referenced against miRNA expressed in *Mus musculus* testes. Seven miRNA commonly misexpressed in *Xenopus* hybrids were found to have homologs in the testis of *Mus* (Table 3.3). MicroCosm results recovered a common association between these seven miRNA with genes having (1) zinc finger proteins and (2) homeobox function (Table 3.4). UniProt directly linked four of the nineteen protein coding genes identified by Microcosm to mammalian testes (mouse, rat, and human; see Table 3.5), two to closely related testes expressed protein, and two more to testes expressed associates. One of these two proteins is EED, a member of a protein

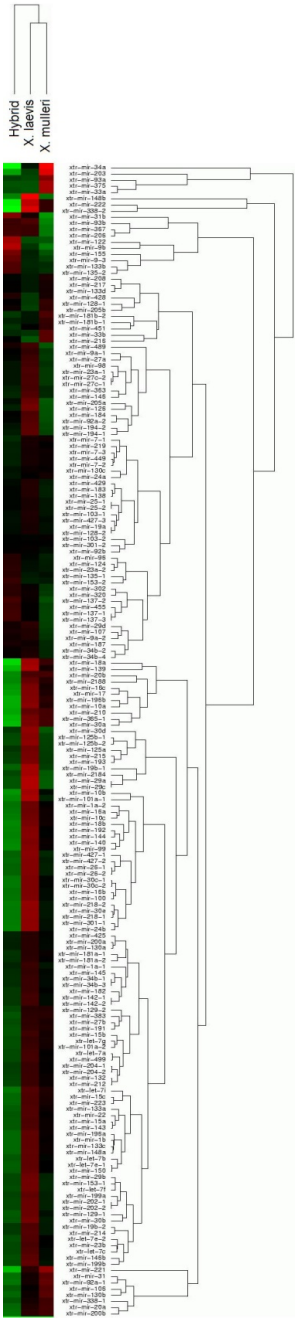

Figure 3.1 Global expression for *X. laevis, X. muelleri*, and their interspecific hybrid. The hybrid exhibits widespread underexpression of microRNA. The degree of underexpression places the hybrid outside of the cluster containing both parental species.

Table 3.2 The top 20 miRNA misexpressed in the testes of *Xenopus* comparisons.

***X. laevis*:hybrid**	**Fold Change**	**Direction**	**p-value**	***X. muelleri*:hybrid**	**Fold Change**	**Direction**	**p-value**
xtr-miR-338-2	14.34	Up	0	**xtr-miR-34a**	21.2	Up	3.36E-13
xtr-miR-222	10.52	Up	0	**xtr-miR-203**	8.53	Up	0
xtr-miR-148b	7.84	Up	0.00027	**xtr-miR-221**	6.61	Up	1.11E-12
xtr-miR-18a	7.54	Up	3.01E-07	**xtr-miR-338-2**	5.74	Up	0
xtr-miR-139	6.81	Up	9.12E-09	**xtr-miR-9b**	5.28	Dn	0
xtr-miR-30a	5.48	Up	0.00102	**xtr-miR-222**	4.92	Up	0
xtr-miR-10a	5.34	Up	0.00272	**xtr-miR-31b**	4.87	Up	0
xtr-miR-10b	5.34	Up	0.00128	**xtr-miR-18a**	4.4	Up	9.01E-05
xtr-miR-101a-1	5.24	Up	0.00605	**xtr-miR-30a**	4.12	Up	0.00604
xtr-miR-196b	5.19	Up	2.40E-11	**xtr-miR-122**	3.96	Dn	3.10E-06
xtr-miR-365-1	5.02	Up	0	**xtr-miR-375**	3.94	Up	0
xtr-miR-17	4.71	Up	0.00776	**xtr-miR-365-1**	3.83	Up	0
xtr-miR-16c	4.4	Up	0.01579	**xtr-miR-33a**	3.68	Up	1.25E-05
xtr-miR-99	4.36	Up	0.02453	**xtr-miR-210**	3.59	Up	0.02424
xtr-miR-218-2	4.35	Up	0.00603	**xtr-miR-196b**	3.33	Up	5.39E-08
xtr-miR-24b	4.3	Up	0	**xtr-miR-92a-1**	3.28	Up	0
xtr-miR-218-1	4.24	Up	0.00603	**xtr-miR-20b**	3.13	Up	0.01698
xtr-miR-192	4.19	Up	0.00212	**xtr-miR-10a**	3.1	Up	0.04259
xtr-miR-140	4.18	Up	0.01049	**xtr-miR-139**	3.06	Up	0.00015
xtr-miR-30e	4.17	Up	0.01441	**xtr-miR-155**	3.05	Dn	0

complex associated with epigenetic repression of transcription of a number of genes, including Hox-c8, by methylating 'Lys-9' and 'Lys-27' of Histone-3. It is also thought to be part of an additional complex that recruits methyltransferase, thereby linking two epigenetic repression systems. Mager et al. (2003; reviewed in Ferguson-Smith and Reik, 2003) found evidence linking EED to the maintenance of silenced alleles on paternal chromosomes. In chapter 2 of this dissertation, EED associated miRNA was recovered from multiple miRNA expression clusters of fertile non-hybrid frogs (*S. tropicalis*, *X. laevis*, *X. muelleri*), demonstrating its involvement in multiple pathways of proper testicular cell function and/or spermatogenesis. Therefore, it makes sense that miRNA associated with this gene would be misexpressed in sterile males.

3.3 Discussion

Zinc is an important structural component in different proteins associated with nucleic acid biding and/or gene regulation (reviewed in Berg, 1990), alterations to zinc-stabilized structures (zinc fingers and zinc-bridges) could result in reduced stability of protein tertiary and quaternary structure (Sakai-Kato et al., 2009). Therefore, reduced Zn function is likely to result in ineffective sperm and male infertility. The association of Zn and zinc fingers and mammalian male fertility has been known for quite some time, as protamine 2 (P2) is a zinc finger protein specific to this clade, and mammalian males with decreased protamine P2 levels are often infertile (de Yebra et al, 1998). P2 is not the only zinc finger protein found in mammalian testes: others include ZPF-95 and ZPF-96, which appear to be expressed at different times during spermatogenesis (Weissig et al., 2003), and BORIS, a male specific protein associated with

Table 3.3 The top 20 microRNA misexpressed in the testes of *Xenopus* comparisons; eight of the forty that were shared between the two comparisons, and seven of the eight that were shared between *Xenopus* and *Mus*.

***X. laevis*: hybrid**	***X. muelleri*: hybrid**	**Shared between *Xenopus***	***Xenopus: Mus musculus***
xtr-mir-338-2	xtr-miR-34a	xtr-mir-338-2	mir-338

Table 3.3--*Continued*

***X. laevis*: hybrid**	***X. muelleri*: hybrid**	**Shared between *Xenopus***	***Xenopus: Mus musculus***
xtr-mir-222	xtr-miR-203	xtr-mir-222	mir-222
xtr-mir-148b	xtr-miR-221	xtr-mir-18a	mir-18(a)
xtr-mir-18a	xtr-miR-338-2	xtr-mir-139	mir-30a(b,c,d,e)
xtr-mir-139	xtr-miR-9b	xtr-mir-30a	mir-10a(b)
xtr-mir-30a	xtr-miR-222	xtr-mir-10a	mir-196a(b)
xtr-mir-10a	xtr-miR-31b	xtr-mir-196b	
xtr-mir-10b	xtr-miR-18a	xtr-mir-365-1	
xtr-mir-101a-1	xtr-miR-30a		
xtr-mir-196b	xtr-miR-122		
xtr-mir-365-1	xtr-miR-375		
xtr-mir-17	xtr-miR-365-1		
xtr-mir-16c	xtr-miR-33a		
xtr-mir-99	xtr-miR-210		
xtr-mir-218-2	xtr-miR-196b		
xtr-mir-24b	xtr-miR-92a-1		
xtr-mir-218-1	xtr-miR-20b		
xtr-mir-192	xtr-miR-10a		
xtr-mir-140	xtr-miR-139		
xtr-mir-30e	xtr-miR-155		

Table 3.4 MiRNAs shared between misexpressed *Xenopus* and *Mus.* Species refers to the microRNA-specific taxon in Microcosm. MiR-18 did not exist as mus-miR-18 nor as xtr-miR-18.

microRNA	**Species**	**Gene**	**Description**	**P-value**
miR-338	*X. tropicalis*	*Zdhhc21*	Zinc finger DHHC domain-containing protein 21	6.31E-11
		TEgg121f24.1	Zinc phosphodiesterase ELAC protein 2	0.000108754
miR-222	*M. musculus*	*Zfp341*	Zinc finger protein 341	8.85E-05
		Zc3hc1	Zinc finger, C3HC type 1	0.000125374
miR-18	unknown	unknown	unknown	unknown
miR-30a (b,c,d,e)	*X. tropicalis*	*Hox-a1*	Hox-A1-prov protein.	7.95405e-06
		Nkx2-2	Homeobox protein NKX-3.2	8.6377e-07
		Slc30a4	Zinc transporter 4 (ZnT-4)	3.0404e-06
	M. musculus	*Msx2*	Homeobox, MSH-like 2	3.68E-07

Table 3.4--*Continued*

microRNA	Species	Gene	Description	P-value
miR-10a(b)	*X. tropicalis*	*Hox-a3*	Homeobox protein Hox-A3 (Hox-1E)	9.09E-09
		Znf668		
		Zc3hc1	Zinc finger protein 433	1.00E-06
		Zfp259	Zinc finger, C3HC type 1	1.86E-05
			Zinc finger protein 259	8.20E-05
miR-196a(b)	*X. tropicalis*	*Hox-c8*	Homeobox protein Hox-C8	5.52E-07
		unnamed	Homeobox protein Hox-A7 (Hox-1A)	8.67E-06
		Hox-d8	Homeobox protein Hox-D8 (Hox-4E)	3.32E-05
		Znf518	Zinc finger protein 518	6.61E-05
miR-365	*X. tropicalis*	*Znf618*	Zinc finger protein 618	3.01E-05
	M. musculus	*Znf644*	Zinc finger protein 644	3.20E-05
		Zbtb24	Zinc finger and BTB domain containing 24	3.56E-05

Table 3.5 UniProt report of microRNA and their protein or gene targets having testicular expression. If the microRNA does not interact directly with the protein or gene, its association is noted.

microRNA	Protein target	Class	Association	Species
miR-30	NKX-2.2	*Hox*	Closely related to NKX2-4	*Homo*
miR-30	ZnT-4	Zn	Direct interaction	*Rattus*
miR-30	*Hox-A1*	*Hox*	Closely related to *Hox-A13*	*Homo*
miR-30	MSX2	*Hox*	Direct interaction	*Homo, Mus*
miR-222	ZC3HC1	Zn	Direct interaction	*Homo*
miR-365	ZBTB24	Zn	Direct interaction	*Mus*
miR-196	HOMEZ	*Hox*	Interacts with *Hox-c8*	*Homo*
miR-196	EED	*Hox*	Interacts with *Hox-c8*	*Homo, Mus*

epigenetic reprogramming of sperm (Loukiniv et al., 2002). Zinc fingers are also important components of the male steroid hormones androgens (reviewed in Freedman, 1992) which are crucial for processes ranging from sexual differentiation (reviewed in Hughes, 2001; Sharpe,

2006), sexual maturation (Buzek and Sanborn, 1988), and spermatogenesis (Chang et al., 2003). The link between Zn and fertility of other vertebrates is not as clearly understood. However, recent demonstration of a role for Zn in fish spermatogenesis (Yamaguchi et al., 2009) suggests that Zn and zinc fingers may be ubiquitously required for vertebrate spermatogenesis to proceed normally. In this study, all six of the miRNA for which gene interactions were known were found to be associated with Zn or were zinc finger proteins; of these, three proteins have known expression in mammalian testes (see Table 3.5).

All homeobox genes contain a highly conserved 180 bp sequence. The homeobox codes for a 60 amino acid domain that bind specific DNA sequences and regulates transcription of other genes (Daboule, 1995). One set of possible targets for homeobox genes are cellular adhesion molecules (CAMs) (Jones et al. 1992). CAMs are essential for embryonic tissues, and are intimately tied to initial boundary formation in tissues, embryonic induction and migration, tissue stabilization, and regeneration (reviewed in Edelman and Crossin, 1991). Ep-CAM is a type of adhesion molecule that has been isolated in fetal and adult testes (Anderson et al., 1999), with Ep-CAM expression in adult males restricted to spermatogonia, undifferentiated germ cells present in the earliest stages of spermatogenesis. In mammals, disruption of Rhox (reproductive homeobox X-linked) genes results in reduced spermatozoa in testes and reduces motile spermatozoa in the epididymis, leading to subfertility (MacLean et al., 2005). Rhox are not alone in terms of homeodomain proteins being critical for male fertility. Sperm-1, which is expressed in germ cells following the first meiotic division and haploid spermatids, has been shown to affect fertility in mice (Pearse et al., 1997). Other known homeobox genes or proteins expressed in the male germ line include Hoxa-4, Hoxb-4, Hoxa-13, Oct, Pem, MH-3 (Hox-A4), Gtx, KIMI, LIM homeobox protein 9, EMX2, and M33 (Chromobox protein homolog 2) (Wang et al., 2000; Innis et al., 2002; Gupta, 2005; reviewed in Wilson and Davies, 2007).

Of the six miRNA for which gene interactions were known, three of them (miR-30, miR-10, miR-196) were found to be associated with nine different homeobox proteins (see Table 3.4), one of which has known expression in mammalian testes, two of which are closely related to proteins known to be expressed in testes, and another two who are associated with proteins known to be expressed in the testes (see Table 3.5). Although it is unclear how most of the miRNA and homoedomain proteins identified in this study interact, proteins closely related to Hox-A1 participate in fetal testicular development and subsequent adult fertility (Kondo et al., 1997; Hsieh-Li, 1995; Satokata, 1995; Innis et al., 2002) and proteins closely linked to Hox-C8 are involved in the epigenetic process of histone methylation (Cao and Zheng, 2004; Montgomery et al., 2005).

3.4 Conclusions

The sperm of sterile hybrids is significantly larger than the sperm of either *X. laevis* or *X. muelleri* (Malone et al., 2007). This is consistent with defective chromatin packaging in the later stages of spermatogenesis, when DNA -incorporated protamines replace histones, thus allowing for hyper-condensentation of sperm, resulting in its characteristic morphology. This process is epigenetically guided, and results from hyper-acetylation during spermiogenesis: sperm compaction and elongation occurs as a function of protamine incorporation and histone degradation (Green et al., 1994; Kistler et al., 1996; Meistrich, 1989). Many protamine or protamine-like proteins contain zinc finger domains; therefore, perturbation to the regulation of these genes or any changes that may affect Zn affinity to these domains is likely to contribute to a disruption in chromatin packaging and could cause morphological changes in sperm. Consistent with epigenetic regulation of spermatogenesis, P2 protamine content in infertile mammalian males is not necessarily associated with any mutations at the level of DNA sequence (de Yebra et al., 1998). MiRNAs, one pathway of epigenetic regulation, are thought to be critical in the regulation of gene expression at mitotic, meiotic and post-meiotic stages of spermatogenesis (He et al., 2009).

For this study, miRNA were identified that are misexpressed in sterile male hybrid *Xenopus,* and these were compared these to miRNA profiles of fertile mice. In doing so, seven miRNA were identified; these seven should be considered as potentially critical for spermatogenesis in tetrapods. Of these miRNA, two have already been linked to male fertility and/or testes. In mammals, miRNA-222 has known association with the KIT (Mast/stem cell growth factor receptor) receptor and its ligand KL (Gabbianelli et al., 2010) as well as the paralogous C-kit and its ligand SCF (stem cell factor) (He et al., 2005). These proteins participate in gametogenesis, and the KIT-KL/C-kit-SCF complexes are involved in the survival and proliferation of spermatogonia (Yoshinaga et al, 1991; Dym et al., 1995; Dirami et al., 1996), as well as in critical interactions between spermatocytes and Sertoli cells (Vincent et al., 1998). Males with mutated Kit or c-kit genes have reduced or non-existent fertility; mice with mutated Kit have a block at the pre-meiotic stages of spermatogenesis (Kissel et al., 2000) and humans with mutated c-kit experience increased apoptosis and subsequent subfertility (Sandlow et al., 1996; Feng et al., 1999). KIT proteins are not restricted to mammals; they are also present in, and critical for, spermatogenesis in Anurans (the clade containing all frogs and toads) (Raucci and Di Fiore, 2007). Although function remains unknown, miRNA-30b and miRNA-30c have both been found abundant in mouse testes, each occupying >2% of the entire cloned population (Mishima et al., 2008), while miRNA-30a has been found expressed in human testes (Liu et al., 2004). MiRNA-30 is associated with homeobox proteins and Zn transport, both of which are critical for male fertility; therefore, it may explain why testicular expression of this miRNA is conserved across tetrapods.

CHAPTER 4

REPEAT AND NON-REPEAT ASSOCAITED PIWI-LIKE (PILRNA) IN THE TESTIS OF FERTILE *X. LAEVIS, X. MUELLERI,* AND THEIR STERILE INTERSPECIFIC HYBRID *XENOPUS LAEVIS* AND *XENOPUS MUELLERI*

In mammals, three germline PIWI proteins (*Mili, Miwi2*, and *Miwi*) have been demonstrated as critical for ensuring male, but not female, fertility (Kuramochi-Miyagawa et al., 2004; Lacham-Kaplan, 2004). *Mili* and *Miwi2* are both expressed early in germline development, though both are not expressed in females. *Mili* is initially expressed in both sexes starting at ~13.5 dpc (days post coitum), and although it continues to be expressed in males, it ceases to be expressed in females shortly following birth (Aravin et al., 2007; Kuramochi-Miyagawa et al., 2001; Unhavaithaya et al., 2009). *Miwi2*, on the other hand, is only ever expressed in developing males, and can be detected at ~14.5 dpc until shortly after birth (Aravin et al., 2007; Kuramochi-Miyagawa et al., 2001). In mammals, the onset of expression of these proteins corresponds to a window of time during which the developing germline undergoes genome-wide de- and remethylation (Reik et al., 2001), thereby making the germ cells susceptible to double-stranded breaks, ectopic recombination, and disruptive insertions. During this time of germ cell genome susceptibility, PIWI-piRNA containing piRNP (piwi-ribonucleoprotein) complexes silence mobile elements via RISC and RITS like mechanisms (Aravin et al., 2007; Aravin et al., 2008; Aravin et al., 2009; Kuramochi-Miyagawa et al., 2008). Although it is now clear that two Tudor-domain containing proteins (TDRD1 and TDRD9) are required for piRNP complexes to properly function (e.g. proper cellular compartmentalization and biogenesis of piRNA) (Reuter et al. 2009; Shoji et al. 2009; Vagin et al. 2009), the exact details of these interactions are not yet fully understood (see Siomi et al., 2010). Fitting their

role in mobile element silencing, mammalian piRNA are enriched in with sequences complementary to LINES, SINES, and LTRs (Aravin et al., 2007).

As with other PIWI mutants, male *Miwi*-/- mice are sterile; these individuals arrest spermatogenesis at the round spermatid stage (Deng and Lin, 2002; Kuramochi-Miyagawa, 2004). *Miwi* interacting pachytene piRNA is present in the germline ~14 dpp (days post-partum), and reaches adult levels in a matter of days (Girard et al, 2006). Unlike *Mili/Miwi2* associated prepachytene piRNA, only ~17% of this later population of piRNA match to mobile elements or other repeats (Aravin et a., 2007a; Girard et al., 2006; Grivna et al., 2006a) and is not believed to actively participate in mobile element post-transcriptional silencing. Although not much is known about MIWI-piRNA, MIWI has been associated with polysomes in an mRNA dependent fashion, suggesting that MIWI-piRNA complexes may be involved in translation (Grivna et al., 2006b). Additionally, MIWI has been demonstrated to regulate the expression certain miRNA (Grivna et al., 2006b; Vasileva et al., 2009), suggesting an association between these later piRNA populations, MIWI, and miRNA in spermatogenesis. In *Xenopus* oocytes, Xiwi1 has been found to bind both miRNA and piRNA (Armasin et al., 2009; Leu et al., 2009), though a clear association with translation was not recovered (Lau et al., 2009).

Although first noted in mice (Aravin et al, 2006; Carmel et al., 2006; Girard et al., 2006), piRNA have now been recovered from a number of vertebrates, including rat (Lau et al., 2006), zebra fish (Houwing et al., 2007), platypus (Devor and Samollow, 2008; Devor et al., 2008; Murchison et al, 2008; Warren et al., 2008), human (Girard et al., 2006), and *Xenopus* (Armisen et al., 2009; Leu et al., 2009). To date, it has yet to be demonstrated that the PIWI pathway is critical in female vertebrate fertility (see Siomi et al., 2010; see Zamudio and Bourc'his, 2010). In non-vertebrates, however, lack of female deposited piRNA have been associated with mobilization of P and I element mediated hybrid dysgenesis in hybrid *Drosophila melanogaster* (Brennecke et al., 2008).

The male specific consequences of dysfunction in the PIWI pathway are consistent with the faster male evolution hypothesis; therefore, *Xenopus* provides an excellent system with which to investigate a possible link between the PIWI pathways and male hybrid sterility. As such, the objective of this study is not to confirm or deny the role of mobile element transposition in the germline of hybrid male *Xenopus*; but, rather, to describe some of the effects of hybridization on this class of molecule, and to provide a foundation for further studies regarding its role in reproductive isolation in vertebrates.

4.1 Materials and Methods

4.1.1 Preparation of smRNA Library and Summarization of Reads

Total RNA was obtained from the ground testes of freshly sacrificed adult *X. laevis*, *X. muelleri* and *X. laevis* x *X. muelleri* hybrids using Ambion RNA extraction kits (IACUC protocol number A08.002). To obtain <40 nt RNAs, the samples were fractionated using an Ambion flashPAGE fractionator. Libraries of small RNA cDNAs were constructed and sequenced using ABI's SOLiD sequencing next generation technology (outsourced at the University of Oklahoma Medical Sciences core facility). All samples were run on a 35 nucleotide array and a GeneSifter (Perkin Elmer, 2011) pipeline was used to identify and quantify components of the smRNA sample, including microRNAs (employing miRBase v. 15; Griffiths-Jones, 2006, 2010; Griffiths-Jones et al., 2006, 2008; Kozoma and Griffiths-Jones, 2011), piRNA (employing RNAdb; Pang et al., 2005, 2006), and known genic regions. Additionally, all remaining reads were mapped against the *S. tropicalis* reference genome (Hellsten et al., 2010), the only Xenopodinae genome available thus far.

4.1.2 Characterization and Identification of Repeat Associated smRNA

So as to identify repeat associated components of the smRNA libraries, linker sequences were stripped off of each sequence with a Ruby script, and sequences were BLASTed (Altschul et. al, 1997, 2000; Altschul and Koonin, 1998) against Xenopodinae repeats located in the most recent Repbase edition (Jurka et al., 1996; 2005). BLAST analysis was

done employing the nucleotide (blastn) and translated nucleotide (tblastx) algorithms in BLAST 3.3.25+. Group membership and element type for each recovered repeat was obtained from Repbase (*ibid*). Quantification of these RNA was conducted using both raw counts and counts which had been adjusted to account for the sizes of the corresponding smRNA libraries.

Visualization of the distribution of smRNA along the length of both a Class I retroelements and Class II DNA element was achieved by constructing a dot density figure with each of the taxa. Furthermore, strand orientation was used to infer the originating source of repeat associated smRNA. In *Drosophila* and post-natal mammals, these RNA are derived from genomic clusters; as such, these smRNA are enriched for the antisense orientation (Aravin et al., 2008). However, in prenatal mammals, primary repeat associated piRNA are produced from the mRNA of active elements. In order to infer the origin of primary repeat associated piRNA in *Xenopus*, strand orientation was employed. If these were primarily in sense orientation, it was inferred that mRNA of mobile elements are processed as primary repeat associated piRNA; conversely, if these were primarily in antisense orientation, it was inferred that piRNA clusters provide the template for primary processing. Ping-pong amplification was investigated by quantifying the number of overlapping bases between mapped piRNA. It is expected that should ping-pong amplification be occurring, the data would show a spike at ~10 nt. If this spike was recovered, sequences were checked to confirm that those which overlapped by 10 nt did, in fact, map to opposite strands.

4.1.3 Expression of Repeat Associated smRNA in Hybrid and Parental Taxa

Observed counts of repeat associated smRNA were compared using a X^2 test, and individual contribution to a significant statistic was determined by standardized scores (z-scores). Z-scores were also used to identify which upregulated or down-regulated repeats, in the hybrid relative to both parental taxa, were statistically significant (α=0.05, z cut-off ±1.96). These calculations were performed by first obtaining the absolute value of the sum of differences between the hybrid and the parental taxa. For instance, if *X. muelleri* expression

was 40 transcripts, *X. laevis* 25, and the hybrid one, the z-score was based upon |(1-25)+(1-40)|=63. The sum of expression for all remaining repeats, whose smRNA was intermediate to both parents, or less than one and equal to the other, was compared to the sum of mid-parent values (MP); it is inferred that deviation from MP indicates buffering from "genomic shock" (McClintock, 1984). Genomic shock is a phenomenon in hybrids whereby the merging of disparate genomes results in the widespread misregulation of genes, aberrant chromatin structure, and derepression of mobile elements. Despite evolution via allopolyploidization (Bürki and Fischberg, 1985; Evans, 2006; Tymowska and Fischberg, 1982), the interspecific hybrid of *X. laevis* and *X. muelleri* is homoploid (Koroma, 2011). Therefore, MP value was calculated as (XL+XM)/2. X^2 tests were performed using GraphPad's Quick Calcs online calculator (http://www.graphpad.com/quickcalcs/chisquared1.cfm). When necessary, expression data was rounded to the nearest integer. Z-scores were calculated using Systat version 12 (Systat Corporation, 2009).

4.1.4 Quantification and Target Identification of Pachytene piRNA from Adult Testis

Expression profiles for piRNA were analyzed in GeneSifter (Perkin Elmer, 2011) employing likelihood ratio tests (LRT). This test assesses the probability of observed differences in expression for a given transcript between two cDNA libraries based upon sampling variability (null hypothesis) as opposed to true heterogeneity (alternative hypothesis). This test works by comparing the two likelihoods, subtracting the log of the null from the log of the alternate hypothesis, and multiplying this difference by two: 2(LLA-LL0). Probability is determined using a X^2 with df=n-1. For statistically significant piRNA, fold change was calculated, and a list of the top misexpressed piRNA was compiled.

In order to identify putative gene targets for the piRNA, sequences were retrieved from RNAdb using the corresponding identifier, and BLASTed against the NCBI EST library for the corresponding species (mouse, rat, or human). The EST with the lowest e-value was then

BLASTed against Xenopodinae entries in the NCBI nucleotide collection, employing the tblastx algorithm.

4.1.5 Semi-Quantitative TaqMan® qRT-PCR of Repeat Associated smRNA

TaqMan® dual-labeled hydrolysis probes for putative piRNA were custom ordered from Applied Biosystems, and sequences were selected from pre-processed smRNA libraries. Sequences for which probes were constructed were selected based on two sets of selection criterion; these were (I) abundant in two libraries, but absent in the other, and did not correspond to any known repeat associated smRNA (as determined by Repbase) or (II) preliminarily associated with a Class I or Class II transposon (Table 4.1). Sequences were not aligned, and as such, the sequence of each probe was exactly as it was recovered from the library/libraries it was originally recovered from.

Table 4.1 Repeat associated *TaqMan*® smRNA probes used in this study.

Name	**Sequence**	**Len**	**Target**
133-493	CAGTCCAGTATTACTTTTTGTAACCCGA	28	hat_n2, hat_n2a, hat_10, harb_n7
58-1020	TAGGAATTCACAGATAATGAACCAATACT	29	harb_n3
44-1192	AACCATAACTTTCATTGCACAACTGCAAACCC	32	hatn12
31-1682	TAGGAATTCACAGATAATGAACCAATAC	28	hatn2_ harbn3
UNK-25	GCCGGGCGCTGTGGCGTGTGCCTGTAATCCAGCTA	35	Unknown
Gypsy	TCCAAGAAGTTTTCCTCCGCCGAGCAGAATTATG	34	Gypsy-1_I_ST
Harbinger	TCAGAGACTCTAGTCCAACACTTTCATCAGCCATG	35	Harbinger

Thirty individuals, corresponding to five hybrid females, five hybrid males, five *X. laevis* females, five *X. laevis* males, five *X. muelleri* females, and five *X. muelleri* males, were systematically assayed for each probe. Because intrinsic factors could not be corrected for,

extrinsic factors were reduced as much as possible. Prior to each assay, RNA was quality checked by spectrophotometer employing 2 uL of sample on a Nano-Drop 1000 (Thermo Scientific) using quality-control measurements of ~3.0 for 260/280 OD ratio (lower than this could indicate protein, phenol, or other contaminants) and ~1.8-3.2 for 230/260 OD ratio (lower than this could indicate co-purified contaminants). Additionally, all samples and a given probe/primer set were processed as a lot, using a single tube of master-mix for the RT reaction and a single tube for the qRT-PCR reaction. Both cDNA preparation and qRT-PCR was conducted on an ABI-7300 Real-Time PCR system (Applied Biosystems), employing the manufacturer's software for Ct calculation. This system and TaqMan® dual-labeled hydrolysis probes have been designed side-by-side, and ABI documentation states that when used in combination with the TaqMan® system, the 7300 is capable of detecting 10 starting copies of a cDNA in a 50 uL reaction with 99.7% accuracy (Applied Biosystems, 2011).

Statistical analyses were conducted on log-transformed computer-generated inverse calculations of Ct values. These inverse calculations were performed to reflect the common association of low-values as being underrepresented, and high-values as being over-represented. To perform the log transformations, the observed Ct was subtracted from 45 (the maximum number of cycles per qRT-PCR run). Samples for which there was no cDNA detection were set to 45, and after correction became zero. To allow for log-transformation, one was added to all inverse values.

For each of the seven probes, ANOVAs were run to test for differences associated with (I) "species," (II) sex, and (III) sex nested within "species". Analyses were performed in SYSTAT 13 (SYSTAT Software Inc., Chicago, IL), and graphs were constructed in both SYSTAT 13 and standard graphing software. Normalizer genes were not employed, as identification of normalizers for smRNA are difficult to find, and no normalizer has been found to be better than an incorrectly selected normalizer (Peltier and Litham, 2008), at least when employing highly sensitive TaqMan® probes. To date, there has not been a single published

study regarding the selection of normalizers for siRNA or piRNA. As such, it should not be surprising that two highly ranking journals (Molecular Cell, Impact Factor 14.447; Nature, Impact Factor 36.101) have each published papers that include qRT-PCR analysis of siRNA using (I) standard detection and an unvalidated miRNA control (Das et al., 2008) or (II) TaqMan® detection without a control (Tam et al., 2008). For the TaqMan® study published in Nature, the extent to which error was controlled for, according to their Methods section, was "Quantitative PCR was performed using TaqMan probes" (Tam et al., 2008).

4.2 Results

4.2.1 Preparation of smRNA Library and Summarization of Reads

SOLiD sequencing of the small cDNA library from testes of recovered 21,738,343, 18,991,070, and 15,585,086 reads for *X. laevis*, *X. muelleri*, and hybrid samples, respectively. From the GeneSifter pipeline, only a minority of the reads corresponded to known rRNAs, snRNAs, microRNAs, piRNA, introns/exons, or other parts of the reference genome (Table 4.2).

Table 4.2 Reads recovered from small cDNA libraries for *X. laevis*, *X. muelleri*, and *S. tropicalis*. The low representation of rRNA and snRNA for each library shows that the quality of the reads is high, despite the low overall mapping to the reference (*S. tropicalis*) genome.

	X. laevis	***X. muelleri***	***S. tropicalis***
Total Reads	21,738,343	18,991,070	18,273,999
rRNA & snRNA	9,832	3,750	2,889
miRNA	394,550	197,914	201,082
Map to genome	2,068,742	1,429,957	2,360,902
Remainder unmapped	18,370,543	16,728,211	15,732,721

Read lengths ranged from ~18-35 nt (Figure 4.1); for each, the highest density was ~28 nt. Although these data mostly conformed to those expected from descriptions of previously published piRNA libraries, the right and left tails were longer than expected given the frequently cited size of ~25-31 nt for MIWI, MILI, and MIWI2 immunoprecipitated piRNA (Aravin et al.,

2006; Aravin et al., 2007; Aravin et al., 2008; Brennecke et al., 2007; Girard et al., 2006; Grivna et al., 2006; Thompson and Lin, 2009; Vagin et al., 2006). The observed distribution of ~18-35 nt is, however, consistent with >25-33< piwi-like RNA ("pilRNA," Ro et al., 2007a) recovered from male and female reproductive tissues in juvenile and adult mice and macaque (Ro et al., 2007a,b; Ro et al., 2007; Yan et al., 2011), as well as from somatic tissues having no association with reproduction (Yan et al., 2011). Because Ro et al., (2007, 2007a,b) were the

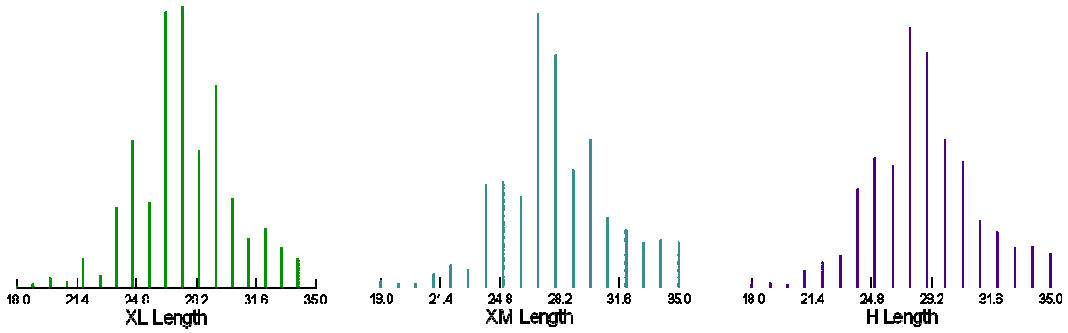

Figure 4.1 Dot histograms of sequence lengths for repeat associated smRNA.

first to describe repeat associated pilRNA, which they dubbed "rapiRNA," the repeat associated smRNA in this study will be referred to as "rapiRNA."

Although overall the proportions of the libraries that matched to known references were low, this should not be unexpected. MirBase currently contains only 208 Xenopodinae miRNA, and all piRNA in RNAdb are all mammalian. Additionally, while the genome of *S. tropicalis* has been declared finished, its current state is both unassembled and unannotated. Thus, given the subpar state of this "genome," other smRNA studies have recovered only ~55.625% match between stringently filtered *S. tropicalis* reads and its own genome (Lau et al., 2009). As is always the case, the number of mismatches between two sequences (or genomes) can only be expected to increase as distance to MRCA (Most Recent Common Ancestor) increases [~53-64 MYA between *S. tropicalis* and both *X. laevis* and *X. muelleri*, as well as an additional <21-41 MYA between *X. laevis* and *X. muelleri* (Evans et al., 2004; Chain and Evans, 2006)]. Given

the low matching potential between the *S. tropicalis* genome and smRNA libraries derived from *X. laevis*, *X. muelleri*, and their interspecies hybrid , the relative quality of the libraries needs to be assessed in other terms. As such, it is fair to state that due to the sheer abundance of rRNA in the cell, a low quality library built on degraded RNA should contain a large proportion of rRNA reads. This was not the case for any of the *Xenopus* libraries. Moreover, there was an unexpectedly high degree of homology between the reads in the *Xenopus* libraries and those obtained from the mammalian RNAdb library (4.39%-4.36%; Table 4.2). It was not expected that there would be any homology, as prior comparisons of piRNA populations in the germlines of mouse and rat [MRCA~ 16-23 MYA (Springer, 2003), as opposed to ~350 MYA for amphibians-mammals (Wang and Gu, 1999)] were reported to be very low (Lau et al., 2006), and had been treated as inconsequential.

4.2.2 Characterization and Identification of Repeat Associated smRNA

RapiRNA was primarily in the antisense orientation, though the proportions varied among all three taxa. *X. lavis* was 72.13% antisense and 27.87% sense; *X. muelleri* was 66.15% antisense and 33.85% sense; and the hybrid was 63.10% antisense and 36.90% sense. Given these, it appears that in *Xenopus,* primary rapiRNA are derived from genomic clusters. Furthermore, while the hybrid did have the highest proportion of sense strands, this alone could not be used to infer upregulation of repeat elements and activation of the Ping-Pong amplification loop in it relative to its parental taxa, as the difference between the hybrid and *X. muelleri* was less than the difference between *X. muelleri* and *X. laevis.* To infer Ping-Pong in the hybrid based upon this observation would demand that there also be recognition of Ping-Pong amplification in *X. muelleri* as compared to *X. laevis.* Along the length of a given transcript, rapiRNA was dense and overlapping in some regions, while relatively sparse in others (Figure 4.2). While the density of hits did vary between the two parental taxa, interestingly, the general locations of hits were the same; visually, the hybrid shared features with both parental taxa.

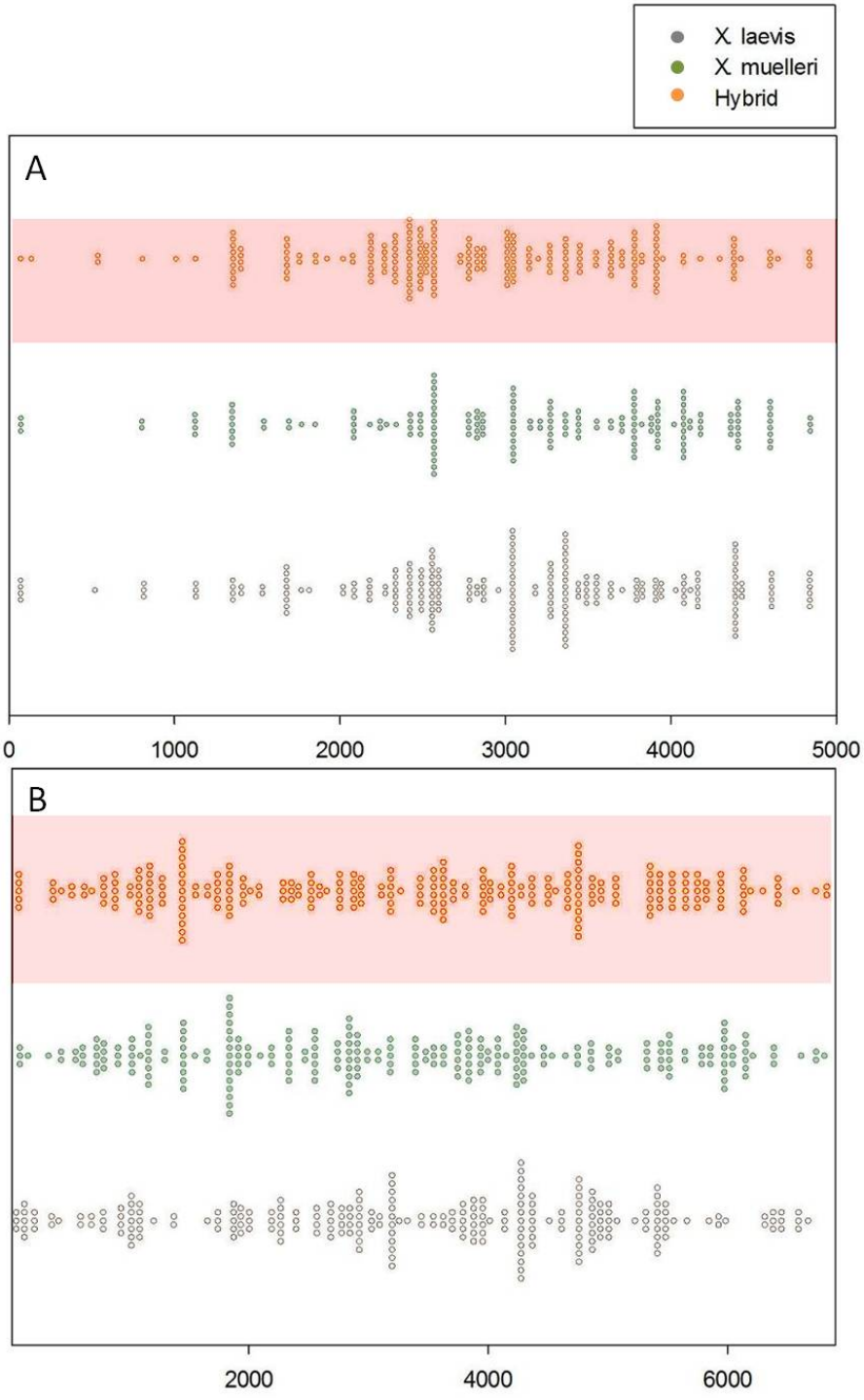

Figure 4.2 Distribution of rapiRNA along the length of (a) hat-N9_XT and (b) XEN1-I.

BLAST analysis against all 557 Xenopodinae repeat elements resulted in 5839 non-redundant rapiRNA for *X. laevis*, 5749 for *X. muelleri*, and 6832 for the hybrid (available upon request: mersee@gmail.com). From among these, 859 reads were shared between *X. laevis* and *X. muelleri*, 736 shared between *X. laevis* and the hybrid, and 657 shared between *X. muelleri* and the hybrid; 268 were common to all three (Figure 4.3). Given these, 77.27% of the reads found in the *X. laevis* library were unique to this individual, while 78.29% were unique to *X. muelleri*, and 84.53% unique to the hybrid. When the raw counts were normalized to account for differences in library size, these became 4186 (XL), 4718 (XM), and 6832 (H). It should be noted that, overall, the low number of BLAST hits is in large part a function of the fact that ~94% of Xenopodinae repeats have been described from the genome of *S. tropicalis*; a separate library from this species recovered >20,000 repeat associated smRNA (request). In total, the cDNA libraries recovered homology to 467 of the 526 currently described Xenopodinae repeat families, a fraction of each which was unique to one of the three taxa (Table 4.3). Of the 467 repeats identified in the libraries, 291 were Class I retroelements (64.31%),140 Class II DNA elements (29.98%),29 simple repeats (6.21%), and seven which have yet to be characterized (1.5%) (Figure 4.4). The high abundance of rapiRNA associated with Class II DNA elements from these libraries was different from what has been reported from mouse [no Class II were reported, though this is not to say there were not any (Aravin et al., 2007)]; however, they were consistent with previously published Xenopodinae oocyte data (Armisen et al. 2009; Lau et al., 2009). RapiRNA for simple repeats have not previously reported, though as with DNA element associated rapiRNA, this is not to say they have not been recovered.

Table 4.3 Counts of smRNA, number of repeats, and number of unique repeats.

Taxon	No. smRNA	No. Repeats	No. Unique
X. laevis	4186	359	17
X. muelleri	4718	381	25
Hybrid	6832	386	30

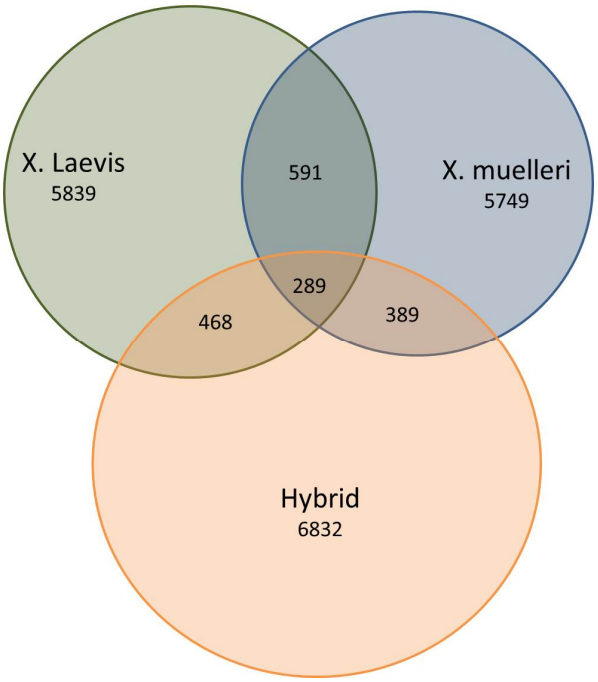

Figure 4.3 Scaled diagram displaying total counts as well as shared counts of smRNA.

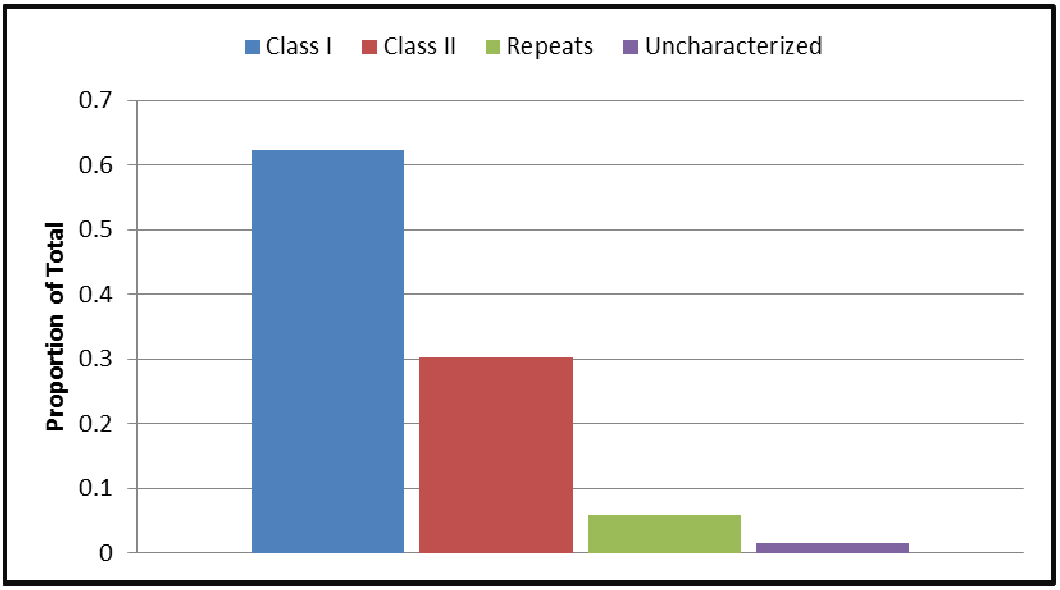

Figure 4.4 Distribution of repeat classes associated with rapiRNA. Uncharacterized repeats are lacking detailed annotation data in Repbase.

4.2.3 Expression of Repeat Associated smRNA in Hybrid and Parental Taxa

Analysis of overall rapiRNA expression among the three libraries was highly significant ($p \leq 0.0001$, X^2 test statistic 747.671, 2 df). The z-scores for these placed the hybrid as most deviant from zero (z=1.13 Hybrid, -0.307 *X. laevis*, -0.377 *X. muelleri*); as such, it can be concluded that upregulation of hybrid expression was most influential to the overall statistic. A total of 118 repeats were upregulated in the hybrid, and an additional 32 were present only in the hybrid. Z-scores for these 150 elements (obtained from the absolute value of the sum of differences between hybrid and parental species (|(H-XL)+(H-XM)|)) recovered nine with $z \geq 1.96$ (Figure 4.5 a). Of these, average fold change ranged from 1.69 (Gypsy-15_XT-I) to 4.72 (Gypsy-9-I_XT); for one (L2-5_XT), associated rapiRNA was not present in one of the parental taxa, so fold change could not be calculated. Across the upregulated and gained rapiRNA, mean hybrid expression (23.71 transcripts) was between ~43-48% higher than mean expression in the parental taxa (*X. laevis* 13.56, *X. muelleri* 12.32). Interestingly, eight of the nine upregulated or gained rapiRNA were Class I repeats (one was unclassified).

53 repeats were down-regulated in the hybrid, while another 30 were lost. Of these 83 elements, five had $z \geq 1.96$ (none were ≤ -1.96; Figure 4.5 b). Average fold change ranged from 1.83 (XL1723R) to 4.25 (Chap4a_XT); for one (Chap4sat_XT), rapiRNA was lost in the hybrid, so fold change could not be calculated. Across the down-regulated and lost rapiRNA, mean hybrid expression (6.7 transcripts) was approximately 50% of what was observed in the parental species (*X. laevis* 13.19, *X. muelleri* 13.61). Interestingly, four of the five down-regulated or lost rapiRNA were Class II repeats (one was Class I).

Quantification of 234 repeats having rapiRNA expression equal to one of the parental species and less than the other, or midway between both, revealed that overall expression did not deviate from MP, with hybrid sum of expression 2719 transcripts and average expression 11.62 and MP sum of expression 2740.5 and average expression 11.71. Despite this lack of

overall deviation, the hybrid was closer to *X. laevis* (sum 2710, average 11.58) than *X. muelleri* (sum 2771, average 11.84).

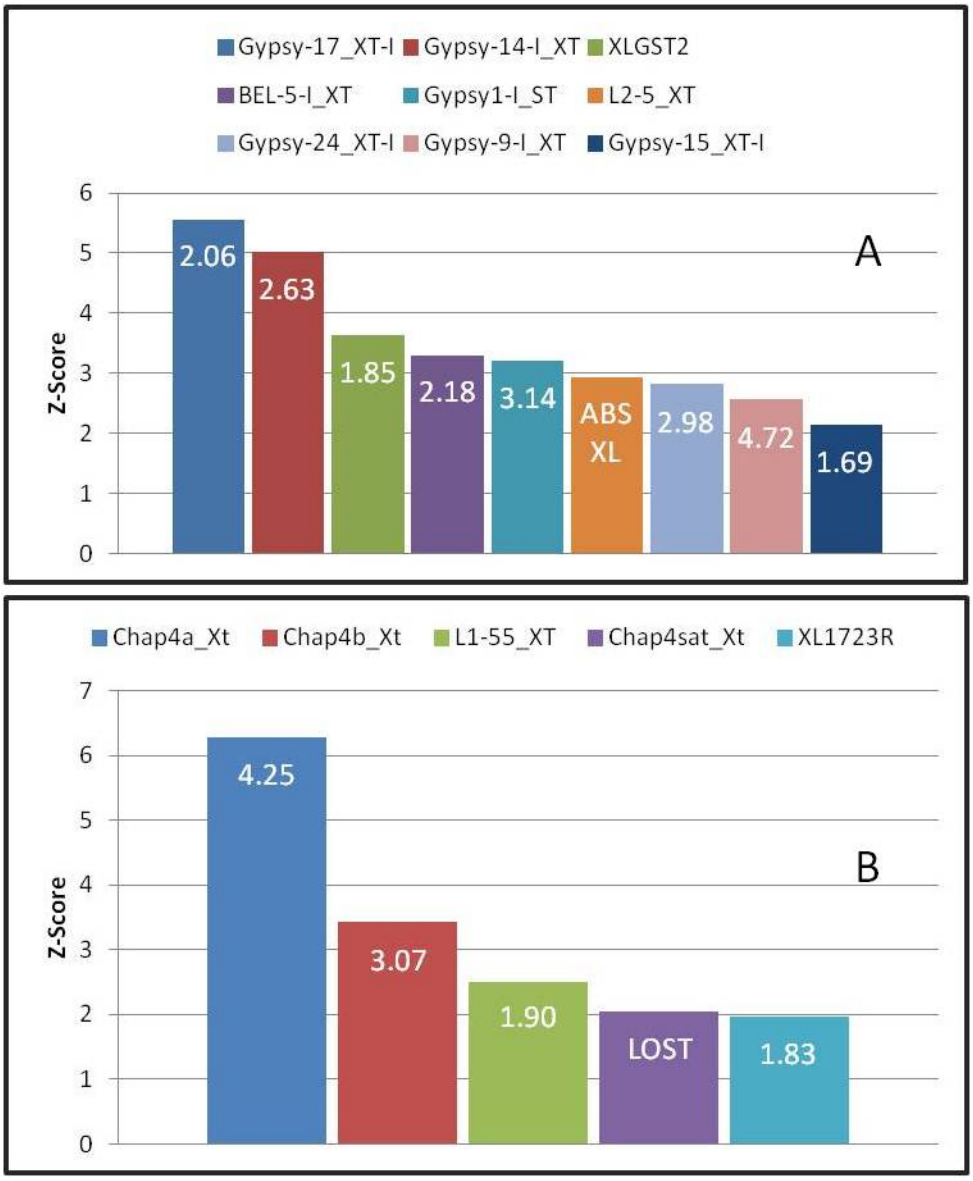

Figure 4.5 Up and down-regulated rapiRNA having z ≥ 1.96. (a) Nine upregualed repeat elements. Values inside of the bar are fold change; one was absent in *X. laevis* so fold change could not be calculated. (b) Five down-regulated repeat elements; one was lost in the hybrid so fold change could not be calculated.

4.2.4 Investigation of Ping-Pong Amplification in Upregulated rapiRNA

Differences between each of the four possible 5' and 3' ends of the nine statistically significant upregulated rapiRNA were calculated and graphed. Because the Ping-Pong amplification loop is characterized by a 10 bp overlap between the 5' ends of primary and secondary rapiRNA (Aravin et al., 2007), a spike in the distribution of pairwise differences at ~10 bp is commonly used as evidence for repeat silencing via the rapiRNA pathway. For none of the upregulated rapiRNA was a spike at ~10 bp evident (Figure 4.6); in fact, this distance was found to have the lowest density of hits for two of the nine (XLGST2 and Gypsy1-I_ST). Given these data, it does not appear that the upregulation of rapiRNA in the hybrid is a function of mobilized repeats interacting with primary rapiRNA.

4.2.5 Misregulated Pachytene pilRNA

GeneSifter enabled LRT analyses of pilRNA expression recovered 1272 pilRNA that were misexpressed between the hybrid and *X. muelleri;* of these, 554 were upregulated and 719 were down-regulated . 1205 were misexpressed relative to *X. laevis,* 532 of which were upregulated and 674 down-regulated. Of these, 346 were shared between lists; these represented seven of the top 10 misexpressed in the *X. laevis* comparison, but only one of the top 10 in the *X. muelleri* comparison. Of all whose fold change exceeded 10, only one was relative to *X. laevis* (Table 4.4).

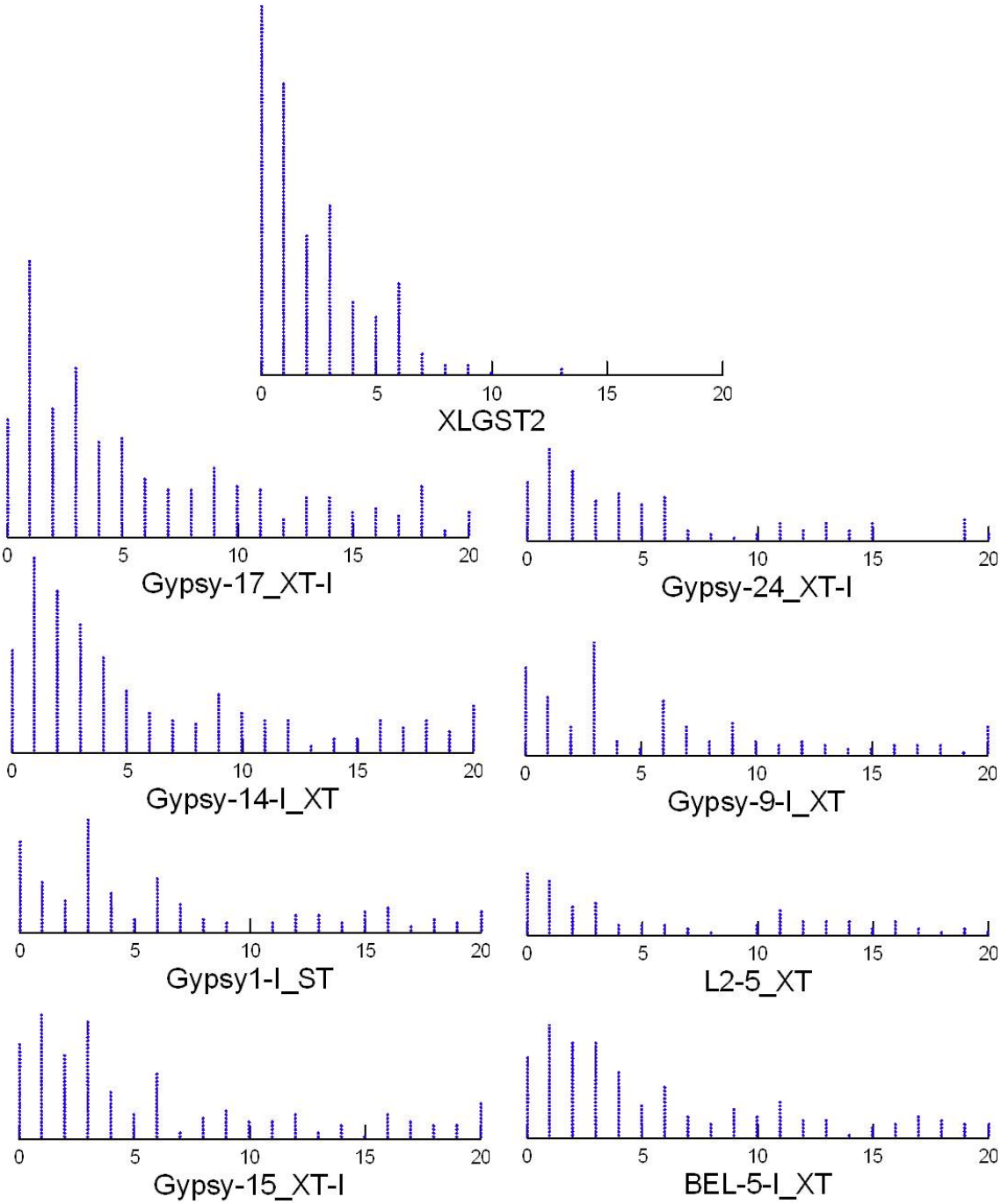

Figure 4.6 Distribution of differences between all combinations of 5' and 3' ends for each of the nine repeat elements whose rapiRNA was found to be upregulated.

Table 4.4 Top misexpressed pilRNA between hybrid and parental species. Fold change is relative to the hybrid.

RNAdb Accession	***piRNA GenBank***	***Parent***	***Fold Change***	***EST GenBank***	***Gene Name***	***Nucleotide GenBank***	***E-Value***
PIR223928	*piR130602*	*XM*	*- 455.57*	*CD550734*	*map3k13* *zswim5*	*XM_002936531.1* *NM_001091107.1*	*0.23* *0.32*
PIR243940	*piR150659*	*XM*	*+35.41*	*CK595579*	*rbbp7* *rbbp4*	*NM_203954.1* *BC088588.1*	*1e-99* *3e-99*
PIR231607	*piR138281*	*XM*	*+28.62*	*BE625944.1*	*nudt2*	*NM_001098713.1*	*0.086*
PIR36898	*piR43899*	*XM*	*+26.08*	*BU561565.1**			
PIR70687	*piR77688*	*XL*	*+22.35*	*CV126717.1*	*senp(1-7)***	*BC098074.1*	*9e-82*
PIR268604	*piR175323*	*XM*	*+19.80*	*BE113411.1*	*angpt1*	*NM_001092334.1*	*0.23*

Table 4.4 --*Continued*

RNAdb Accession	***piRNA GenBank***	***Parent***	***Fold Change***	***EST GenBank***	***Gene Name***	***Nucleotide GenBank***	***E-Value***
PIR270267	*piR176986*	*XM*	*-13.27*	*AW525913.1*	*def6*	*NM_001089966.1*	*0.23*
PIR259466	*piR166185*	*XM*	*+13.17*	*CK595036.1*	*zfyve9 SMAD Anchor*	*BC170474.1 NM_001085697.1*	*1e-102*
PIR236990	*piR143664*	*XM*	*+11.97*	*BF389343.1*	*ranbp2*	*XM_002937659.1*	*0.55*
PIR42239	*piR-49240*	*XM*	*+11.85*	*BF525121.1*	*hdac(1-5; 7,8)****	*NM_001016883.2*	*2e-149*

**This piRNA was highly transcribed (103 ESTs), and the EST hits (>300) were most all very high matches to "uncharacterized proteins." Although it did not flag Repbase as a mobile element, this behavior was suspect; as such, it was not pursued further.*

***This EST hit 7 senp genes, all with e-values <1e-07.*

****This EST hit hdac 1-5,7,8, all with e-values <9e-05*

4.2.6 Semi-Quantitative TaqMan® qRT-PCR of rapiRNA

Results for UNK-25 were non-significant in all comparisons (Figure 4.7): species (p=0.884, F-ratio 0.340); sex (p=0.416, F-ratio 0.686); and sex[species] (p=0.730, F-ratio 0.318).

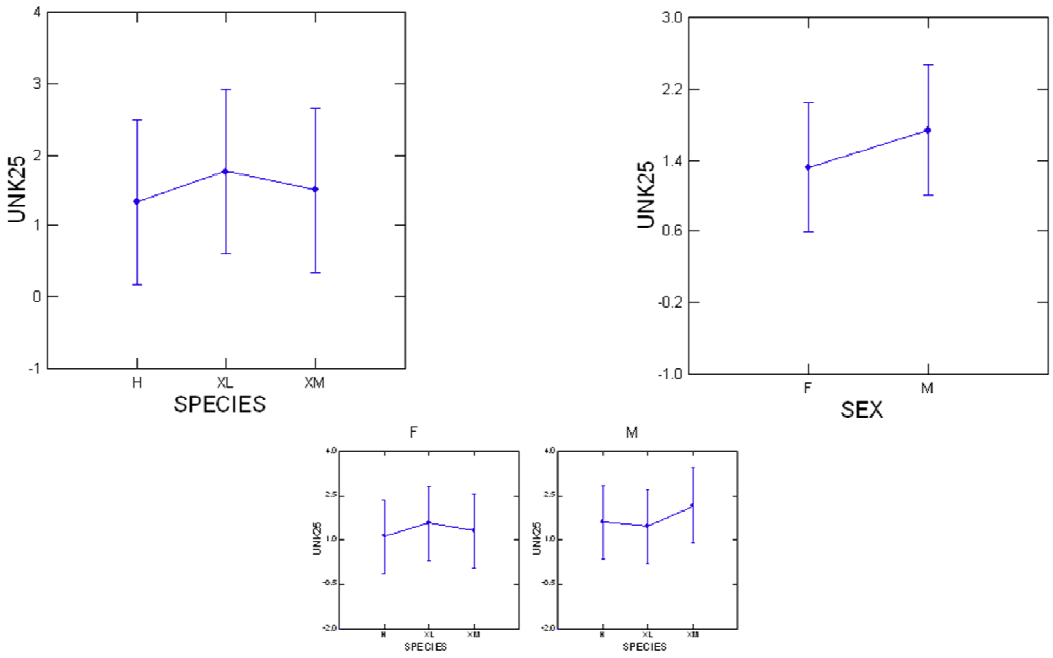

Figure 4.7 ANOVAS for TaqMan® probe UNK-25.

Results for P31-1682 were significant in two of the three comparisons (Figure 4.8): species (p=0.002, F-ratio 7.965) and sex (p=0.007, F-ratio 8.670). For species, X. laevis had highest expression and X. muelleri lowest. The post-hoc test revealed that X. laevis and hybrids differ somewhat (p=0.086), X. muelleri and hybrids differ somewhat (p=0.057), and X. laevis and X. muelleri differ significantly (p=0.002). For sex, males had higher expression than females. Sex[species] was not significant (p=0.629, F-ratio 0.473).

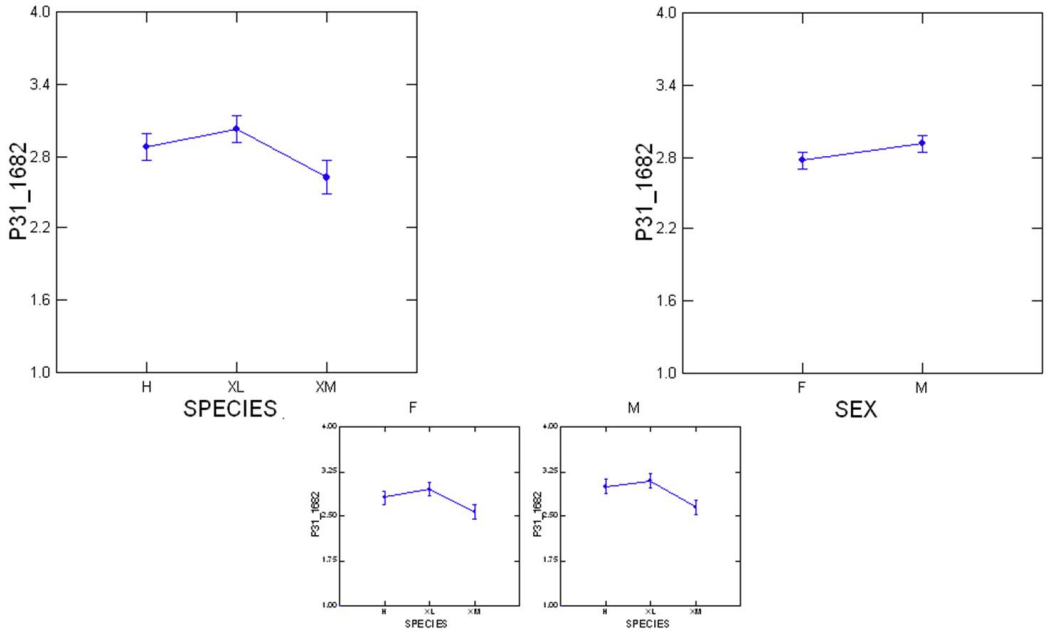

Figure 4.8 ANOVA for TaqMan® probe P31-1682.

Results for P133_493 were significant in one of the three comparisons (Figure 4.9): species (p=0.002, F-ratio of 8.014). For this comparison, hybrids had highest expression and X. muelleri the lowest. The post-hoc test revealed that X. laevis and hybrids do not differ (p=0.777), that X. muelleri and hybrids do differ (p=0.003), and that X. laevis and X. muelleri also differ (p=0.014). Both sex (p= 0.183, F-ratio 1.882) and sex[species] (p=0.837, F-ratio 0.181) were non-significant.

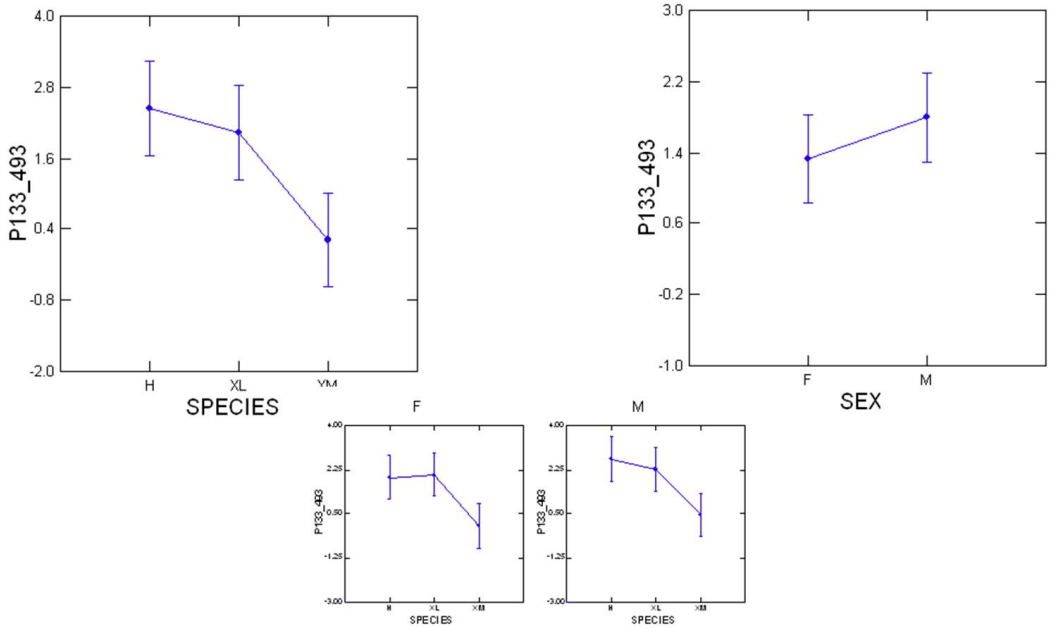

Figure 4.9 ANOVAs for TaqMan® probe P133_493.

Results for P44_1192 were non-significant in all comparisons (Figure 4.10): species (p=0.208, F-ratio 1.678); sex (p=0.145, F-ratio 2.264); and sex[species] (p=0.930, F-ratio 0.147).

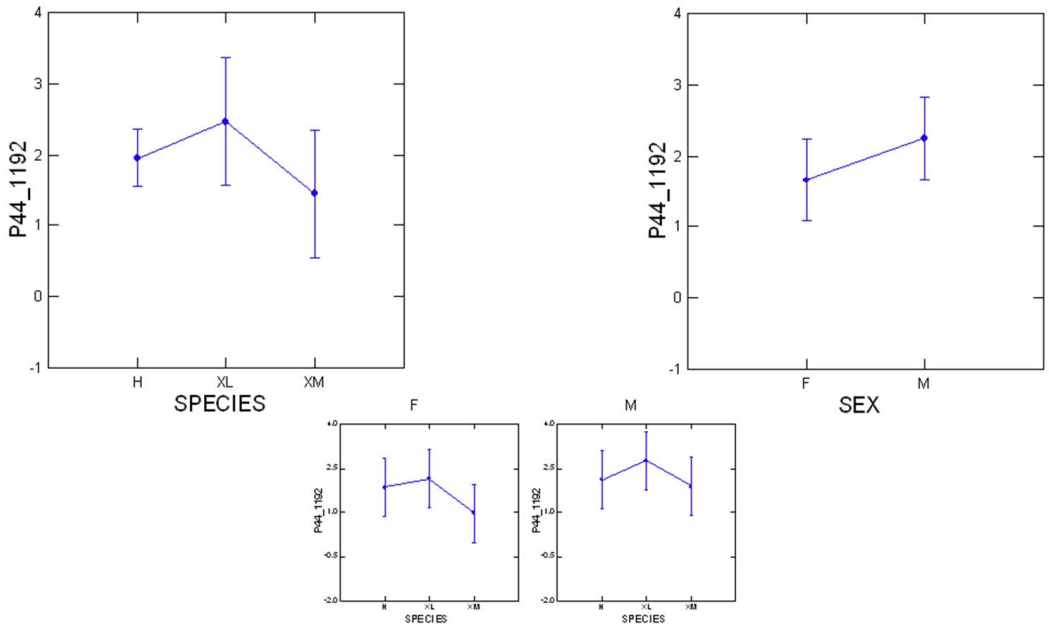

Figure 4.10 ANOVAs for TaqMan® probe P44_1192.

Results for P58_120 were significant in all comparisons (Figure 4.11): species ($p<0.001$, F-ratio 22.194), sex ($p=0.004$, F-ratio 5.388), and sex[species] ($p=0.004$, F-ratio 5.942). For species, *X. laevis* expression was highest and *X. muelleri* lowest. The post-hoc test revealed that expression profiles for none of the taxa was the same ($p<0.001$ in all comparisons). For sex, males had higher expression than females. The post-hoc test for sex[species] revealed that only four of 15 pairwise comparisons were non-significant (Table 4.6): *X. laevis* females to hybrid males ($p=1.0$); *X. laevis* females to *X. laevis* males ($p=0.999$); *X. muelleri* females to *X. muelleri* males ($p=1.0$); and hybrid males to *X. laevis* males ($p=1.0$).

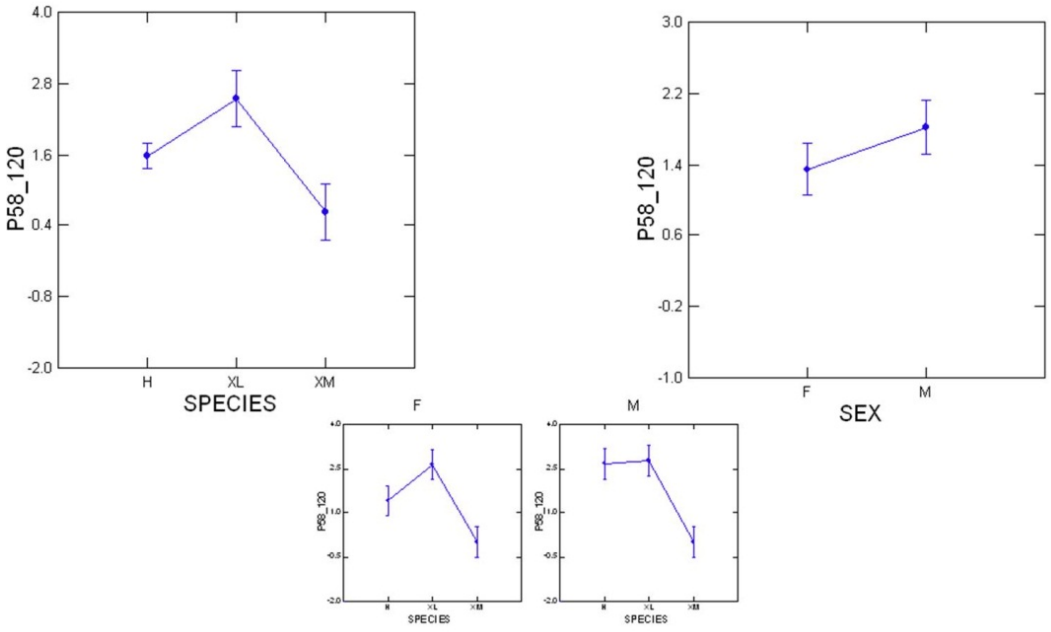

Figure 4.11 ANOVAs for TaqMan® probe P58_120.

Results for Harbinger were non-significant two of the three comparisons (Figure 4.12): species (p=0.237, F-ratio 1.527) and sex (p=0.408, F-ratio 0.709). The third comparison, sex[species], was marginally significant (p=0.058, F-ratio 3.125). The post-hoc test revealed that while the overall statistic was marginally significant, none of the pairwise comparisons was even close to significant (range p=0.269 to p=1.0).

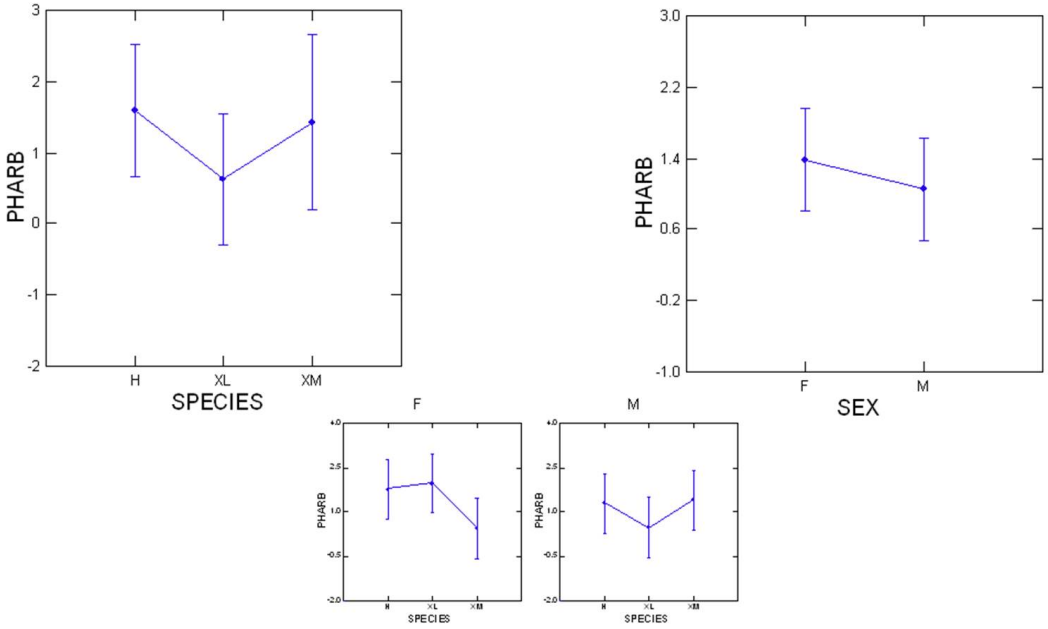

Figure 4.12 ANOVAs for TaqMan® probe Harbinger.

Results for Gypsy were significant in two of the three comparisons (Figure 4.13): species (p=0.001, F-ratio 9.399) and sex (p=0.023, F-ratio 5.921). For species, *X. laevis* had the lowest expression. Pairwise analysis revealed that *X. laevis* and hybrids differ (p=0.002) and that *X.laevis* and *X. muelleri* differ (p=0.004); however, *X. muelleri* and hybrids do not differ (p=0.155). For sex, females had less expression than males. Sex[species] was non-significant (p=0.243, F-ratio 1.501).

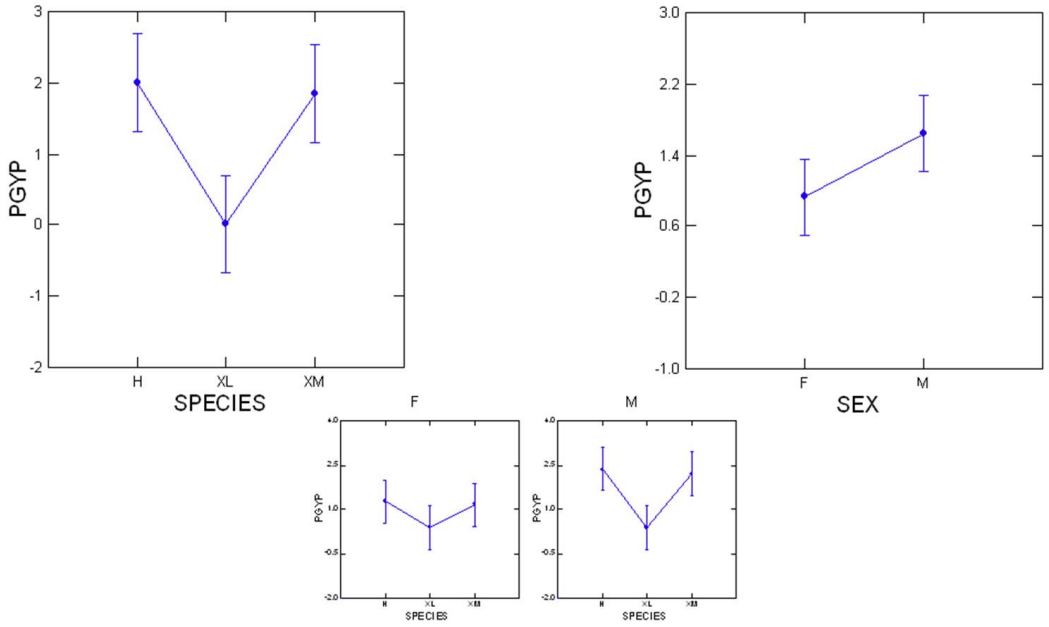

Figure 4.13 ANOVAs for TaqMan® probe Gypsy.

4.3 Discussion

Across 467 Xenopodinae specific repeats identified from rapiRNA, counts from the libraries of *X. laevis, X. muelleri,* and their interspecific hybrid ranged from 5749 to 6832; this low number of recovered mobile element hits was specific to *Xenopus,* as a separate library from *S. tropicalis* resulted in >20,000 hits. The low recovery of mobile elements in *Xenopus* relative to *Silurana* may be explained by differential radiation of mobile elements along these lineages coupled with a high mutation rate; the result of which is low sequence identity and very low power in BLAST analyses. Differential expansion and subsequent evolution of mobile elements has been found in *Gossypium* (Hawkins et al., 2006), as well as in *Hellianthus* (Ungerer et al., 2009). Molecular evolution of mobile elements in hybrids can be incredibly rapid; for instance, those from the genome of *Hellianthus deserticola*, a plant of hybrid origin dating 170,000-63000 years (Gross et al., 2001), have been shown to have an accelerated rate of molecular evolution, with pairwise divergences reaching ≤59.9% between it, its parental taxa, and other hybrid taxa originating from the same parents.

Expression of rapiRNA was also found to differ significantly among *X. laevis, X. muelleri,* and their interspecific hybrid ($p \leq 0.0001$; available upon request: mersee@gmail.com). Analysis showed that among these, the hybrid expressed the most rapiRNA, while *X. muelleri* expressed the fewest (Figure 4.3). Among all rapiRNA recovered, 30 were lost in the hybrid, while another 32 were gained. Interestingly, eight of the nine statistically significant upregulated or gained elements were Class I retroelements, while four of the five statistically significant down-regulated or lost elements were Class II DNA transposons. This class of element is over-represented in Xenopodinae genomes, where they account for ≤72% of all mobile elements (Hellsten et al. 2010); consequentially, they are highly expressed in rapiRNA libraries (Armisen et al., 2009; Lau et al., 2009). While the data presented here does add to our current understanding in the dynamics of Class II associated rapiRNA, this class of molecule remains elusive. In addition to their unique response to genomic shock, it has previously been noted

that the mechanism they employ to differentiate non-transposon mRNA from rapiRNA is not understood (Lau et al., 2009).

Despite an upregulation of rapiRNA recovered in the hybrid small RNA library, as well as an increase in the number of reads having sense orientation, this study did not recover evidence of an active Ping Pong Amplification Cycle, as there was no spike present at an ~10bp overlap between transcripts recovered for a given repeat element (Figure 4.6). Without evidence of ping pong amplification, the recovered increase in sense orientation strands is baffling, and at present time, there are no data available with which to infer the meaning of this observation. However, given that the upregulation of rapiRNA is independent of self-amplification, its upregulation suggests that the PIWI pathway is robust to the effects of genomic shock. Other smRNA have also been found to behave in a surprisingly stable manner, which has led to the conclusion that some classes of smRNA may actually serve as buffer to this phenomena (Ha et al., 2009). Interestingly, according to this same study, siRNA, which also serve in an RNAi silencing pathway of mobile elements, appear to not respond as robustly to genomic shock as do miRNA. The differences between siRNA and piRNA, therefore, beg the question as to whether this difference is a function of evolutionary constraints specific to the germline.

Xenopodinae females are abundant in rapiRNA, and their germlines also contain smRNA corresponding to the size class of siRNA (Armisen et al., 2009; Lau et al., 2009; Wilczynska et al., 2009). As such, future studies employing the *Xenopus* system should be able to discern if PIWI mediated RNAi is inherently more robust to hybridization than the siRNA induced RNAi pathway. A study of this sort would also highlight specific roles for these molecules in the female germline, as this study recovered consistently lower expression of (ra)piRNA in females as opposed to males (Figures 4.7-4.13), and prior studies have found there to be a great deal of variation between oocytes from the same female (Lau et al., 2009). Together, these data suggest that the PIWI pathway may not be the primary route to mobile

element silencing in female frogs. In mice, whereby females also have both rapiRNA and siRNA, siRNA have been suggested as the primary route for this process (Tam et al., 2008; Watanabe et al., 2008). This may also be the case for clawed frogs. Alternatively, the PIWI pathway in frogs may be unnecessary.

4.3.1 Genome Properties of Clawed frogs *May Render the PIWI Pathway Obsolete*

At present, the PIWI pathway is best characterized in *Drosophila* and mammals. Given specific properties of the genomes of these organisms, it is possible that metazoan reliance on rapiRNA silencing of mobile elements has been overestimated. In *Drosophila,* where retrotransposons comprise ~80% of all mobile elements found in the genome (Feschotte and Pritham, 2007), genomes are highly susceptible to further mobile element insertions, with these responsible for up to 50-85% of all spontaneous mutations (see Yoder et al., 1997). The strong significance of mobile elements in *Drosophila* genome dynamics is a result of the fact that their genomes are not methylated (Bird and Taggert, 1980), and that their heterochromatin does not demand transcriptional silencing (see Yasuhara et al., 2005). Given these genomic properties, piRNA (or similar molecules) are necessary to ensure genomic integrity and persistence of the species. The same is true for mammals, where mobile elements in the genome can be up to ~90% retroemelents (Feschotte and Pritham, 2007), and although these genomes are highly methylated (Bird and Taggert, 1980), their germlines undergo global demethylation early in development (see Reik et al., 2001), such that by the time primordial germ cells enter the fetal gonads, imprints and other epigenetic marks have been removed. It is during this time that the PIWI pathway is active in silencing mobile elements in the male germline (Aravin et al., 2007).

Unlike flies and mammals, the genomes of Clawed frogs are comprised of <28% retroransposons (Hellsten et al. 2010). Furthermore, while their genomes are methylated (Koroma et al., 2011), they do not undergo global demethylation (Stancheva and Meehan, 2000; Stancheva et al., 2001; Stancheva et al., 2002), though they do have a period of reduced methyation beginning just prior to developmental stage 12.5. Concurrent with the reduction in

methylation, the protein Vasa increases in expression (Ikenishi and Tanka, 1997, 2001); this protein is part of the PIWI pathway, where it is a critical component of piRISC and the Ping-Pong Amplification Cycle (Malone et al., 2009) that mediates cytoplasmic destruction of mobile element mRNA. While precipitation of Vasa prior to stage ~12.5 leads to noticeably reduced germ plasm by developmental stage 37/38, and a reduced (or lack of) primordial germ cells by stage 44 (Ikenishi and Tanka, 1997), there are several key issues that indicate the PIWI pathway, as currently understood, does not behave in Clawed frogs the same as it behaves in mice and flies. First, the loss of germ cells is not male specific, as is the case with mice. Second, the timing of events is off. In *Xenopus,* by developmental stage 19, zygotic methyltransferase is upregulated, and between stages ~23-36, remethylation takes place (Stancheva et al., 2002). Therefore, by the time primordial germ cells differentiate between stages ~44-53 (Villalpando and Merchant-Larios, 1990), the genome is intact and mobile elements should no longer be a threat. Furthermore, larvae whose genomes continue to be undermethylated during later stages of development undergo widespread apoptosis and subsequent deformation (Stancheva and Meeha, 2000). Therefore, it is very likely that the mobile element silencing component of the PIWI pathway is not critical for ensuring reproductive potential in frogs. This, however, is not to say that other aspects of the PIWI pathway, such as those associated with the presently not understood role of pachytene piRNA, are not critical for fertility.

4.3.2 Misregulated Pachytene pilRNA and the PIWI Pathway: Male Hybrid Sterility

Of the genes found to be likely associates of the top misregulated pilRNA, several have defined roles in reproduction, while others are known to be expressed in reproductive tissues. Franco et al. (2005) found *Rbbp7* to be misexpressed in mice having telomere dysfunction, and proposed that this gene is involved in epigenetic alterations of telomeric heterochromatin. Combined with other misexpressed genes, the net result was testicular atrophy and impaired fertility. *Rbbp7* has also been found expressed in Sertoli cells, where their role has yet to be

characterized (Griswold, 1998). These cells are responsible for primordial germ cell capture, triggering testis development, and also later ensure the successful maturation of germ cells into spermatozoa (Griswold, 1998). *Senp* proteins are highly expressed in testis, where they catalyze various aspects of the Sumo pathway (Xu and Au, 2005). This pathway is involved in the post-translational modification of lysine residues on proteins, and plays a crucial role in a number of cellular processes such as nuclear transport, DNA replication and repair, mitosis and signal transduction (Apweiler et al., 2004; Bairoch et al., 2005; Wu et al., 2006). Furthermore, *Senp1* is associated with transcription, as it not only deconjugates *Sumo1* from *Hdac1*, thereby decreasing its transcriptional repression activity (Cheng et al., 2005), but also interacts with JUN (Apweiler et al., 2004; Bairoch et al., 2005; Wu et al., 2006). This protein is part of the JNK pathway, and is associated with increases the activity of transcription in key genes associated maintaining homeostasis of junction-associated proteins and proteases which are critical for maintaining the blood-testis barrier (BTB) (Wong et al., 2005). Perturbations to gene expression will result in degradation of cellular junctions within the seminiferous epithelium of Sertoli cells, affecting both Sertoli-Sertoli and Sertoli-germ cell interactions (Wong et al., 2005); the result of this is cellular damage and reduced sperm. Interestingly, *Map3K13*, a member of the JUN mediated JNK pathway and Hhac1-5, 7, 8 were identified in this study as other putative targets of misregulated piRNA.

4.3.2.1 Chromatoid Bodies: Linking piRNA, miRNA, and Gene/Protein Regulation

Chromatoid bodies are a type of cytoplasmic germ body rich in ribonuclear proteins, the collection of which are called “nuage,” that aggregate in the juxtanuclear area of the pachytene spermatocytes post-meiotically (Fawcett et al., 1970). These structures are enriched in RNA as well as RNA-binding proteins, and are linked to post-transcriptional gene regulation (Kotaja et al., 2006). Only recently, are the extent to which chromatoid bodies are involved in male reproduction becoming clear (reviewed in Meikar et al., 2011).

Cytoplasmic P-bodies are inversely associated with ribosome activity, and have been known for a while to be associated with post-transcriptional modifications of mRNA, thereby serving in the post-transcriptional regulation of protein production (reviewed in Parker and Sheth, 2007). Haraguchi et al. (2005) showed that CBs are the site of degradation for unnecessary mRNA and proteins, while Grivna et al. (2006) and Kotaja et al. (2006) linked the PIWI pathway to chromatoid bodies and these regulatory functions. Furthermore, Grivna et al. (2006) were the first to associate piRNA, and components of the PIWI pathway, to miRNA regulation/expression. Vasileva et al. (2009) further established a connection between chromatoid bodies, the PIWI pathway, and processing/regulation of a subset of miRNA. Their research found that TDRD6-/- mutants not only fail in the aggregation of *Mael, Mvh* (murine *Vasa* homolog), and *Miwi*, but also that a subset of miRNA become misexpressed.

Given these previous findings, it may not come as a surprise that genes identified as putative miRNA targets in Chapter 2 of this dissertation, namely *map3k9 kdm2a*, are either part of the same pathways, or have similar function to, genes identified as putative piRNA targets (MAP3K9 and MAP3K13 both interact with JUN in JNK initiation of transcription activation/upregulation, and KDM2A and the HDAC genes are code for transcriptional regulators that act via histone deacetylation). However, it is remarkable that some of the top miRNA identified as being misregulated in PIWI mutants (Vasileva et al., 2009) are among the top miRNA found in Chapter 3 of this dissertation to be misexpressed in the hybrid relative to parental taxa (miR-34a, miR-30a, miR-148, miR-17, miR-18, miR-203, miR-20). Likewise, it is remarkable that many of the top misexpressed piRNA in the hybrid relative to parental taxa (primarily *X. muelleri;* Table 4.4) are associated the Sumo pathway (*senp1-7, angpt1, ranbp2*), which facilitates protein-protein interactions (Melchior, 2000). For instance, in *Xenopus,* SUMOlatyion facilitates the interactions of RANGAP1 and RANBP2 (another protein whose gene identified by the pilRNA) (see Saitoh et al., 1007); RANGAP1 activates the GTPase RAN, which in turn is involved in the nuclear-cytoplasmic transport of mRNA, mRNA processing, cell

cycle regulation, and other aspects of nuclear function. *RanGap1* mislocalization has, in *Drosophila,* been attributed to segregation distortion of chromosomes in the testis during meiosis, therefore leading to meiotic-drive (Kusano et al., 2000). Moreover, *ranbp2* has been associated with male infertility in humans (Koide et al., 2000) and mice (Nagai et al., 2011).

An additional subset of the genes identified as likely targets of the misexpressed piLRNA are associated with *Mael,* the Sin3/Hdac complex (*rbbp4, rbbp7, hdac1/2*), and a SWI/SNF chromatin modifier in the meiotic silencing of unsynapsed chromatin (MSUC) (Costa et al., 2006). Furthermore, Costa et al. (2006) proposed that chromatoid bodies shuttle *Mael* between the nucleus and cytoplasm, in the process transporting miRNA out of the nucleus and facilitating its coupling with MAEL, where together they silence unsynapsed chromosomes. While SUMOlation/Acetylation are two mutually exclusive post-translational histone modifications, it has been shown that a switch between their pathways pathways exists, and that this is initiated by SIRT1 and HIC1 acting in conjunction with HDAC4, the results of which determine SUMOlation (initiation of transcription) or deacetylation (suppression of translation) of the transcription repressor MTA1 (Van Rechem et al., 2010).

4.3.3 Physiological Sex May Not Be All That Matters: Male Hybrid Sterility

Recently, it has been noted that in clawed frogs*, mael, vasa* and *xiwi* are differentially expressed in embryos and early tadpoles which have been exposed to feminizing (up-regulates genes) vs. masculinizing (down-regulates genes) chemical agents (Zhang et al., 2010). Previously, it was noted that physiological sex, irregardless of genetic sex, was the determining factor in F1 male hybrid sterility (Malone and Michalak, 2008). However, given the possible connections between the PIWI pathway and male fertility, as well as the clear association of the PIWI pathway in sex determination, it is plausible that this earlier conclusion was incorrect; up-regulation of *mael, vasa*, and *xiwi*, as well as possibly other components of the PIWI pathway, are likely to have rescued the sterile phenotype associated with genetically male hybrids, while down-regulation of this same pathway may be responsible for genetic females having become

sterile males. This does not negate the observation that this system supports the faster male hypothesis (Wu and Davis, 1993; Wu et al., 1996); however, until it can be demonstrated that there was not rescue, or that components of the rescuing pathway are not located on sex chromosomes, it temporarily returns genetic sex as a possible factor to explain the observed phenenomenon.

4.4 Conclusions

This study recovered a robust response of repeat associated piRNA to the perturbations of genomic shock. However, despite the observed upregualtion of rapiRNA in hybrid *Xenopus,* the timing of gonad differentiation and germline susceptibility to mobile element translocations do not appear to coincide. As such, the mobile element silencing aspects of the PIWI pathway may not be applicable to Xenopodinae frogs, and mobile element translocations do not offer a good explanation for the observed sterility of hybrid male *Xenopus.*

While this study did not recover a link between rapiRNA and genomic shock, this study did recover a link between pilRNA misexpressed in sterile hybrids and transcriptional regulation, as well as overlap in the types of genes/pathways associated with miRNA that has been found to be conserved across Xenopodinae. Furthermore, this study highlights the first evidence to ever have been recovered which links the pachytene portion of the PIWI pathway to hybrid sterility. While not definitive nor conclusive, the data presented in this dissertation are compelling evidence of strong associations between miRNA, pilRNA, and proteins that are critical for ensuring male reproductive potential. As such, their misexpression, perhaps mediated by misexpression of their associated smRNA, is likely to lead to infertility via a number of different pathways. While presently the spermatogenic pathways that link miRNA, piRNA, *Miwi, Mael, Mhv*, and *Tdrd6* are not very well understood, the data from this study demonstrate that these may have a role in hybrid sterility, the result of which is an incredibly exciting hypothesis that warrants further investigation.

APPENDIX A

DAVID/MIRAT! CLUSTERS

DAVID functional clusters of Mir-AT! identified genes. PAM cluster A.

Annotation Cluster 1	**Enrichment Score: 1.9956203768075136**				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
INTERPRO	IPR018200:Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site	4	3.418803419	0.003401341	74841, 78787, 170707, 170822
INTERPRO	IPR001394:Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2	4	3.418803419	0.004179375	74841, 78787, 170707, 170822
GOTERM_MF_FAT	GO:0004221~ubiquitin thiolesterase activity	4	3.418803419	0.00722614	74841, 78787, 170707, 170822
GOTERM_MF_FAT	GO:0016790~thiolester hydrolase activity	4	3.418803419	0.022310528	74841, 78787, 170707, 170822
GOTERM_BP_FAT	GO:0006511~ubiquitin-dependent protein catabolic process	4	3.418803419	0.045890252	74841, 78787, 170707, 170822
Annotation Cluster 2	**Enrichment Score: 1.906215239421441**				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SP_PIR_KEYWORDS	chromatin regulator	6	5.128205128	0.005320509	13626, 15182, 20586, 52808, 13831, 225876
GOTERM_BP_FAT	GO:0016568~chromatin modification	6	5.128205128	0.010865692	13626, 15182, 20586, 52808, 13831, 225876
GOTERM_BP_FAT	GO:0006325~chromatin organization	6	5.128205128	0.033063109	13626, 15182, 20586, 52808, 13831, 225876
Annotation Cluster 3	**Enrichment Score: 1.4355838275952029**				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0070013~intracellular organelle lumen	12	10.25641026	0.033881225	27055, 13006, 18181, 13626, 15182, 20586, 18045, 67619, 237221, 13831, 21422, 225876

GOTERM_CC_FAT	GO:0043233~organelle lumen	12	10.25641026	0.034451006	27055, 13006, 18181, 13626, 15182, 20586, 18045, 67619, 237221, 13831, 21422, 225876
GOTERM_CC_FAT	GO:0031974~membrane-enclosed lumen	12	10.25641026	0.042275378	27055, 13006, 18181, 13626, 15182, 20586, 18045, 67619, 237221, 13831, 21422, 225876
Annotation Cluster 4	**Enrichment Score: 1.3452109506562235**				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	calcium-binding region:2	4	3.418803419	0.02616506	12626, 18799, 27055, 13858
UP_SEQ_FEATURE	calcium-binding region:1	4	3.418803419	0.032678561	12626, 18799, 27055, 13858
INTERPRO	IPR002048:Calcium-binding EF-hand	4	3.418803419	0.043304877	12626, 18799, 27055, 13858
SMART	SM00054:EFh	4	3.418803419	0.051493094	12626, 18799, 27055, 13858
Annotation Cluster 5	**Enrichment Score: 0.9928684287134554**				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0000278~mitotic cell cycle	5	4.273504274	0.049665912	12626, 13006, 70385, 53892, 16765

DAVID functional clusters of Mir-AT! identified genes. PAM cluster B.

Annotation Cluster 1	***Enrichment Score: 3.4547822514947213***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:WD 18	4	0.454545455	3.51E-04	21745, 106618, 319670, 237711
UP_SEQ_FEATURE	repeat:WD 19	4	0.454545455	3.51E-04	21745, 106618, 319670, 237711
UP_SEQ_FEATURE	repeat:WD 21	4	0.454545455	3.51E-04	21745, 106618, 319670, 237711
UP_SEQ_FEATURE	repeat:WD 20	4	0.454545455	3.51E-04	21745, 106618, 319670, 237711

Annotation Cluster 2	***Enrichment Score: 2.85264906068663***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
PIR_SUPERFAMILY	PIRSF002258:collagen alpha 1(IV) chain	4	0.454545455	0.001154495	12830, 94216, 12826, 12829
INTERPRO	IPR001442:Type 4 procollagen, C-terminal repeat	4	0.454545455	0.001383695	12830, 94216, 12826, 12829
GOTERM_CC_FAT	GO:0005587~collagen type IV	4	0.454545455	0.001444908	12830, 94216, 12826, 12829
GOTERM_CC_FAT	GO:0030935~sheet-forming collagen	4	0.454545455	0.001444908	12830, 94216, 12826, 12829
SMART	SM00111:C4	4	0.454545455	0.001635467	12830, 94216, 12826, 12829
Annotation Cluster 3	***Enrichment Score: 2.6423782001858243***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:WD 17	5	0.568181818	1.96E-05	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 16	5	0.568181818	5.67E-05	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 15	5	0.568181818	1.28E-04	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 14	5	0.568181818	0.001506994	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 13	5	0.568181818	0.009182517	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 12	5	0.568181818	0.013205653	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 10	5	0.568181818	0.031332256	21745, 106618, 319670, 237711, 22138
UP_SEQ_FEATURE	repeat:WD 11	5	0.568181818	0.031332256	21745, 106618, 319670, 237711, 22138

Annotation Cluster 4	***Enrichment Score: 2.007414548715792***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
INTERPRO	IPR012680:Laminin G, subdomain 2	7	0.795454545	0.005651483	14456, 170571, 53867, 71355, 12814, 18191, 12819
INTERPRO	IPR001791:Laminin G	7	0.795454545	0.011390866	14456, 170571, 53867, 71355, 12814, 18191, 12819
SMART	SM00282:LamG	7	0.795454545	0.014758328	14456, 170571, 53867, 71355, 12814, 18191, 12819
Annotation Cluster 5	***Enrichment Score: 1.6735769773383498***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0050892~intestinal absorption	4	0.454545455	0.005204239	11428, 69060, 27409, 64602
GOTERM_BP_FAT	GO:0022600~digestive system process	4	0.454545455	0.01926863	11428, 69060, 27409, 64602
Annotation Cluster 6	***Enrichment Score: 1.5570159433549975***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0016702~oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	8	0.909090909	0.024162967	18822, 69113, 214133, 16922, 56401, 107817, 216850, 225876
GOTERM_MF_FAT	GO:0016701~oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	8	0.909090909	0.025949966	18822, 69113, 214133, 16922, 56401, 107817, 216850, 225876
SP_PIR_KEYWORDS	Dioxygenase	8	0.909090909	0.03401458	18822, 69113, 214133, 16922, 56401, 107817, 216850, 225876

Annotation Cluster 7	***Enrichment Score: 1.4414458704687934***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
PIR_SUPERFAMILY	PIRSF037037:Kelch-like_protein_gigaxonin	6	0.681818182	0.024986841	209239, 228003, 66689, 212390, 224023, 207952
INTERPRO	IPR017096:Kelch-like protein, gigaxonin	6	0.681818182	0.031623228	209239, 228003, 66689, 212390, 224023, 207952
INTERPRO	IPR011705:BTB/Kelch-associated	6	0.681818182	0.05997149	209239, 228003, 66689, 212390, 224023, 207952
Annotation Cluster 8	***Enrichment Score: 1.4356178504836075***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0000289~nuclear-transcribed mRNA poly(A) tail shortening	3	0.340909091	0.010098454	103135, 22695, 17350
GOTERM_BP_FAT	GO:0000288~nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	3	0.340909091	0.042114767	103135, 22695, 17350
Annotation Cluster 9	***Enrichment Score: 1.3940587621225018***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	80	9.090909091	0.034555097	108143, 56248, 71819, 270685, 17685, 22375, 170759, 76858, 26426, 192292, 270163, 246729, 11947, 12566, 22138, 226539, 27409, 217715, 16576, 14960, 17169, 225994, *381393*, 14772, 13206, 78894, 70218, 69922, 26563, 272396, 13548, 93843, 30955, 23938, 230073, 13845, 17350, 26419, 66588, 73707, 19361, 21745, 71743, 74104,

					226564, 15165, 94045, 641170, 26921, 13835, 51797, 227231, 18669, 26448, 12323, 76408, 14158, 22241, 12545, 13404, 192170, 56379, 108888, 330177, 50883, 16882, 20877, 12144, 104721, 241633, 68277, 270151, 78943, 269152, 74653, 51789, 320119, 269400, 12988, 268930, 19090
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	80	9.090909091	0.041531268	108143, 56248, 71819, 270685, 17685, 22375, 170759, 76858, 26426, 192292, 270163, 246729, 11947, 12566, 22138, 226539, 27409, 217715, 16576, 14960, 17169, 225994, 381393, 14772, 13206, 78894, 70218, 69922, 26563, 272396, 13548, 93843, 30955, 23938, 230073, 13845, 17350, 26419, 66588, 73707, 19361, 21745, 71743, 74104, 226564, 15165, 94045, 641170, 26921, 13835, 51797, 227231, 18669, 26448, 12323, 76408, 14158, 22241, 12545, 13404, 192170, 56379, 108888, 330177, 50883, 16882, 20877, 12144, 104721, 241633, 68277, 270151, 78943, 269152, 74653, 51789, 320119, 269400, 12988, 268930, 19090
GOTERM_MF_FAT	GO:0001882~nucleoside binding	80	9.090909091	0.04580748	108143, 56248, 71819, 270685, 17685, 22375, 170759, 76858, 26426, 192292, 270163, 246729, 11947, 12566, 22138, 226539,

					27409, 217715, 16576, 14960, 17169, 225994, 381393, 14772, 13206, 78894, 70218, 69922, 26563, 272396, 13548, 93843, 30955, 23938, 230073, 13845, 17350, 26419, 66588, 73707, 19361, 21745, 71743, 74104, 226564, 15165, 94045, 641170, 26921, 13835, 51797, 227231, 18669, 26448, 12323, 76408, 14158, 22241, 12545, 13404, 192170, 56379, 108888, 330177, 50883, 16882, 20877, 12144, 104721, 241633, 68277, 270151, 78943, 269152, 74653, 51789, 320119, 269400, 12988, 268930, 19090
Annotation Cluster 10	***Enrichment Score: 1.3085075102776171***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0070013~intracellular organelle lumen	62	7.045454545	0.039825119	56194, 20602, 14230, 329650, 108143, 74195, 56248, 17130, 70088, 75416, 16906, 72050, 26754, 268903, 16588, 12566, 226539, 17345, 18572, 208727, 225876, 434632, 100608, 11865, 21425, 68153, 217431, 67278, 14281, 67838, 207932, 13548, 55989, 241764, 28019, 21665, 68083, 234407, 21745, 225432, 433759, 97112, 56036, 77286, 30942, 641170, 68479, 107094, 11611, 13163, 192170, 71911, 67397, 56280, 104394, 77053, 68272, 67154, 12040, 13619,

					102423, 70767, 545124, 20018
GOTERM_CC_FAT	GO:0043233~organelle lumen	62	7.045454545	0.042137516	56194, 20602, 14230, 329650, 108143, 74195, 56248, 17130, 70088, 75416, 16906, 72050, 26754, 268903, 16588, 12566, 226539, 17345, 18572, 208727, 225876, 434632, 100608, 11865, 21425, 68153, 217431, 67278, 14281, 67838, 207932, 13548, 55989, 241764, 28019, 21665, 68083, 234407, 21745, 225432, 433759, 97112, 56036, 77286, 30942, 641170, 68479, 107094, 11611, 13163, 192170, 71911, 67397, 56280, 104394, 77053, 68272, 67154, 12040, 13619, 102423, 70767, 545124, 20018
Annotation Cluster 11	***Enrichment Score: 1.1908721093253019***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:WD 5	16	1.818181818	0.03967891	68083, 21745, 97387, 73683, 94186, 217431, 216961, 319481, 237711, 27979, 240667, 106618, 319670, 98193, 22138, 57773
Annotation Cluster 12	***Enrichment Score: 1.1435917015794765***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
PIR_SUPERFAMILY	PIRSF002306:tubulin	4	0.454545455	0.043506607	22151, 73710, 22154, 56427
INTERPRO	IPR000217:Tubulin	4	0.454545455	0.050718133	22151, 73710, 22154, 56427
INTERPRO	IPR003008:Tubulin/FtsZ, GTPase domain	4	0.454545455	0.050718133	22151, 73710, 22154, 56427

Annotation Cluster 13	***Enrichment Score: 1.1207499044906828***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0008406~gonad development	9	1.022727273	0.032274547	12190, 12124, 12125, 268903, 11705, 17685, 217715, 56449, 13404
Annotation Cluster 14	***Enrichment Score: 1.1202523362784838***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0003841~1-acylglycerol-3-phosphate O-acyltransferase activity	3	0.340909091	0.032445308	68262, 67469, 52123
Annotation Cluster 15	***Enrichment Score: 1.0818070689751418***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	short sequence motif:JAMM motif	3	0.340909091	0.047280612	26754, 76630, 59029
Annotation Cluster 16	***Enrichment Score: 1.0509034743194163***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0000956~nuclear-transcribed mRNA catabolic process	5	0.568181818	0.04894117	319618, 103135, 22695, 17350, 641170, 192170
Annotation Cluster 17	***Enrichment Score: 0.995491218385854***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
PIR_SUPERFAMILY	PIRSF006060:AA_transporter	4	0.454545455	0.024023133	30962, 330836, 11989, 328059
INTERPRO	IPR002293:Amino acid/polyamine transporter I	4	0.454545455	0.028221426	30962, 330836, 11989, 328059

DAVID functional clusters of Mir-AT! identified genes. PAM cluster C.

Annotation Cluster 1	***Enrichment Score: 1.3226599595125694***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0032420~stereocilium	3	0.789473684	0.028538934	17913, 17921, 56226
GOTERM_CC_FAT	GO:0032421~stereocilium bundle	3	0.789473684	0.040161619	17913, 17921, 56226
Annotation Cluster 2	***Enrichment Score: 1.2201361871898928***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0009083~branched chain family amino acid catabolic process	3	0.789473684	0.007405606	12039, 12036, 11992
GOTERM_BP_FAT	GO:0009081~branched chain family amino acid metabolic process	3	0.789473684	0.025701612	12039, 12036, 11992
Annotation Cluster 3	***Enrichment Score: 0.6315422980535269***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	region of interest:Framework-4	3	0.789473684	0.055912228	626347, 626583, 676201

DAVID functional clusters of Mir-AT! identified genes. PAM cluster D.

Annotation Cluster 1	***Enrichment Score: 1.409501961255964***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:ANK 7	5	1.547987616	0.012406319	77097, 80859, 65257, 69635, 77318
UP_SEQ_FEATURE	repeat:ANK 6	5	1.547987616	0.043660221	77097, 80859, 65257, 69635, 77318

Annotation Cluster 2	***Enrichment Score: 1.3339246791720898***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0005524~ATP binding	32	9.907120743	0.028018771	50995, 66590, 208144, 329509, 22193, 330721, 100535, 17688, 13872, 110082, 212880, 27404, 333050, 11306, 101540, 117147, 215303, 18706, 67054, 68058, 23802, 69635, 110197, 18198, 78894, 21856, 110854, 244562, 18754, 26417, 79059, 21974
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	32	9.907120743	0.032238191	50995, 66590, 208144, 329509, 22193, 330721, 100535, 17688, 13872, 110082, 212880, 27404, 333050, 11306, 101540, 117147, 215303, 18706, 67054, 68058, 23802, 69635, 110197, 18198, 78894, 21856, 110854, 244562, 18754, 26417, 79059, 21974
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	32	9.907120743	0.056935188	50995, 66590, 208144, 329509, 22193, 330721, 100535, 17688, 13872, 110082, 212880, 27404, 333050, 11306, 101540, 117147, 215303, 18706, 67054, 68058, 23802, 69635, 110197, 18198, 78894, 21856, 110854, 244562, 18754, 26417, 79059, 21974
Annotation Cluster 3	***Enrichment Score: 1.200108563393097***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
INTERPRO	IPR002219:Protein kinase C, phorbol ester/diacylglycerol binding	4	1.238390093	0.044892882	18754, 110197, 333050, 101540
GOTERM_MF_FAT	GO:0019992~diacylglycerol	4	1.238390093	0.059833858	18754, 110197, 333050, 101540

	binding				
Annotation Cluster 4	***Enrichment Score: 0.9924576506500113***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0030286~dynein complex	4	1.238390093	0.015037558	330355, 110082, 56087, 13426

DAVID functional clusters of Mir-AT! identified genes. PAM cluster E1.

Annotation Cluster 1	***Enrichment Score: 1.884818136558638***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:13	8	0.822199383	0.001453975	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:12	8	0.822199383	0.001738216	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:11	8	0.822199383	0.003858525	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:9	8	0.822199383	0.008495577	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:10	8	0.822199383	0.008495577	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:8	8	0.822199383	0.028206288	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
UP_SEQ_FEATURE	repeat:7	8	0.822199383	0.055245982	20020, 13858, 50774, 81910, 14030, 12659, 57435, 16694
Annotation Cluster 2	***Enrichment Score: 1.8795365985856494***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:21	5	0.513874615	0.007630659	20020, 81910, 14030, 12659, 57435

UP_SEQ_FEATURE	repeat:19	5	0.513874615	0.009579745	20020, 81910, 14030, 12659, 57435
UP_SEQ_FEATURE	repeat:20	5	0.513874615	0.009579745	20020, 81910, 14030, 12659, 57435
UP_SEQ_FEATURE	repeat:18	5	0.513874615	0.009579745	20020, 81910, 14030, 12659, 57435
UP_SEQ_FEATURE	repeat:16	5	0.513874615	0.014382702	20020, 81910, 14030, 12659, 57435
UP_SEQ_FEATURE	repeat:17	5	0.513874615	0.017264124	20020, 81910, 14030, 12659, 57435
UP_SEQ_FEATURE	repeat:15	5	0.513874615	0.04184417	20020, 81910, 14030, 12659, 57435

Annotation Cluster 3	***Enrichment Score: 1.8666234753896018***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0043066~negative regulation of apoptosis	21	2.158273381	0.011649525	208449, 11684, 97541, 69635, 16149, 17869, 12211, 12367, 12824, 14102, 74268, 29810, 11798, 225372, 13871, 11799, 18049, 21812, 16183, 98558, 233575
GOTERM_BP_FAT	GO:0043069~negative regulation of programmed cell death	21	2.158273381	0.014391859	208449, 11684, 97541, 69635, 16149, 17869, 12211, 12367, 12824, 14102, 74268, 29810, 11798, 225372, 13871, 11799, 18049, 21812, 16183, 98558, 233575
GOTERM_BP_FAT	GO:0060548~negative regulation of cell death	21	2.158273381	0.014986647	208449, 11684, 97541, 69635, 16149, 17869, 12211, 12367, 12824, 14102, 74268, 29810, 11798, 225372, 13871, 11799, 18049, 21812, 16183, 98558, 233575

Annotation Cluster 4	***Enrichment Score: 1.837392134350211***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:29	4	0.411099692	0.008956029	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:24	4	0.411099692	0.012311551	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:28	4	0.411099692	0.012311551	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:23	4	0.411099692	0.016291401	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:25	4	0.411099692	0.016291401	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:26	4	0.411099692	0.016291401	20020, 81910, 14030, 57435

UP_SEQ_FEATURE	repeat:27	4	0.411099692	0.016291401	20020, 81910, 14030, 57435
UP_SEQ_FEATURE	repeat:22	4	0.411099692	0.020906828	20020, 81910, 14030, 57435
Annotation Cluster 5	***Enrichment Score: 1.7815884765199443***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0016702~oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	9	0.924974306	0.014850988	72898, 11684, 112406, 16922, 11689, 227696, 56720, 66627, 104263
GOTERM_MF_FAT	GO:0016701~oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	9	0.924974306	0.016123869	72898, 11684, 112406, 16922, 11689, 227696, 56720, 66627, 104263
SP_PIR_KEYWORDS	Dioxygenase	9	0.924974306	0.018880347	72898, 11684, 112406, 16922, 11689, 227696, 56720, 66627, 104263
Annotation Cluster 6	***Enrichment Score: 1.589198781592727***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0006915~apoptosis	34	3.494347379	0.016772978	208449, 68441, 107047, 106869, 12684, 66437, 17869, 234683, 72168, 67809, 12033, 74268, 29810, 11798, 13871, 11799, 11796, 17948, 26401, 170770, 327959, 69635, 16985, 12367, 14103, 14102, 13368, 74143, 13419, 11785, 57913, 18436, 19090, 23882
GOTERM_BP_FAT	GO:0012501~programmed cell death	34	3.494347379	0.02100619	208449, 68441, 107047, 106869, 12684, 66437, 17869, 234683, 72168, 67809, 12033, 74268, 29810, 11798, 13871, 11799, 11796, 17948, 26401, 170770, 327959, 69635, 16985, 12367, 14103, 14102, 13368, 74143, 13419, 11785, 57913, 18436, 19090, 23882

GOTERM_BP_FAT	GO:0008219~cell death	34	3.494347379	0.048467005	208449, 68441, 107047, 106869, 12684, 66437, 17869, 234683, 72168, 67809, 12033, 74268, 29810, 11798, 13871, 11799, 11796, 17948, 26401, 170770, 327959, 69635, 16985, 12367, 14103, 14102, 13368, 74143, 13419, 11785, 57913, 18436, 19090, 23882
Annotation Cluster 7	***Enrichment Score: 1.501120215965657***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
INTERPRO	IPR004344:Tubulin-tyrosine ligase	4	0.411099692	0.030940425	319953, 67534, 239591, 330010
GOTERM_MF_FAT	GO:0004835~tubulin-tyrosine ligase activity	4	0.411099692	0.031638699	319953, 67534, 239591, 330010
UP_SEQ_FEATURE	domain:TTL	4	0.411099692	0.032054903	319953, 67534, 239591, 330010
Annotation Cluster 8	***Enrichment Score: 1.4730477526847434***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0007189~activation of adenylate cyclase activity by G-protein signaling pathway	7	0.71942446	0.003796351	11556, 19219, 18390, 83428, 14680, 13490, 19226
GOTERM_BP_FAT	GO:0010578~regulation of adenylate cyclase activity involved in G-protein signaling	7	0.71942446	0.003796351	11556, 19219, 18390, 83428, 14680, 13490, 19226
GOTERM_BP_FAT	GO:0010579~positive regulation of adenylate cyclase activity by G-protein signaling pathway	7	0.71942446	0.003796351	11556, 19219, 18390, 83428, 14680, 13490, 19226
GOTERM_BP_FAT	GO:0007190~activation of adenylate cyclase activity	7	0.71942446	0.013069366	11556, 19219, 18390, 83428, 14680, 13490, 19226
GOTERM_BP_FAT	GO:0031281~positive regulation of cyclase	7	0.71942446	0.014644074	11556, 19219, 18390, 83428, 14680, 13490, 19226

	activity				
GOTERM_BP_FAT	GO:0051349~positive regulation of lyase activity	7	0.71942446	0.014644074	11556, 19219, 18390, 83428, 14680, 13490, 19226
GOTERM_BP_FAT	GO:0045762~positive regulation of adenylate cyclase activity	7	0.71942446	0.014644074	11556, 19219, 18390, 83428, 14680, 13490, 19226
Annotation Cluster 9	***Enrichment Score: 1.463960062933707***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:BIR 2	3	0.308324769	0.034358954	11798, 11796, 17948
UP_SEQ_FEATURE	repeat:BIR 3	3	0.308324769	0.034358954	11798, 11796, 17948
UP_SEQ_FEATURE	repeat:BIR 1	3	0.308324769	0.034358954	11798, 11796, 17948
Annotation Cluster 10	***Enrichment Score: 1.3705447340486716***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0007188~G-protein signaling, coupled to cAMP nucleotide second messenger	8	0.822199383	0.015919232	11556, 19219, 18390, 83428, 14680, 13490, 19773, 19226
GOTERM_BP_FAT	GO:0007187~G-protein signaling, coupled to cyclic nucleotide second messenger	8	0.822199383	0.024686679	11556, 19219, 18390, 83428, 14680, 13490, 19773, 19226
GOTERM_BP_FAT	GO:0019933~cAMP-mediated signaling	8	0.822199383	0.026771041	11556, 19219, 18390, 83428, 14680, 13490, 19773, 19226
GOTERM_BP_FAT	GO:0019935~cyclic-nucleotide-mediated signaling	8	0.822199383	0.038988859	11556, 19219, 18390, 83428, 14680, 13490, 19773, 19226

Annotation Cluster 11	***Enrichment Score: 1.1971272461224376***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	87	8.941418294	0.053872543	239591, 76722, 11836, 98985, 72508, 71819, 72168, 235661, 226604, 230577, 243659, 446099, 16548, 192292, 17948, 23962, 11981, 54721, 97541, 69635, 12750, 16337, 27410, 224824, 13204, 14287, 68291, 14635, 54369, 18639, 107932, 102448, 56495, 21975, 270076, 71586, 23938, 16578, 216965, 257632, 13841, 19882, 338372, 68087, 71742, 330010, 15166, 56325, 69719, 16560, 13871, 21812, 225028, 67475, 216150, 26401, 56456, 224805, 270672, 12468, 56520, 74376, 16913, 19088, 217207, 216578, 12041, 66368, 435766, 68278, 53608, 70762, 60525, 59030, 16886, 240216, 12361, 216019, 27103, 78903, 13139, 19139, 50490, 18436, 19090, 23954, 13424
Annotation Cluster 12	***Enrichment Score: 0.9038203070314017***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SP_PIR_KEYWORDS	gpi-anchor biosynthesis	5	0.513874615	0.025056808	14731, 231293, 329777, 14755, 233575
GOTERM_BP_FAT	GO:0006506~GPI anchor biosynthetic process	5	0.513874615	0.054277382	14731, 231293, 329777, 14755, 233575

DAVID functional clusters of Mir-AT! identified genes. PAM cluster E2.

Annotation Cluster 1	***Enrichment Score: 1.6555118190233722***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_MF_FAT	GO:0016279~protein-lysine N-methyltransferase	3	2.41935484	0.01009143	74729, 14056, 208266

	activity				
GOTERM_MF_FAT	GO:0016278~lysine N-methyltransferase activity	3	2.41935484	0.01009143	74729, 14056, 208266
GOTERM_MF_FAT	GO:0018024~histone-lysine N-methyltransferase activity	3	2.41935484	0.01009143	74729, 14056, 208266
GOTERM_MF_FAT	GO:0042054~histone methyltransferase activity	3	2.41935484	0.01690106	74729, 14056, 208266
GOTERM_MF_FAT	GO:0008170~N-methyltransferase activity	3	2.41935484	0.02744796	74729, 14056, 208266
GOTERM_MF_FAT	GO:0008276~protein methyltransferase activity	3	2.41935484	0.02861706	74729, 14056, 208266
Annotation Cluster 2	***Enrichment Score: 1.4100410568557769***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
INTERPRO	IPR018029:C2 membrane targeting protein	4	3.22580645	0.0255862	78771, 224860, 53420, 18796
INTERPRO	IPR000008:C2 calcium-dependent membrane targeting	4	3.22580645	0.04287180	78771, 224860, 53420, 18796
SMART	SM00239:C2	4	3.22580645	0.05366601	78771, 224860, 53420, 18796
Annotation Cluster 3	***Enrichment Score: 1.2888202640116764***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0042981~regulation of apoptosis	8	6.45161290	0.04917677	12981, 225372, 269275, 72168, 21933, 329244, 16193, 18129
GOTERM_BP_FAT	GO:0043067~regulation of programmed cell	8	6.45161290	0.05197673	12981, 225372, 269275, 72168, 21933, 329244, 16193, 18129

GOTERM_BP_FAT	death GO:0010941~regulation of cell death	8	6.45161290	0.05320719	12981, 225372, 269275, 72168, 21933, 329244, 16193, 18129
Annotation Cluster 4	***Enrichment Score: 1.0626006227925668***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SP_PIR_KEYWORDS	Microtubule	5	4.03225807	0.03568868	66884, 73804, 74764, 71819, 17318

DAVID functional clusters of Mir-AT! identified genes. PAM cluster E3.

Annotation Cluster 1	***Enrichment Score: 2.447636529115684***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0032956~regulation of actin cytoskeleton organization	6	1.986754967	0.001987345	242894, 19353, 218397, 13829, 18643, 13609
GOTERM_BP_FAT	GO:0032970~regulation of actin filament-based process	6	1.986754967	0.002138946	242894, 19353, 218397, 13829, 18643, 13609
GOTERM_BP_FAT	GO:0044087~regulation of cellular component biogenesis	6	1.986754967	0.010681134	242894, 19353, 218397, 13829, 18643, 13609
Annotation Cluster 2	***Enrichment Score: 2.0943214380544632***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0030833~regulation of actin filament polymerization	5	1.655629139	0.004464855	242894, 19353, 218397, 13829, 18643
GOTERM_BP_FAT	GO:0008064~regulation of actin polymerization or	5	1.655629139	0.00651393	242894, 19353, 218397, 13829, 18643

	depolymerization				
GOTERM_BP_FAT	GO:0030832~regulation of actin filament length	5	1.655629139	0.006987241	242894, 19353, 218397, 13829, 18643
GOTERM_BP_FAT	GO:0032271~regulation of protein polymerization	5	1.655629139	0.010307279	242894, 19353, 218397, 13829, 18643
GOTERM_BP_FAT	GO:0043254~regulation of protein complex assembly	5	1.655629139	0.016117311	242894, 19353, 218397, 13829, 18643
Annotation Cluster 3	***Enrichment Score: 1.3728777146340554***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SP_PIR_KEYWORDS	sh3 domain	8	2.649006623	0.034665831	13043, 218038, 234214, 338372, 218397, 213649, 59009, 15162
SMART	SM00326:SH3	8	2.649006623	0.046315808	13043, 218038, 234214, 338372, 218397, 213649, 59009, 15162
INTERPRO	IPR001452:Src homology-3 domain	8	2.649006623	0.047395378	13043, 218038, 234214, 338372, 218397, 213649, 59009, 15162
Annotation Cluster 4	***Enrichment Score: 1.1442294063930478***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0051494~negative regulation of cytoskeleton organization	4	1.324503311	0.034758594	76499, 13829, 18643, 13609
Annotation Cluster 5	***Enrichment Score: 1.1029182838781797***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0001739~sex chromatin	3	0.993377483	0.041066798	13626, 20587, 58186
GOTERM_CC_FAT	GO:0000803~sex	3	0.993377483	0.044724336	13626, 20587, 58186

	chromosome				
Annotation Cluster 6	***Enrichment Score: 1.0455102174052162***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SMART	SM00131:KU	3	0.993377483	0.037999616	209351, 433502, 20732
INTERPRO	IPR002223:Proteinase inhibitor I2, Kunitz metazoan	3	0.993377483	0.03802338	209351, 433502, 20732

DAVID functional clusters of Mir-AT! identified genes. PAM cluster E4.

Annotation Cluster 1	***Enrichment Score: 1.5415000501609881***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0000280~nuclear division	12	1.98019802	0.02633587	16319, 20962, 67956, 11799, 76499, 22152, 208628, 17120, 67141, 232679, 66578, 66934
GOTERM_BP_FAT	GO:0007067~mitosis	12	1.98019802	0.02633587	16319, 20962, 67956, 11799, 76499, 22152, 208628, 17120, 67141, 232679, 66578, 66934
GOTERM_BP_FAT	GO:0000087~M phase of mitotic cell cycle	12	1.98019802	0.030062137	16319, 20962, 67956, 11799, 76499, 22152, 208628, 17120, 67141, 232679, 66578, 66934
GOTERM_BP_FAT	GO:0048285~organelle fission	12	1.98019802	0.032725373	16319, 20962, 67956, 11799, 76499, 22152, 208628, 17120, 67141, 232679, 66578, 66934
Annotation Cluster 2	***Enrichment Score: 1.4693899209841048***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0043242~negative regulation of protein complex disassembly	5	0.825082508	0.019319891	20740, 76499, 208158, 17318, 12344

GOTERM_BP_FAT	GO:0043244~regulation of protein complex disassembly	5	0.825082508	0.037907668	20740, 76499, 208158, 17318, 12344
GOTERM_BP_FAT	GO:0051494~negative regulation of cytoskeleton organization	5	0.825082508	0.053345586	20740, 76499, 208158, 17318, 12344
Annotation Cluster 3	***Enrichment Score: 1.2960445149266488***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_CC_FAT	GO:0070013~intracellular organelle lumen	47	7.755775578	0.042204724	64656, 83701, 14049, 53975, 56095, 21417, 100121, 17869, 241525, 94066, 66078, 18205, 67902, 107094, 66645, 66071, 17984, 117147, 18572, 66101, 224805, 15387, 66448, 11689, 20586, 70791, 94112, 22218, 12041, 12632, 68272, 67710, 16418, 382053, 20466, 12040, 18949, 17126, 70673, 68776, 15476, 80509, 67619, 14056, 22431, 69241, 27393
GOTERM_CC_FAT	GO:0043233~organelle lumen	47	7.755775578	0.044469614	64656, 83701, 14049, 53975, 56095, 21417, 100121, 17869, 241525, 94066, 66078, 18205, 67902, 107094, 66645, 66071, 17984, 117147, 18572, 66101, 224805, 15387, 66448, 11689, 20586, 70791, 94112, 22218, 12041, 12632, 68272, 67710, 16418, 382053, 20466, 12040, 18949, 17126, 70673, 68776, 15476, 80509, 67619, 14056, 22431, 69241, 27393

Annotation Cluster 4	***Enrichment Score: 1.183386545693678***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	active site:Pros-phosphohistidine intermediate	3	0.495049505	0.020467795	171567, 79059, 54369
GOTERM_MF_FAT	GO:0004550~nucleoside diphosphate kinase activity	3	0.495049505	0.032150273	171567, 79059, 54369
GOTERM_BP_FAT	GO:0006241~CTP biosynthetic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0009209~pyrimidine ribonucleoside triphosphate biosynthetic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0009208~pyrimidine ribonucleoside triphosphate metabolic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0006228~UTP biosynthetic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0046051~UTP metabolic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0046036~CTP metabolic process	3	0.495049505	0.033730414	171567, 79059, 54369
GOTERM_BP_FAT	GO:0006183~GTP biosynthetic process	3	0.495049505	0.040431014	171567, 79059, 54369
GOTERM_BP_FAT	GO:0046039~GTP metabolic process	3	0.495049505	0.040431014	171567, 79059, 54369
GOTERM_BP_FAT	GO:0009148~pyrimidine nucleoside triphosphate biosynthetic process	3	0.495049505	0.040431014	171567, 79059, 54369
GOTERM_BP_FAT	GO:0009220~pyrimidine ribonucleotide biosynthetic process	3	0.495049505	0.04758398	171567, 79059, 54369
GOTERM_BP_FAT	GO:0009218~pyrimidine ribonucleotide metabolic process	3	0.495049505	0.04758398	171567, 79059, 54369

Annotation Cluster 5	***Enrichment Score: 1.1766325249959473***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0001556~oocyte maturation	3	0.495049505	0.033730414	56484, 67141, 54611
Annotation Cluster 6	***Enrichment Score: 1.0651705919969658***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0009201~ribonucleoside triphosphate biosynthetic process	7	1.155115512	0.058041463	114143, 140494, 171567, 66043, 79059, 54369, 11972
GOTERM_BP_FAT	GO:0009206~purine ribonucleoside triphosphate biosynthetic process	7	1.155115512	0.058041463	114143, 140494, 171567, 66043, 79059, 54369, 11972
Annotation Cluster 7	***Enrichment Score: 0.9512776882973525***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0031343~positive regulation of cell killing	4	0.660066007	0.029535595	14972, 15018, 12479, 54698
GOTERM_BP_FAT	GO:0001912~positive regulation of leukocyte mediated cytotoxicity	4	0.660066007	0.029535595	14972, 15018, 12479, 54698
GOTERM_BP_FAT	GO:0001910~regulation of leukocyte mediated cytotoxicity	4	0.660066007	0.058235634	14972, 15018, 12479, 54698
GOTERM_BP_FAT	GO:0031341~regulation of cell killing	4	0.660066007	0.058235634	14972, 15018, 12479, 54698
Annotation Cluster 8	***Enrichment Score: 0.924746738578303***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0006164~purine nucleotide biosynthetic	9	1.485148515	0.051330871	114143, 140494, 171567, 66043, 79059, 18159, 224129,

	process				54369, 11972
Annotation Cluster 9	***Enrichment Score: 0.7933604380470283***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0001916~positive regulation of T cell mediated cytotoxicity	3	0.495049505	0.027516354	14972, 15018, 12479
GOTERM_BP_FAT	GO:0001914~regulation of T cell mediated cytotoxicity	3	0.495049505	0.040431014	14972, 15018, 12479
GOTERM_BP_FAT	GO:0002711~positive regulation of T cell mediated immunity	3	0.495049505	0.055156727	14972, 15018, 12479
Annotation Cluster 10	***Enrichment Score: 0.7727633746934233***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	region of interest:Alpha-3	3	0.495049505	0.033588489	15007, 14972, 15018
SP_PIR_KEYWORDS	mhc i	3	0.495049505	0.042625505	15007, 14972, 15018
UP_SEQ_FEATURE	region of interest:Alpha-2	3	0.495049505	0.049135698	15007, 14972, 15018
UP_SEQ_FEATURE	region of interest:Alpha-1	3	0.495049505	0.049135698	15007, 14972, 15018
Annotation Cluster 11	***Enrichment Score: 0.6821682207681469***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0045059~positive thymic T cell selection	3	0.495049505	0.021824664	12479, 16149, 22637
GOTERM_BP_FAT	GO:0043368~positive T cell selection	3	0.495049505	0.033730414	12479, 16149, 22637

DAVID functional clusters of Mir-AT! identified genes. PAM cluster E5.

Annotation Cluster 1	***Enrichment Score: 1.275528304728059***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
SP_PIR_KEYWORDS	atp-binding	22	10.4761905	0.03167769	66590, 108888, 20874, 54402, 74451, 74104, 381921, 380698, 225997, 328099, 68015, 17687, 268281, 18669, 60525, 20187, 83560, 26409, 20133, 23939, 76826, 244859
Annotation Cluster 2	***Enrichment Score: 1.156270400343773***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0006874~cellular calcium ion homeostasis	4	1.90476191	0.05876174	16963, 14738, 59091, 19226
Annotation Cluster 4	***Enrichment Score: 1.0910513596513698***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0007584~response to nutrient	4	1.90476191	0.03770392	214601, 54402, 13198, 17687
Annotation Cluster 5	***Enrichment Score: 1.00572511760775***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:HEAT 5	3	1.42857143	0.04490314	76499, 75786, 54160
Annotation Cluster 6	***Enrichment Score: 0.856396376884771***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
GOTERM_BP_FAT	GO:0030005~cellular di-, tri-valent inorganic cation homeostasis	5	2.38095238	0.04160361	16963, 328695, 14738, 59091, 19226
GOTERM_BP_FAT	GO:0055066~di-, tri-valent inorganic cation homeostasis	5	2.38095238	0.05394031	16963, 328695, 14738, 59091, 19226

Annotation Cluster 7	***Enrichment Score: 0.5369290300199814***				
Category	**Term**	**Count**	**%**	**P-value**	**Genes**
UP_SEQ_FEATURE	repeat:WD 13	3	1.42857143	0.01913323	213773, 319670, 217325
UP_SEQ_FEATURE	repeat:WD 12	3	1.42857143	0.02315367	213773, 319670, 217325
UP_SEQ_FEATURE	repeat:WD 10	3	1.42857143	0.03703438	213773, 319670, 217325
UP_SEQ_FEATURE	repeat:WD 11	3	1.42857143	0.03703438	213773, 319670, 217325
UP_SEQ_FEATURE	repeat:WD 9	3	1.42857143	0.05624868	213773, 319670, 217325

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389 -3402. doi:10.1093/nar/25.17.3389

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*, 350-355.

Anderson, R., Schaible, K., Heasman, J., & Wylie, C. (1999). Expression of the homophilic adhesion molecule, Ep-CAM, in the mammalian germ line. *Journal of Reproduction and Fertility*, *116*(2), 379 -384. doi:10.1530/jrf.0.1160379

Apweiler, R. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, *32*, 115D-119. doi:10.1093/nar/gkh131

Aravin, A.A., & Hannon, G. J. (2008). Small RNA Silencing Pathways in Germ and Stem Cells. *Cold Spring Harbor Symposia on Quantitative Biology*. doi:10.1101/sqb.2008.73.058

Aravin, Alexei A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science*, *318*(5851), 761 -764. doi:10.1126/science.1146484

Aravin, Alexei A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., et al. (2003). The Small RNA Profile during *Drosophila melanogaster* Development. *Developmental Cell*, *5*(2), 337-350. doi:10.1016/S1534-5807(03)00228-4

Aravin, Alexei A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., et al. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, *31*(6), 785-799. doi:10.1016/j.molcel.2008.09.003

Aravin, Alexei 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. doi:10.1126/science.1142612

Aravin, Alexei A., van der Heijden, G. W., Castañeda, J., Vagin, V. V., Hannon, G. J., & Bortvin, A. (2009). Cytoplasmic Compartmentalization of the Fetal piRNA Pathway in Mice. *PLoS Genet*, *5*(12), e1000764. doi:10.1371/journal.pgen.1000764

Armisen, J., Gilchrist, M. J., Wilczynska, A., Standart, N., & Miska, E. A. (2009). Abundant and dynamically expressed miRNAs, piRNAs, and other small RNAs in the vertebrate *Xenopus tropicalis*. *Genome Research*, *19*(10), 1766 -1775. doi:10.1101/gr.093054.109

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64-71. doi:10.1038/nature07242

Bairoch, A. (2004). The Universal Protein Resource (UniProt). *Nucleic Acids Research*, *33*, D154-D159. doi:10.1093/nar/gki070

Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., et al. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, *441*(7089), 87-90. doi:10.1038/nature04696

Berg, J. M. (1990). Zinc Finger Domains: Hypotheses and Current Knowledge. *Annual Review of Biophysics and Biophysical Chemistry*, *19*, 405-421. doi:10.1146/annurev.bb.19.060190.002201

Bird, A. P., & Taggart, M. H. (1980). Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Research*, *8*(7), 1485 -1497. doi:10.1093/nar/8.7.1485

Bowers, John E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, *422*(6930), 433-438. doi:10.1038/nature01521

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., & Cohen, S. M. (2003). bantam Encodes a Developmentally Regulated microRNA that Controls Cell Proliferation and Regulates the Proapoptotic Gene hid in *Drosophila*. *Cell*, *113*(1), 25-36. doi:10.1016/S0092-8674(03)00231-9

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science*, *322*(5906), 1387 -1392. doi:10.1126/science.1165171

Brothers, A. N., & Delph, L. F. (2010). Haldane's Rule Is Extended to Plants with Sex Chromosomes. *Evolution*, *64*(12), 3643-3648. doi:10.1111/j.1558-5646.2010.01095.x

Brunet, F. G., Crollius, H. R., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., Laudet, V., et al. (2006). Gene Loss and Evolutionary Rates Following Whole-Genome Duplication in Teleost Fishes. *Molecular Biology and Evolution*, *23*(9), 1808 -1816. doi:10.1093/molbev/msl049

Bürki, E., & Fischberg, M. (1985). Evolution of globin expression in the genus *Xenopus* (Anura: Pipidae). *Molecular Biology and Evolution*, *2*(3), 270 -277.

Buzek, S. W., & Sanborn, B. M. (1988). Increase in testicular androgen receptor during sexual maturation in the rat. *Biology of Reproduction*, *39*(1), 39 -49. doi:10.1095/biolreprod39.1.39

Cao, R., & Zhang, Y. (2004). SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Molecular Cell*, *15*(1), 57-67. doi:10.1016/j.molcel.2004.06.020

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. doi:10.1016/j.devcel.2007.03.001

Chain, F. J. J., & Evans, B. J. (2006). Multiple Mechanisms Promote the Retained Expression of Gene Duplicates in the Tetraploid Frog *Xenopus laevis*. *PLoS Genetics*, *2*(4). doi:10.1371/journal.pgen.0020056

Chang, C., Chen, Y.-T., Yeh, S.-D., Xu, Q., Wang, R.-S., Guillou, F., Lardy, H., et al. (2004). Infertility with defective spermatogenesis and hypotestosteronemia in male mice lacking the androgen receptor in Sertoli cells. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(18), 6876 -6881. doi:10.1073/pnas.0307306101

Cheng, J., Wang, D., Wang, Z., & Yeh, E. T. H. (2004). SENP1 Enhances Androgen Receptor-Dependent Transcription through Desumoylation of Histone Deacetylase 1. *Molecular and Cellular Biology*, *24*(13), 6021 -6028. doi:10.1128/MCB.24.13.6021-6028.2004

Comai, L., Madlung, A., Josefsson, C., & Tyagi, A. (2003). Do the different parental "heteromes" cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1434), 1149 -1155. doi:10.1098/rstb.2003.1305

Comings, D. E., & MacMurray, J. P. (2000). Molecular Heterosis: A Review. *Molecular Genetics and Metabolism*, *71*(1-2), 19-31. doi:10.1006/mgme.2000.3015

Cuzin, F., & Rassoulzadegan, M. (2010). Non-Mendelian epigenetic heredity: gametic RNAs as epigenetic regulators and transgenerational signals. *Essays in Biochemistry*, *48*, 101-106. doi:10.1042/bse0480101

Daboule, D. (n.d.). *Guide to the homeobox genes.* New York, NY: Oxford Univ. Press.

Dadoune, J. (2009). Spermatozoal RNAs: What about their functions? *Microscopy Research and Technique*, *72*(8), 536-551. doi:10.1002/jemt.20697

Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A., Buhecha, H. R., et al. (2008). Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the Caenorhabditis elegans Germline. *Molecular Cell*, *31*(1), 79-90. doi:10.1016/j.molcel.2008.06.003

Deng, W., & Lin, H. (2002). miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Developmental Cell*, *2*(6), 819-830. doi:10.1016/S1534-5807(02)00165-X

Devor, E. J., & Samollow, P. B. (2008). In Vitro and In Silico Annotation of Conserved and Nonconserved MicroRNAs in the Genome of the Marsupial *Monodelphis domestica*. *Journal of Heredity*, *99*(1), 66 -72. doi:10.1093/jhered/esm085

Devor, E. J., Huang, L., & Samollow, P. B. (2008). piRNA-like RNAs in the marsupial *Monodelphis domestica* identify transcription clusters and likely marsupial transposon targets. *Mammalian Genome*, *19*, 581-586. doi:10.1007/s00335-008-9109-x

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863 -14868.

Elmén, J., Lindow, M., Silahtaroglu, A., Bak, M., Christensen, M., Lind-Thomsen, A., Hedtjärn, M., et al. (2008). Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Research*, *36*(4), 1153 -1162. doi:10.1093/nar/gkm1113

Emery, B. R., & Carrell, D. T. (2006). The effect of epigenetic sperm abnormalities on early embryogenesis. *Asian J Androl*, *8*(2), 131-142.

Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J., & Cannatella, D. C. (2004). A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution*, *33*(1), 197-213. doi:10.1016/j.ympev.2004.04.018

Evgen'ev, M. B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.-H., & Corces, V. G. (1997). Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proceedings of the National Academy of Sciences*, *94*(1), 196 -201.

Farh, K. K.-H., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B., et al. (2005). The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. *Science*, *310*(5755), 1817 -1821. doi:10.1126/science.1121158

Fawcett, D.W., Eddy, W.M. & Phillips, D.M. (1970). Observations on the Fine Structure and Relationships of the Chromatoid Body in Mammalian Spermatogenesis. *Biology of Reproduction*, *2*(1), 129 -153. doi:10.1095/biolreprod2.1.129

Felli, N., Fontana, L., Pelosi, E., Botta, R., Bonci, D., Facchiano, F., Liuzzi, F., et al. (2005). MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(50), 18081 -18086. doi:10.1073/pnas.0506216102

Feng, H. L., Sandlow, J. I., Sparks, A. E. ., Sandra, A., & Zheng, L. J. (1999). Decreased expression of the c-kit receptor is associated with increased apoptosis in subfertile human testes. *Fertility and Sterility*, *71*(1), 85-89. doi:10.1016/S0015-0282(98)00401-4

Ferguson-Smith, A. C., & Reik, W. (2003). The need for Eed. *Nat Genet*, *33*(4), 433-434. doi:10.1038/ng0403-433

Feschotte, C. (2008). The contribution of transposable elements to the evolution of regulatory networks. *Nature reviews. Genetics*, *9*(5), 397-405. doi:10.1038/nrg2337

Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual review of genetics*, *41*, 331-368. doi:10.1146/annurev.genet.40.110405.090448

Fischer, W. J., Koch, W. A., & Elepfandt, A. (2000). Sympatry and hybridization between the clawed frogs *Xenopus laevis laevis* and *Xenopus muelleri* (Pipidae). *Journal of Zoology*, *252*(1), 99-107. doi:10.1111/j.1469-7998.2000.tb00824.x

Francis, T. & (2011, November 7). Biological Systems in Vertebrates V4: Reproduction of Amphibians (eBook) - Taylor & Francis. Text, . Retrieved November 7, 2011, from http://www.taylorandfrancis.com/books/details/9781439842263/

Franco, S., Canela, A., Klatt, P., & Blasco, M. A. (2005). Effectors of mammalian telomere dysfunction: a comparative transcriptome analysis using mouse models. *Carcinogenesis*, *26*(9), 1613 -1626. doi:10.1093/carcin/bgi107

Freedman. L. P. (1992). Anatomy of the Steroid Receptor Zinc Finger Region. *Endocrine Reviews*, *13*(2), 129 -145. doi:10.1210/edrv-13-2-129

Gabbianelli, M., Testa, U., Morsilli, O., Pelosi, E., Saulle, E., Petrucci, E., Castelli, G., et al. (2010). Mechanism of human Hb switching: a possible role of the kit receptor/miR 221-222 complex. *Haematologica*, *95*(8), 1253 -1260. doi:10.3324/haematol.2009.018259

Genome analysis of the platypus reveals unique signatures of evolution. (2008).*Nature*, *453*(7192), 175-183. doi:10.1038/nature06936

Girard, Angélique, Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199-202. doi:10.1038/nature04917

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, *128*(4), 635-638. doi:10.1016/j.cell.2007.02.006

Grandjean, V., & Rassoulzadegan, M. (2009). Epigenetic inheritance of the sperm: an unexpected role of RNA. *Gynécologie, Obstétrique & Fertilité*, *37*(6), 558-561. doi:10.1016/j.gyobfe.2009.04.005

Green, G. R., Balhorn, R., Poccia, D. L., & Hecht, N. B. (1994). Synthesis and processing of mammalian protamines and transition proteins. *Molecular Reproduction and Development*, *37*(3), 255-263. doi:10.1002/mrd.1080370303

Griffiths-Jones, S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, *34*, D140-D144. doi:10.1093/nar/gkj112

Griffiths-Jones, S. (2011). miRBase: microRNA Sequences and Annotation. doi:10.1002/0471250953.bi1209s29

Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2007a). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, *36*, D154-D158. doi:10.1093/nar/gkm952

Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2007b). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, *36*, D154-D158. doi:10.1093/nar/gkm952

Griffiths-Jones, Sam. (2011). miRBase: The MicroRNA Sequence Database. *MicroRNA Protocols* (Vol. 342, pp. 129-138). New Jersey: Humana Press. Retrieved from http://www.springerlink.com/content/pp2761057r192h42/#section=83853&page=1

Griswold, M. D. (1998, August). The central role of Sertoli cells in spermatogenesis. Article, . Retrieved November 3, 2011, from https://digital.lib.washington.edu/researchworks/handle/1773/4296

Grivna, S. T., Beyret, E., Wang, Z., & Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development*, *20*(13), 1709 -1714. doi:10.1101/gad.1434406

Grivna, S. T., Pyhtila, B., & Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proceedings of the National Academy of Sciences*, *103*(36), 13415 -13420. doi:10.1073/pnas.0605506103

Gross, B. L., Schwarzbach, A. E., & Rieseberg, L. H. (2003). Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany*, *90*(12), 1708 -1719. doi:10.3732/ajb.90.12.1708

Guo, M., Davis, D., & Birchler, J. A. (1996). Dosage Effects on Gene Expression in a *Maize* Ploidy Series. *Genetics*, *142*(4), 1349 -1355.

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K. D., Chapman, E. J., Carrington, J. C., et al. (2009). Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences*, *106*(42), 17835 -17840. doi:10.1073/pnas.0907003106

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, *12*, 101-109. doi:10.1007/BF02983075

Haraguchi, C. M., Mabuchi, T., Hirata, S., Shoda, T., Hoshi, K., Akasaki, K., & Yokota, S. (2005). Chromatoid Bodies: Aggresome-like Characteristics and Degradation Sites for Organelles of Spermiogenic Cells. *Journal of Histochemistry & Cytochemistry*, *53*(4), 455 -465. doi:10.1369/jhc.4A6520.2005

Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A., & Wendel, J. F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Research*, *16*(10), 1252 -1261. doi:10.1101/gr.5282906

He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G. A., et al. (2005). The role of microRNA genes in papillary thyroid carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 19075 -19080. doi:10.1073/pnas.0509603102

He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G. A., et al. (2005). The role of microRNA genes in papillary thyroid carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 19075 -19080. doi:10.1073/pnas.0509603102

He, Lin, & Hannon, G. J. (2004). MicroRNAs: SMALL RNAs WITH A BIG ROLE IN GENE REGULATION. *Nature Review Genetics*, *5*, 522-531.

He, Lin, Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., et al. (2005). A microRNA polycistron as a potential human oncogene. *Nature*, *435*(7043), 828-833. doi:10.1038/nature03552

He, Z., Kokkinaki, M., Pant, D., Gallicano, G. I., & Dym, M. (2009). Small RNA molecules in the regulation of spermatogenesis. *Reproduction*, *137*(6), 901 -911. doi:10.1530/REP-08-0494

Heald, R., Tournebize, R., Blank, T., Sandaltzopoulos, R., Becker, P., Hyman, A., & Karsenti, E. (1996). Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature*, *382*(6590), 420-425. doi:10.1038/382420a0

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., et al. (2010). The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science*, *328*(5978), 633 -636. doi:10.1126/science.1183670

Hervé, S. (2009). Redefining MicroRNA Targets. *Current Biology*, *19*(10), 870-873. doi:10.1016/j.cub.2009.03.059

Hnilica, L. S., Stein, G. S., & Stein, J. L. (1989). *Histones and other basic nuclear proteins*. CRC Press.

Hore, T. A., Rapkins, R. W., & Graves, J. A. M. (2007). Construction and evolution of imprinted loci in mammals. *Trends in Genetics*, *23*(9), 440-448. doi:10.1016/j.tig.2007.07.003

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., et al. (2007). A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell*, *129*(1), 69-82. doi:10.1016/j.cell.2007.03.026

Hsieh-Li, H. M., Witte, D. P., Weinstein, M., Branford, W., Li, H., Small, K., & Potter, S. S. (1995). Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility. *Development*, *121*(5), 1373 -1385.

Hughes, I. (n.d.). Mini-Review: Sex differentiation. *Endocrinology*, *142*, 3281–3287.

Ikenishi, K., & Tanaka, T. S. (1997). Involvement of the protein of *Xenopus* vasa homolog (Xenopus vasa-like gene 1, XVLG1) in the differentiation of primordial germ cells. *Development, Growth & Differentiation*, *39*(5), 625-633. doi:10.1046/j.1440-169X.1997.t01-4-00010.x

Ikenishi, K., & Tanaka, T. S. (2000). Spatio-temporal expression of *Xenopus* vasa homolog, XVLG1, in oocytes and embryos: The presence of XVLG1 RNA in somatic cells as well as germline cells. *Development, Growth & Differentiation*, *42*(2), 95-103. doi:10.1046/j.1440-169x.2000.00493.x

Innis, J. W., Goodman, F. R., Bacchelli, C., Williams, T. M., Mortlock, D. P., Sateesh, P., Scambler, P. J., et al. (2002). A HOXA13 allele with a missense mutation in the homeobox and a dinucleotide deletion in the promoter underlies Guttmacher syndrome. *Human Mutation*, *19*(5), 573-574. doi:10.1002/humu.9036

Jones, F. S., Prediger, E. A., Bittner, D. A., De Robertis, E. M., & Edelman, G. M. (1992). Cell adhesion molecules as targets for Hox genes: neural cell adhesion molecule promoter activity is modulated by cotransfection with Hox-2.5 and -2.4. *Proceedings of the National Academy of Sciences*, *89*(6), 2086 -2090.

Josefsson, C., Dilkes, B., & Comai, L. (2006). Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. *Current Biology*, *16*(13), 1322-1328. doi:10.1016/j.cub.2006.05.045

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, *110*, 462-467. doi:10.1159/000084979

Jurka, Jerzy, Klonowski, P., Dagman, V., & Pelton, P. (1996). Censor—a program for identification and elimination of repetitive elements from DNA sequences. *Computers & Chemistry*, *20*(1), 119-121. doi:10.1016/S0097-8485(96)80013-1

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*(6983), 617-624. doi:10.1038/nature02424

Kimmins, S., & Sassone-Corsi, P. (n.d.). Chromatin remodelling and epigenetic features of germ cells. *Nature*, *434*, 583-589.

Kimura, M., & Ohta, T. (1974). On Some Principles Governing Molecular Evolution. Proceedings of the National Academy of Sciences, 71(7), 2848 -2852.

Kissel, H., Timokhina, I., Hardy, M. P., Rothschild, G., Tajima, Y., Soares, V., Angeles, M., et al. (2000). Point mutation in Kit receptor tyrosine kinase reveals essential roles for Kit signaling in spermatogenesis and oogenesis without affecting other Kit responses. *EMBO J*, *19*(6), 1312-1326. doi:10.1093/emboj/19.6.1312

Kistler, W. S., Henriksén, K., Mali, P., & Parvinen, M. (1996). Sequential Expression of Nucleoproteins during Rat Spermiogenesis. *Experimental Cell Research*, *225*(2), 374-381. doi:10.1006/excr.1996.0188

Klattenhoff, C., & Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. *Development*, *135*(1), 3 -9. doi:10.1242/dev.006486

Kobel, H. R., & Pasquier, L. (1975). Production of large clones of histocompatible, fully identical clawed toads (*Xenopus*). *Immunogenetics*, *2*, 87-91. doi:10.1007/BF01572278

Kobel, H. R., Pasquier, L. D., & Tinsley, R. C. (1981). Natural hybridization and gene introgression between Xenopus gilli and Xenopus laevis laevis (Anura: Pipidae). *Journal of Zoology*, *194*(3), 317-322. doi:10.1111/j.1469-7998.1981.tb04584.x

Koide, S. S., Wang, L., & Kamada, M. (2000). Antisperm Antibodies Associated with Infertility: Properties and Encoding Genes of Target□Antigens. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, *224*(3), 123 -132.

Kondo, T., Zakany, J., Innis, J. W., & Duboule, D. (1997). Of fingers, toes and penises. *Nature*, *390*(6655), 29. doi:10.1038/36234

Koroma, A.P. (2010). Comparison of the DNA methylation patterns between interspecific Xenopus F1 hybrids and their parental species (Doctoral Dissertation). Retrieved from ProQuest Dissertation and Theses database. (UMI number 3456791).

Koroma, A. P., Jones, R., & Michalak, P. (2011). A snapshot of DNA methylation changes associated with hybridization in Xenopus. *Physiological Genomics*. doi:10.1152/physiolgenomics.00110.2011

Kozomara, A., & Griffiths-Jones, S. (2010). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, *39*, D152-D157. doi:10.1093/nar/gkq1027

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development*, *131*(4), 839 -849. doi:10.1242/dev.00973

Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., Sato, M., et al. (2001). Two mouse piwi-related genes: miwi and mili. *Mechanisms of Development*, *108*(1-2), 121-133. doi:10.1016/S0925-4773(01)00499-3

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI

and MIWI2 in murine fetal testes. *Genes & Development*, *22*(7), 908 -917. doi:10.1101/gad.1640708

Lacham-Kaplan, O. (2004). In vivo and in vitro differentiation of male germ cells in the mouse. *Reproduction*, *128*(2), 147 -152. doi:10.1530/rep.1.00220

Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E., & Blower, M. D. (2009). Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. *EMBO J*, *28*(19), 2945-2958. doi:10.1038/emboj.2009.237

Lau, N. C., Robine, N., Martin, R., Chung, W.-J., Niki, Y., Berezikov, E., & Lai, E. C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Research*, *19*(10), 1776 -1785. doi:10.1101/gr.094896.109

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. doi:10.1126/science.1130164

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, *120*(1), 15-20. doi:10.1016/j.cell.2004.12.035

Lewis, B. P., Shih, I.-hung, Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of Mammalian MicroRNA Targets. *Cell*, *115*(7), 787-798. doi:10.1016/S0092-8674(03)01018-3

Liu, C.-G., Calin, G. A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C. D., et al. (2004). An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(26), 9740 -9744. doi:10.1073/pnas.0403293101

Loukinov, D. I., Pugacheva, E., Vatolin, S., Pack, S. D., Moon, H., Chernukhin, I., Mannan, P., et al. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein

involved in reading imprinting marks in the soma. *Proceedings of the National Academy of Sciences*, *99*(10), 6806 -6811. doi:10.1073/pnas.092123699

Lynch, M., & Conery, J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, *290*(5494), 1151 -1155. doi:10.1126/science.290.5494.1151

Lynch, M., & Conery, J. S. (2003). The Origins of Genome Complexity. *Science*, *302*(5649), 1401 -1404. doi:10.1126/science.1089370

Lynch, M., & Force, A. (2000). The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics*, *154*(1), 459 -473.

Lynch, M., O'Hely, M., Walsh, B., & Force, A. (2001). The Probability of Preservation of a Newly Arisen Gene Duplicate. *Genetics*, *159*(4), 1789 -1804.

Maatouk, D. M., Loveland, K. L., McManus, M. T., Moore, K., & Harfe, B. D. (2008a). Dicer1 Is Required for Differentiation of the Mouse Male Germline. *Biology of Reproduction*, *79*(4), 696 -703. doi:10.1095/biolreprod.108.067827

Maatouk, D. M., Loveland, K. L., McManus, M. T., Moore, K., & Harfe, B. D. (2008b). Dicer1 Is Required for Differentiation of the Mouse Male Germline. *Biology of Reproduction*, *79*(4), 696 -703. doi:10.1095/biolreprod.108.067827

MacLean II, J. A., Chen, M. A., Wayne, C. M., Bruce, S. R., Rao, M., Meistrich, M. L., Macleod, C., et al. (2005). Rhox: A New Homeobox Gene Cluster. *Cell*, *120*(3), 369-382. doi:10.1016/j.cell.2004.12.022

Mager, J., Montgomery, N. D., de Villena, F. P.-M., & Magnuson, T. (2003). Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nat Genet*, *33*(4), 502-507. doi:10.1038/ng1125

Malone, J. H., & Michalak, P. (2008). Physiological Sex Predicts Hybrid Sterility Regardless of Genotype. *Science*, *319*(5859), 59. doi:10.1126/science.1148231

Malone, J. H., Chrzanowski, T. H., & Michalak, P. (2007). Sterility and Gene Expression in Hybrid Males of *Xenopus laevis* and *X. muelleri*. *PLoS ONE*, *2*(8), e781. doi:10.1371/journal.pone.0000781

Martin-ponthieu, A., Wouters-tyrou, D., Pudlo, B., Buisine, E., & Sautière, P. (1994). Isolation and characterization of a small putative zinc finger protein from cuttlefish epididymal sperm cells. *European Journal of Biochemistry*, *220*(2), 463-468. doi:10.1111/j.1432-1033.1994.tb18644.x

Martins, R. P., & Krawetz, S. A. (2007). Nuclear organization of the protamine locus. *Reproduction in Domestic Ruminants*, *6*(1), 1-12. doi:10.5661/RDR-VI-1

Massirer, K. B., & Pasquinelli, A. E. (2006). The evolving role of microRNAs in animal gene expression. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *28*(5), 449-452. doi:10.1002/bies.20406

Mattick, John S, Amaral, P. P., Dinger, M. E., Mercer, T. R., & Mehler, M. F. (2009). RNA regulation of epigenetic processes. *BioEssays*, *31*(1), 51-59. doi:10.1002/bies.080099

Meister, M., & Tuschl, T. (n.d.). Mechanisms of gene silencing by double-stranded RNA Gunter Meister. *Nature*, *431*, 343-349.

Michalak, Pawel, & Malone, J. H. (2008). Testis-derived microRNA profiles of African clawed frogs (*Xenopus*) and their sterile hybrids. *Genomics*, *91*(2), 158-164. doi:10.1016/j.ygeno.2007.10.013

Microcosm Targets. (2011, November 5). Retrieved November 5, 2011, from http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/

Mishima, T., Takizawa, T., Luo, S.-S., Ishibashi, O., Kawahigashi, Y., Mizuguchi, Y., Ishikawa, T., et al. (2008). MicroRNA (miRNA) cloning analysis reveals sex differences in miRNA expression profiles between adult mouse testis and ovary. *Reproduction*, *136*(6), 811 -822. doi:10.1530/REP-08-0349

Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S., & Papaioannou, V. E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*, *68*(5), 869-877. doi:10.1016/0092-8674(92)90030-G

Montgomery, N. D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S. J., Otte, A. P., & Magnuson, T. (2005). The Murine Polycomb Group Protein Eed Is Required for Global Histone H3 Lysine-27 Methylation. *Current Biology*, *15*(10), 942-947. doi:10.1016/j.cub.2005.04.051

Muller, HJ. (1940). *Muller: Bearing of the Drosophila work on systematics - Google Scholar*. Huxley. Retrieved from http://scholar.google.com/scholar?cluster=5313316019279883711&hl=en&as_sdt=0,44

Murchison, E. P., Kheradpour, P., Sachidanandam, R., Smith, C., Hodges, E., Xuan, Z., Kellis, M., et al. (2008). Conservation of small RNA pathways in platypus. *Genome Research*, *18*(6), 995 - 1004. doi:10.1101/gr.073056.107

Nagai, M., Moriyama, T., Mehmood, R., Tokuhiro, K., Ikawa, M., Okabe, M., Tanaka, H., et al. (2011). Mice lacking Ran binding protein 1 are viable and show male infertility. *FEBS Letters*, *585*(5), 791-796. doi:10.1016/j.febslet.2011.02.002

Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag.

Orr, H A, & Presgraves, D. C. (2000). Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *22*(12), 1085-1094. doi:10.1002/1521-1878(200012)22:12<1085::AID-BIES6>3.0.CO;2-G

Orr, H. Allen. (1997). Haldane's Rule. *Annual Review of Ecology and Systematics*, *28*, 195-218.

Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P., & Krawetz, S. A. (2004). Reproductive biology: Delivering spermatozoan RNA to the oocyte. *Nature*, *429*(6988), 154. doi:10.1038/429154a

Pallante, P., Visone, R., Ferracin, M., Ferraro, A., Berlingieri, M. T., Troncone, G., Chiappetta, G., et al. (2006). MicroRNA deregulation in human thyroid papillary carcinomas. *Endocrine-Related Cancer*, *13*(2), 497 -508. doi:10.1677/erc.1.01209

Pang, K. C. (2004). RNAdb--a comprehensive mammalian noncoding RNA database. *Nucleic Acids Research*, *33*, D125-D130. doi:10.1093/nar/gki089

Pang, K. C., Stephen, S., Dinger, M. E., Engstrom, P. G., Lenhard, B., & Mattick, J. S. (2007). RNAdb 2.0--an expanded database of mammalian non-coding RNAs. *Nucleic Acids Research*, *35*, D178-D182. doi:10.1093/nar/gkl926

Papaioannou, M. D., & Nef, S. (2010). microRNAs in the Testis: Building Up Male Fertility. *J Androl*, *31*(1), 26-33. doi:10.2164/jandrol.109.008128

Papaioannou, M. D., Pitetti, J.-L., Ro, S., Park, C., Aubry, F., Schaad, O., Vejnar, C. E., et al. (2009). Sertoli cell Dicer is essential for spermatogenesis in mice. *Developmental Biology*, *326*(1), 250-259. doi:10.1016/j.ydbio.2008.11.011

Paterson, A. H., Bowers, J. E., & Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(26), 9903 -9908. doi:10.1073/pnas.0307901101

Pearse, R. V., Drolet, D. W., Kalla, K. A., Hooshmand, F., Bermingham, J. R., & Rosenfeld, M. G. (1997). Reduced fertility in mice deficient for the POU protein□sperm-1. *Proceedings of the National Academy of Sciences*, *94*(14), 7555 -7560.

Peltier, H. J., & Latham, G. J. (2008). Normalization of microRNA expression levels in quantitative RT-PCR assays: Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA*, *14*(5), 844 -852. doi:10.1261/rna.939908

Perkin Elmer, Inc. (2011, November 5). GeneSifter Analysis Edition (GSAE). Retrieved November 5, 2011, from http://www.geospiza.com/Products/AnalysisEdition.shtml

Petrov, D. A., Schutzman, J. L., Hartl, D. L., & Lozovskaya, E. R. (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis. *Proceedings of the National Academy of Sciences*, *92*(17), 8050 -8054.

Pfeiffer, P., Goedecke, W., & Obe, G. (2000). Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis*, *15*(4), 289 -302. doi:10.1093/mutage/15.4.289

Picker, M. D. (1985). Hybridization and Habitat Selection in *Xenopus gilli* and *Xenopus laevis* in the South-Western Cape Province. *Copeia*, *1985*(3), 574-580. doi:10.2307/1444746

Picker, Mike D., & de Villiers, A. L. (1989). The distribution and conservation status of *Xenopus gilli* (Anura: Pipidae). *Biological Conservation*, *49*(3), 169-183. doi:10.1016/0006-3207(89)90034-7

Reed, S. C., & Stanley, H. P. (1972). Fine structure of spermatogenesis in the South African clawed toad *Xenopus laevis daudin*. *Journal of Ultrastructure Research*, *41*(3-4), 277-295. doi:10.1016/S0022-5320(72)90070-6

Reik, W., Dean, W., & Walter, J. (2001). Epigenetic Reprogramming in Mammalian Development. *Science*, *293*(5532), 1089 -1093. doi:10.1126/science.1063443

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., & Pillai, R. S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol*, *16*(6), 639-646. doi:10.1038/nsmb.1615

Ro, S., Park, C., Song, R., Nguyen, D., Jin, J., Sanders, K. M., McCarrey, J. R., et al. (2007). Cloning and expression profiling of testis-expressed piRNA-like RNAs. *RNA*, *13*(10), 1693 -1702. doi:10.1261/rna.640307

Ro, S., Song, R., Park, C., Zheng, H., Sanders, K. M., & Yan, W. (2007). Cloning and expression profiling of small RNAs expressed in the mouse ovary. *RNA*, *13*(12), 2366 -2380. doi:10.1261/rna.754207

Russo, V. E. a. (1996). *Epigenetic Mechanisms of Gene Regulation*. Cold Spring Harbor Laboratory Press.

Schilthuizen, M., Giesbers, M. C. W. G., & Beukeboom, L. W. (2011). Haldane's rule in the 21st century. *Heredity*, *107*(2), 95-102.

Schlecht, U., Demougin, P., Koch, R., Hermida, L., Wiederkehr, C., Descombes, P., Pineau, C., et al. (2004). Expression Profiling of Mammalian Male Meiosis and Gametogenesis Identifies Novel Candidate Genes for Roles in the Regulation of Fertility. *Molecular Biology of the Cell*, *15*(3), 1031 -1043. doi:10.1091/mbc.E03-10-0762

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., et al. (2009). The TDRD9-MIWI2 Complex Is Essential for piRNA-Mediated Retrotransposon Silencing in the Mouse Male Germline. *Developmental Cell*, *17*(6), 775-787. doi:10.1016/j.devcel.2009.10.012

Siegfried, K. R., & Nüsslein-Volhard, C. (2008). Germ line control of female sex determination in zebrafish. *Developmental Biology*, *324*(2), 277-287. doi:10.1016/j.ydbio.2008.09.025

Siomi, M. C., Mannen, T., & Siomi, H. (2010). How does the Royal Family of Tudor rule the PIWI-interacting RNA pathway? *Genes & Development*, *24*(7), 636 -646. doi:10.1101/gad.1899210

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., & Ruvkun, G. (2000). The lin-41 RBCC Gene Acts in the C. elegans Heterochronic Pathway between the let-7 Regulatory RNA and the LIN-29 Transcription Factor. *Molecular Cell*, *5*(4), 659-669. doi:10.1016/S1097-2765(00)80245-2

Sperling, E. A., Vinther, J., Moy, V. N., Wheeler, B. M., Sémon, M., Briggs, D. E. G., & Peterson, K. J. (2009). MicroRNAs resolve an apparent conflict between annelid systematics and their fossil record. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1677), 4315 -4322. doi:10.1098/rspb.2009.1340

Stancheva, I., & Meehan, R. R. (2000). Transient depletion of xDnmt1 leads to premature gene activation in *Xenopus* embryos. *Genes & Development*, *14*(3), 313 -327. doi:10.1101/gad.14.3.313

Stancheva, I., El-Maarri, O., Walter, J., Niveleau, A., & Meehan, R. R. (2002). DNA Methylation at Promoter Regions Regulates the Timing of Gene Activation in *Xenopus laevis* Embryos. *Developmental Biology*, *243*(1), 155-165. doi:10.1006/dbio.2001.0560

Stancheva, I., Hensey, C., & Meehan, R. R. (2001). Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in *Xenopus* embryos. *EMBO J*, *20*(8), 1963-1973. doi:10.1093/emboj/20.8.1963

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., & Cohen, S. M. (2005). Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution. *Cell*, *123*(6), 1133-1146. doi:10.1016/j.cell.2005.11.023

Susan, G. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in Genetics*, *21*(7), 399-404. doi:10.1016/j.tig.2005.05.008

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, *453*(7194), 534-538. doi:10.1038/nature06904

Thomson, T., & Lin, H. (2009). The Biogenesis and Function PIWI Proteins and piRNAs: Progress and Prospect. *Annual review of cell and developmental biology*, *25*, 355-376. doi:10.1146/annurev.cellbio.24.110707.175327

Tymowska, J., & Fischberg, M. (1982). A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species *Xenopus epitropicalis* Fischberg and Picard with those of *Xenopus tropicalis* Gray (Anura, Pipidae). *Cytogenetic and Genome Research*, *34*, 149-157. doi:10.1159/000131803

Ungerer, M. C., Strakosh, S. C., & Zhen, Y. (2006). Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology: CB*, *16*(20), R872-873. doi:10.1016/j.cub.2006.09.020

Unhavaithaya, Y., Hao, Y., Beyret, E., Yin, H., Kuramochi-Miyagawa, S., Nakano, T., & Lin, H. (2009). MILI, a PIWI-interacting RNA-binding Protein, Is Required for Germ Line Stem Cell Self-renewal and Appears to Positively Regulate Translation. *Journal of Biological Chemistry*, *284*(10), 6507 -6519. doi:10.1074/jbc.M809104200

Vagin, V. V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., et al. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes & Development*, *23*(15), 1749 -1762. doi:10.1101/gad.1814809

Vasileva, A., Tiedau, D., Firooznia, A., Müller-Reichert, T., & Jessberger, R. (2009). Tudor domain protein Tdrd6 is required for spermiogenesis, chromatoid body architecture and regulation of miRNA expression. *Current biology□: CB*, *19*(8), 630-639. doi:10.1016/j.cub.2009.02.047

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. S., & Moazed, D. (2004). RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. *Science*, *303*(5658), 672 -676. doi:10.1126/science.1093686

Vieira, J., Vieira, C. P., Hartl, D. L., & Lozovskaya, E. R. (1998). Factors Contributing to the Hybrid Dysgenesis Syndrome in Drosophila Virilis. *Genetics Research*, *71*(02), 109-117. doi:null

Vigodner, M., & Morris, P. L. (2005). Testicular expression of small ubiquitin-related modifier-1 (SUMO-1) supports multiple roles in spermatogenesis: Silencing of sex chromosomes in spermatocytes, spermatid microtubule nucleation, and nuclear reshaping. *Developmental Biology*, *282*(2), 480-492. doi:10.1016/j.ydbio.2005.03.034

Villalpando, I, & Merchant-Iarios, H. (1990). Determination of the sensitive stages for gonadal sex-reversal in *Xenopus laevis* tadpoles. *Int. J. Dev. Bio.*, *34*, 281-285.

Wahid, F., Khan, T., Hwang, K., & Kim, Y. (2010). Piwi-interacting RNAs (piRNAs) in animals: The story so far. *African Journal of Biotechnology*, *8*(17). Retrieved from http://www.ajol.info/index.php/ajb/article/view/62111

Wang, Jianquan, Saxe, J. P., Tanaka, T., Chuma, S., & Lin, H. (2009). Mili Interacts with Tudor Domain-Containing Protein 1 in Regulating Spermatogenesis. *Current Biology*, *19*(8), 640-644. doi:10.1016/j.cub.2009.02.061

Wang, Y., Lee, A. T. C., Ma, J. Z. I., Wang, J., Ren, J., Yang, Y., Tantoso, E., et al. (2008). Profiling MicroRNA Expression in Hepatocellular Carcinoma Reveals MicroRNA-224 Up-regulation and

Apoptosis Inhibitor-5 as a MicroRNA-224-specific Target. *Journal of Biological Chemistry*, *283*(19), 13205 -13215. doi:10.1074/jbc.M707629200

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, *453*(7194), 539-543. doi:10.1038/nature06908

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. A., Holstein, T. W., Heber, S., & Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution & Development*, *11*(1), 50-68. doi:10.1111/j.1525-142X.2008.00302.x

Wilczynska, A., Minshall, N., Armisen, J., Miska, E. A., & Standart, N. (2009). Two Piwi proteins, Xiwi and Xili, are expressed in the *Xenopus* female germline. *RNA*, *15*(2), 337 -345. doi:10.1261/rna.1422509

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*, *2*(5), 333-341. doi:10.1038/35072009

Wu, C. H. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*, *34*, D187-D191. doi:10.1093/nar/gkj161

Wu, C.-I., & Davis, A. W. (1993). Evolution of Postmating Reproductive Isolation: The Composite Nature of Haldane's Rule and Its Genetic Bases. *The American Naturalist*, *142*(2), 187-212.

Wu, C.-I., Johnson, N. A., & Palopoli, M. F. (1996). Haldane's rule and its legacy: Why are there so many sterile males? *Trends in Ecology & Evolution*, *11*(7), 281-284. doi:10.1016/0169-5347(96)10033-1

Xu, E. Y., Moore, F. L., & Pera, R. A. R. (2001). A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans. *Proceedings of the National Academy of Sciences*, *98*(13), 7414 -7419. doi:10.1073/pnas.131090498

Xu, P., Vernooy, S. Y., Guo, M., & Hay, B. A. (2003). The Drosophila MicroRNA Mir-14 Suppresses Cell Death and Is Required for Normal Fat Metabolism. *Current Biology*, *13*(9), 790-795. doi:10.1016/S0960-9822(03)00250-1

Xu, Z., & Au, S. W. N. (2005). Mapping residues of SUMO precursors essential in differential maturation by SUMO-specific protease, SENP1. *Biochemical Journal*, *386*(Pt 2), 325-330. doi:10.1042/BJ20041210

Yan, Z., Hu, H. Y., Jiang, X., Maierhofer, V., Neb, E., He, L., Hu, Y., et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Research*. doi:10.1093/nar/gkr298

Yang, J., Medvedev, S., Yu, J., Tang, L. C., Agno, J. E., Matzuk, M. M., Schultz, R. M., et al. (2005). Absence of the DNA-/RNA-binding protein MSY2 results in male and female infertility. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(16), 5755 -5760. doi:10.1073/pnas.0408718102

Yasuhara, J. C., DeCrease, C. H., & Wakimoto, B. T. (2005). Evolution of heterochromatic genes of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10958 -10963. doi:10.1073/pnas.0503424102

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*(8), 335-340. doi:10.1016/S0168-9525(97)01181-5

Young, J. J., Cherone, J. M., Doyon, Y., Ankoudinova, I., Faraji, F. M., Lee, A. H., Ngo, C., et al. (2011). Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1102030108

Zhang, D., Duarte-Guterman, P., Langlois, V. S., & Trudeau, V. L. (2010). Temporal expression and steroidal regulation of piRNA pathway genes (mael, piwi, vasa) during *Silurana (Xenopus)* tropicalis embryogenesis and early larval development. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *152*(2), 202-206. doi:10.1016/j.cbpc.2010.04.005

Zhao, Y., Ransom, J. F., Li, A., Vedantham, V., von Drehle, M., Muth, A. N., Tsuchihashi, T., et al. (2007). Dysregulation of Cardiogenesis, Cardiac Conduction, and Cell Cycle in Mice Lacking miRNA-1-2. *Cell*, *129*(2), 303-317. doi:10.1016/j.cell.2007.03.030

Zhou, X., Zuo, Z., Zhou, F., Zhao, W., Sakaguchi, Y., Suzuki, T., Suzuki, T., et al. (2010). Profiling Sex-Specific piRNAs in Zebrafish. *Genetics*, *186*(4), 1175 -1185. doi:10.1534/genetics.110.122234

BIOGRAPHICAL INFORMATION

M.J. Madison-Villar is a first generation college graduate from a Jehovah's Witness background. Her greatest rebellion (so far) is having gone to college (rather than popping out God-fearing babies) where she dared to learn about, and graciously accept, the truth of evolution; her most interesting accomplishment is hitchhiking from Denali National Park, AK to Missoula, MT (or perhaps from Prague, Czech Republic to Garmisch, Germany while unable to speak to most of her rides?). She has resided in a dry cabin whose outhouse was not placed over a hole into the ground; warded off bears with waving of hands and shouting of "hey bear!"; slept in an REI brand tent in a city park outside of some random town in SE Italy; lived in Southern California without getting a boob or nose job; and has yet to say "y'all" without being entirely sarcastic. When not mashing up the gonads of small animals in eppindorf tubes or laboring to find the meaning of the fruits from said smashed gonads, she can often be found lying on her couch (or bed…or floor) with two small dogs, staring at the celing (or television), and dreaming of running away to Belize (or anywhere else with an ocean and trees in which to live), at a coffee shop, or enjoying the outdoors.