DISSERTATION

THE EVOLUTION OF PLASTICITY IN THE TRANSCRIPTOME OF THE TRINIDADIAN GUPPY

Submitted by

Miles Whedbee

Department of Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2023

Doctoral Committee:

Advisor: Kim Hoke

Dan Sloan
Taiowa Montgomery
Asa Ben-Hur

ABSTRACT

THE EVOLUTION OF PLASTICITY IN THE TRANSCRIPTOME OF THE TRINIDADIAN GUPPY

Phenotypic plasticity is a ubiquitous feature of all living systems, and there is much interest in how plasticity influences long term evolutionary trajectories. One of the major complications with modeling evolutionary trajectories is that plasticity itself is known to evolve. The evolution of plasticity has mainly been focused on at the level of the whole organism, and it is unclear if plasticity at all levels of biological organization evolve. Models that assume no generational change in plasticity may be overly simplistic; a more nuanced approach could incorporate the evolution of plasticity into the modeling. A first step towards this end is to determine what levels of biological organization plasticity evolves, and then to determine if there are predictable patterns of evolved plasticity.

RNA is an intermediate to DNA and protein, that can undergo changes in response to environmental conditions, thereby modifying the genetic information passed on to non-coding RNAs, functional RNAs, and proteins. Responses to environment include both changes in abundance of RNAs, as well as changes to the composition of the molecules. This dissertation focuses on the evolution of plasticity within the transcriptome of *Poecilia reticulata* (Trinidadian guppy). One of the major known regulators of transcript abundance are small RNAs (sRNAs). Micro RNAs (miRNAs), are a specific type of sRNA that bind transcripts, typically leading to translational silencing. We investigated two forms of plasticity, an abundance measure of plasticity (miRNA differential expression), and a compositional measure of plasticity (A-to-I

RNA editing). A-to-I RNA editing is the chemical nucleotide change from adenosine to inosine, catalyzed by the enzyme ADAR.

We first produced a set of miRNAs in guppies, and confirmed the presence of key biogenesis pathway components, i.e. argonaute proteins in the genome. Tissue-specific miRNA expression patterns were identified for three tissues in Poecilia reticulata (Trinidadian guppy), brain, ovary and testis. We found most discovered miRNAs were located in intergenic regions of the genome. Some miRNAs matched known miRBase sequences, while others were considered novel guppy miRNAs. We observed miRNAs expressed from tandem clusters and analyzed piRNA distribution in ovary samples. This study provides important insights into guppy small RNA expression, laying the groundwork for future investigations into their regulatory roles.

The 3rd chapter of this dissertation revealed many miRNAs with differential expression (DE), including population main effects, rearing condition, and their interactions. Population DE miRNAs showed a wide range of expression levels. Rearing condition main effects were (slightly) less common. We identified miRNAs with evolved expression plasticity, distributed across four categories: reversed, evolved plastic, assimilated, and accommodated. Both populations showed similar numbers of miRNAs exhibiting plasticity.

In the final chapter of this dissertation we characterized the "editome" of guppies. The majority of the edits were consistent with A-to-I editing, with a smaller proportion of C-to-U edits. The intragenic edits were distributed among a number of genes. However, there were no significant differences in editing between populations, rearing conditions, or their interaction. This dissertation revealed significant miRNA expression differences and provided insights into A-to-I editing patterns in guppies.

DEDICATION

I dedicate this dissertation to my mother, the unwavering supporter of my educational journey and the greatest champion of my academic pursuits. Her love, encouragement, and belief in me have been the driving force behind my persistence, and I am forever grateful for her guidance and inspiration.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

That plants and animals have evolved sophisticated mechanisms to regulate gene expression, enabling them to adapt to complex environmental conditions and drive the evolution of multicellular organisms, is a truism (Chen & Rajewsky, 2007). These regulatory mechanisms involve the interplay of multi-transcription factor complexes, microRNAs (miRNAs), and other small non-coding RNAs (sRNAs), which collectively regulate gene expression at the transcriptional and translational levels (Ambros, 2004; Bartel, 2004; Chen & Rajewsky, 2007). The increasing number of transcriptomic studies e.g. (Ghalambor et al., 2015; Manousaki et al., 2013; Pankey, Minin, Imholte, Suchard, & Oakley, 2014) points to the interest in exploring gene expression patterns associated with evolution.

The traditional "central dogma" of molecular biology states that genes primarily code for proteins, and genetic information is mainly carried out in cells by proteins. Consequently, biologists have focused on genetic complexity, e.g. total number of genes in the genome, and the generation of new proteins as key drivers of evolution. However, studies have shown both phylogenetically basal organisms and complex animals like vertebrates possess a majority of protein-coding gene families involved in processes such as transcription and signal transduction, despite differences in tissue and organ complexity (Mattick, 2004; Sempere, Cole, Mcpeek, & Peterson, 2006; Technau et al., 2005). Therefore, the distribution of protein-coding genes across taxa does not correlate with the significant increase in morphological complexity seen in animal evolution. This has led to the hypothesis that the gene expression in complex metazoan genomes necessitates additional regulatory mechanisms, including alternative splicing and non-coding RNAs (Mattick, 2004; Mattick & Makunin, 2006).

Investigating the influence of environmental sensitivity on evolution poses challenges, particularly in linking transcriptome states to overall organismal phenotypes. This involves not only understanding how genotypes impact chromatin states but also how post-transcriptional regulatory mechanisms govern the transcriptome. Small RNAs play a role in modifying DNA methylation and regulating mRNA degradation or translation (V. N. Kim, Han, & Siomi, 2009; Krol, Loedige, & Filipowicz, 2010; Zamore & Ghildiyal, 2009). They serve as important mediators in controlling mRNA and protein abundance, and also help buffer mRNA and protein levels against environmental fluctuations (Wu et al. 2009).

MicroRNAs (miRNAs), a specific class of endogenous sRNAs approximately 22 nucleotides in length, have been extensively studied in various biological processes, including developmental biology, immunology, cardiovascular biology, cancer, and neurobiology (Kawahara, 2008; Pasquinelli & Ruvkun, 2003; Reddy, 2015). They are critical regulators of gene expression, functioning primarily through the binding of target mRNAs and subsequent repression of protein synthesis or mRNA degradation (Bartel, 2009). MiRNA production begins with the processing of RNA polymerase II/III transcripts, either after or during transcription (Ha & Kim, 2014). Approximately half of known miRNAs originate from within genes, mainly from introns and a small number of exons, while the rest are transcribed independently and have their own promoter regions (De Rie et al., 2017; Y. K. Kim & Kim, 2007). MiRNAs can also be transcribed together as clusters, forming families when they share similar seed regions (Tanzer & Stadler, 2004). The biogenesis of miRNAs can be categorized into canonical and non-canonical pathways.

The canonical pathway is the primary route for miRNA biogenesis. In this pathway, primary miRNAs (pri-miRNAs) are transcribed from genes and processed into precursor miRNAs (pre-miRNAs) by the microprocessor complex, which includes DGCR8 and Drosha

(Denli, Tops, Plasterk, Ketting, & Hannon, 2004). DGCR8 recognizes specific motifs within the pri-miRNA (Alarcón, Lee, Goodarzi, Halberg, & Tavazoie, 2015), while Drosha cleaves the hairpin structure of pri-miRNA to generate pre-miRNAs. Exportin 5 then transports the pre-miRNAs to the cytoplasm, where Dicer removes the terminal loop to produce mature miRNA duplexes (Denli et al., 2004; Okada et al., 2009). The directionality of the miRNA strand determines its name (5' or 3'), and both strands can be loaded into AGO proteins (Yoda et al., 2009). The selection of the guide strand is influenced by factors such as thermodynamic stability and the presence of a 5' uracil. The unloaded passenger strand is degraded, resulting in a strand bias (Khvorova, Reynolds, & Jayasena, 2003). MiRNAs play a crucial role in gene silencing by guiding Argonaute (AGO) proteins to specific sites in the 3′ untranslated region (UTR) of mRNAs. Once loaded onto AGO, the miRNA forms the targeting module of the miRNA-induced silencing complex (miRISC) (Swarts et al., 2014). This miRISC complex then facilitates translation repression and degradation of the targeted mRNAs, thereby regulating gene expression (Jonas & Izaurralde, 2015).

Although miRNAs have been widely investigated in the context of individual biological systems, their contribution to phenotypic evolution remains relatively unexplored (Kawahara, 2008; Pasquinelli & Ruvkun, 2003; Reddy, 2015). Understanding how genetic variation in miRNAs and miRNA-binding sites influences phenotypic evolution requires comparative studies across diverse taxa and natural environments. However, current miRNA databases have limited representation of metazoan species and tissue diversity. Therefore, expanding the catalog of miRNA sequences is crucial, necessitating the characterization of the miRNAome in different organisms and tissues (e.g., miRbase and mirGeneDB). Chapter 2 of this dissertation provides the first miRNAome for *Poecilia reticulata* (Trinidadian guppy).

To investigate the role of miRNAs in phenotypic evolution, comparative genomics studies have been conducted across various organisms, including plants, animals, and fungi (Berezikov, 2011; Friedman, Farh, Burge, & Bartel, 2009). These studies have revealed conserved miRNA families and their targets, highlighting their functional importance in development and other biological processes (Friedman et al., 2009; Bartel, 2009). For instance, studies in animals have demonstrated that miRNAs play critical roles in regulating early embryonic development, organogenesis, and tissue homeostasis (Bushati & Cohen, 2007). Furthermore, it has been suggested that changes in miRNA genes and miRNA target sites could contribute to phenotypic variation and adaptation in natural populations (Mencía et al., 2009). Single nucleotide polymorphisms (SNPs) within miRNA genes or their target sites can potentially affect miRNA biogenesis or the binding affinity of miRNAs to their targets, thereby influencing gene expression and phenotypic traits (Mencia et al., 2009). Understanding the functional consequences of such genetic variations in miRNAs and their targets is vital for deciphering the molecular basis of evolutionary processes.

MiRNAs are essential regulators of gene expression with well-established roles in various biological processes. However, their contribution to phenotypic evolution and adaptation in diverse organisms and natural environments is not fully understood. Comparative studies across taxa and the expansion of miRNA catalogs are crucial for unraveling the functional significance of miRNAs in phenotypic evolution. Investigating the fluctuations in levels of miRNAs will help to shed light on (potentially) a major molecular mechanism underlying evolutionary processes. Ultimately, a comprehensive understanding of miRNA-mediated regulatory networks will enhance our knowledge of biological diversity and the adaptive potential of organisms. In addition to miRNA mediated regulation of RNA abundances, transcriptome plasticity can also be characterized by compositional changes to RNAs in response

to varying environmental conditions. RNA editing refers to molecular processes that modify the information content of RNA molecules. These processes have been observed in various cellular compartments, including the nucleus, mitochondria, and plastids (Brennicke, Marchfelder, & Binder, 1999). Many of the RNA editing mechanisms are believed to be relatively recent evolutionary events that have emerged independently. The diversity of RNA editing mechanisms encompasses nucleoside modifications, such as C to U and A to I deaminations, as well as non-templated nucleotide additions and insertions (Brennicke et al., 1999).

A-to-I RNA editing, or the conversion of adenosine to inosine, is the most common form of RNA editing in metazoans, and can generate nucleotide changes to mRNAs as well as play a crucial role in gene expression regulation (Farajollahi & Maas, 2010; Maydanovych & Beal, 2006). A-to-I editing is mediated by adenosine deaminases called ADARs (Nishikura, 2010). Editing has been extensively studied in various organisms and has been associated with diverse biological processes, including alternative splicing, miRNA binding, and transcriptome diversification (Borchert et al., 2009; Fumagalli et al., 2015; Rueter, Dawson, & Emeson, 1999).

In humans, A-to-I editing accounts for the majority of editing events, particularly in noncoding regions (Bahn et al., 2012; Ramaswami, Lin, Piskol, Tan, & Davis, 2012; Zhang & Xiao, 2015). Aberrant editing at specific sites has been linked to phenotypic outcomes and human diseases (Anne-Laure, 2010; Boris Zinshteyn, 2010; Zipeto, Jiang, Melese, & Jamieson, 2015). For example RNA editing may be involved in amyotrophic lateral sclerosis (ALS) by affecting the excessive influx of calcium ions through glutamate receptors, leading to motor neuron death (Maas, Kawahara, Tamburro, & Nishikura, 2006). Specifically, the editing of the Q/R position in the GluR-2 receptor determines its low permeability to calcium ions, which is observed in the majority of neurons expressing AMPA-type glutamate receptors (Kwak & Kawahara, 2005).

While several adaptive explanations have been proposed for A-to-I editing, its precise impact on natural phenotypic variation is still a subject of investigation. To explore the biological implications of A-to-I editing in a natural context, we investigated its variation in response to a major ecological factor (predation pressure), as well as genetic background using the Trinidadian guppy system. Guppies from different environments exhibit local adaptations and developmental plasticity in response to predators. By leveraging the natural phenotypic variation and parallel lineages, we aim to determine if A-to-I editing shows consistent differences based on these factors. Our analysis of A-to-I editing in the whole brain did not reveal associations with population or rearing condition. Moreover, we considered two separate lineages to study parallel evolution, but differential editing was not detected in either group. Consequently, we conclude that A-to-I editing may not play a major role in influencing developmental plasticity or population divergence in guppies.

Comparative studies of miRNAs are more useful when conducted across a diverse array of organisms, which requires annotated miRNAs. This dissertation produced the first Trinidadian guppy miRNAome, adding to the growing database of miRNAomes. We revealed miRNA expression patterns demonstrate similar evolved plasticity as observed in mRNA. We suggest that A-to-I editing, (in terms of edit levels), are not significantly associated with phenotypic variation of populations, rearing conditions, or their interactions. The exploration of miRNAs and A-to-I editing in the context of diverse organisms and natural environments remains critical for understanding the molecular basis of evolutionary processes and the mechanisms underlying adaptation.

CHAPTER 1 WORKS CITED

Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N., & Tavazoie, S. F. (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature 2015 519:7544*, *519*(7544), 482–485. https://doi.org/10.1038/nature14281

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*(7006), 350–355. https://doi.org/10.1038/nature02871

Anne-Laure. (2010). *RNA Biology Plant RNA editing*. https://doi.org/10.4161/rna.7.2.11343

Bahn, J. H., Lee, J.-H., Li, G., Greer, C., Peng, G., & Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research*. https://doi.org/10.1101/gr.124107.111

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function Review. *Biosensors and Bioelectronics*, *116*, 281–297. https://doi.org/10.1016/j.bios.2017.04.001

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, *136*(2), 215–233. https://doi.org/10.1016/j.cell.2009.01.002

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics*, *12*(12), 846–860. https://doi.org/10.1038/nrg3079

Borchert, G. M., Gilmore, B. I., Spengler, R. M., Xing, Y., Lanier, W., Bhattacharya, D., & Davidson, B. L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. *Human Molecular Genetics*, *18*(24), 4801–4807. https://doi.org/10.1093/HMG/DDP443

Boris Zinshteyn, K. N. (2010). NIH Public Access. *Wiley Interdiscip Rev Syst Biol Med*, *1*(2), 202–209. https://doi.org/10.1002/wsbm.10.Adenosine-to-inosine

Brennicke, A., Marchfelder, A., & Binder, S. (1999). RNA editing. *FEMS Microbiology Reviews*, *23*(3), 297–316. https://doi.org/10.1111/J.1574-6976.1999.TB00401.X

Bushati, N., & Cohen, S. M. (2007). microRNA Functions. *Https://Doi.Org/10.1146/Annurev.Cellbio.23.090506.123406*, *23*, 175–205. https://doi.org/10.1146/ANNUREV.CELLBIO.23.090506.123406

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, *8*(2), 93–103. https://doi.org/10.1038/nrg1990

De Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., … De Hoon, M. J. L. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnology 2017 35:9*, *35*(9), 872–878. https://doi.org/10.1038/nbt.3947

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature 2004 432:7014*, *432*(7014), 231–235. https://doi.org/10.1038/nature03049

Farajollahi, S., & Maas, S. (2010). Molecular diversity through RNA editing: a balancing act. *Trends in Genetics*, *26*(5), 221–230.

Friedman, R. C., Farh, K. K. H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92. https://doi.org/10.1101/GR.082701.108

Fumagalli, D., Gacquer, D., Rothé, F., Lefort, A., Libert, F., Brown, D., … Detours, V. (2015). Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. *Cell Reports*, *13*(2), 277–289. https://doi.org/10.1016/J.CELREP.2015.09.032

Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, *525*(7569), 372–375. https://doi.org/10.1038/nature15256

Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, *15*(8), 509–524. https://doi.org/10.1038/nrm3838

Jonas, S., & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics 2015 16:7*, *16*(7), 421–433. https://doi.org/10.1038/nrg3965

Kawahara, Y. (2008). [MicroRNAs in neurobiology]. *Brain and Nerve = Shinkei Kenkyu No Shinpo*, *60*(12), 1437–1444. Retrieved from https://europepmc.org/article/med/19110755

Khvorova, A., Reynolds, A., & Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, *115*(2), 209–216. https://doi.org/10.1016/S0092-8674(03)00801-8

Kim, V. N., Han, J., & Siomi, M. C. (2009). *Biogenesis of small RNAs in animals*. *10*(fEBRuARy). https://doi.org/10.1038/nrm2632

Kim, Y. K., & Kim, V. N. (2007). Processing of intronic microRNAs. *The EMBO Journal*, *26*(3), 775–783. https://doi.org/10.1038/SJ.EMBOJ.7601512

Krol, J., Loedige, I., & Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics*, *11*(9), 597–610. https://doi.org/10.1038/nrg2843

Kwak, S., & Kawahara, Y. (2005). Deficient RNA editing of GluR2 and neuronal death in amyotropic lateral sclerosis. *Journal of Molecular Medicine*, *83*(2), 110–120. https://doi.org/10.1007/S00109-004-0599-Z/FIGURES/7

Maas, S., Kawahara, Y., Tamburro, K. M., & Nishikura, K. (2006). *RNA Biology A-to-I RNA Editing and Human Disease*. https://doi.org/10.4161/rna.3.1.2495

Manousaki, T., Hull, P. M., Kusche, H., MacHado-Schiaffino, G., Franchini, P., Harrod, C., … Meyer, A. (2013). Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, *22*(3), 650–669. https://doi.org/10.1111/MEC.12034

Mattick, J. S. (2004). RNA regulation: a new genetics? *Nature Reviews Genetics 2004 5:4*, *5*(4), 316–323. https://doi.org/10.1038/nrg1321

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15*(suppl_1), R17–R29. https://doi.org/10.1093/HMG/DDL046

Maydanovych, O., & Beal, P. A. (2006). Breaking the central dogma by RNA editing. *Chemical Reviews*, *106*(8), 3397–3411. https://doi.org/10.1021/CR050314A/ASSET/IMAGES/LARGE/CR050314AF00011.JPEG

Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., … Moreno-Pelayo, M. A. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics 2009 41:5*, *41*(5), 609–613. https://doi.org/10.1038/ng.355

Nishikura, K. (2010). Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry*, *79*, 321. https://doi.org/10.1146/ANNUREV-BIOCHEM-060208-105251

Okada, C., Yamashita, E., Lee, S. J., Shibata, S., Katahira, J., Nakagawa, A., … Tsukihara, T. (2009). A high-Resolution structure of the pre-microrna nuclear export machinery. *Science*, *326*(5957), 1275–1279. https://doi.org/10.1126/SCIENCE.1178705/SUPPL_FILE/OKADA.SOM.PDF

Pankey, M. S., Minin, V. N., Imholte, G. C., Suchard, M. A., & Oakley, T. H. (2014). Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(44), E4736–E4742. https://doi.org/10.1073/PNAS.1416574111/SUPPL_FILE/PNAS.1416574111.SAPP.PDF

Pasquinelli, A. E., & Ruvkun, G. (2003). Control of Developmental Timing by MicroRNAs and Their Targets. *Https://Doi.Org/10.1146/Annurev.Cellbio.18.012502.105832*, *18*, 495–513. https://doi.org/10.1146/ANNUREV.CELLBIO.18.012502.105832

Ramaswami, G., Lin, W., Piskol, R., Tan, M. H., & Davis, C. (2012). Accurate identification of human Aluand non-AluRNA editing sites. *Nature Protocols*, *9*(6), 579–581. https://doi.org/10.1038/nmeth.1982.Accurate

Reddy, K. B. (2015). MicroRNA (miRNA) in cancer. *Cancer Cell International*, *15*(1), 1–6. https://doi.org/10.1186/S12935-015-0185-1/FIGURES/1

Rueter, S. M., Dawson, T. R., & Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. *Nature*, *399*(May), 75–80. https://doi.org/10.1038/19992

Sempere, L. F., Cole, C. N., Mcpeek, M. A., & Peterson, K. J. (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *306B*(6), 575–588. https://doi.org/10.1002/JEZ.B.21118

Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., … Van Der Oost, J. (2014). The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology 2014 21:9*, *21*(9), 743–753. https://doi.org/10.1038/nsmb.2879

Tanzer, A., & Stadler, P. F. (2004). Molecular Evolution of a MicroRNA Cluster. *JMB*. https://doi.org/10.1016/j.jmb.2004.03.065

Technau, U., Rudd, S., Maxwell, P., Gordon, P. M. K., Saina, M., Grasso, L. C., … Miller, D. J. (2005). Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends in Genetics*, *21*(12), 633–639. https://doi.org/10.1016/j.tig.2005.09.007

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q., & Tomari, Y. (2009). ATP-dependent human RISC assembly pathways. *Nature Structural & Molecular Biology 2009 17:1*, *17*(1), 17–23. https://doi.org/10.1038/nsmb.1733

Zamore, P. D., & Ghildiyal, M. (2009). Small silencing RNAs: an expanding universe. *Nature Reviews Genetics*, *10*(2), 94–108. https://doi.org/10.1038/nrg2504.Small

Zhang, Q., & Xiao, X. (2015). Genome sequence–independent identification of RNA editing sites. *Nature Methods 2015 12:4*, *12*(4), 347–350. https://doi.org/10.1038/nmeth.3314

Zipeto, M. A., Jiang, Q., Melese, E., & Jamieson, C. H. M. (2015). RNA rewriting, recoding, and rewiring in human disease. *Trends in Molecular Medicine*, *21*(9), 549–559. https://doi.org/10.1016/J.MOLMED.2015.07.001

CHAPTER 2: MIRNA AND PIRNA RNA CHARACTERIZATION IN POECILIA RETICULATA

Introduction

Plants and animals evolved complex, multilayered mechanisms that regulate gene expression (Chen & Rajewsky, 2007). The emergence of complex, multicellular organisms was accompanied by (and perhaps driven by) combinatorial control of gene expression through multi-transcription factor complexes and microRNAs (miRNAs), a class of small non-coding RNAs (sRNAs) that are ~21 nucleotides (nt) long and primarily function as translational repressors in plants and animals (Ambros, 2004; Bartel, 2004; Chen & Rajewsky, 2007). Transcriptional regulation by transcription factors and miRNA mediated regulation are two highly studied layers of transcriptional regulation.

Although much progress has been made in the understanding of the functioning of miRNAs in developmental biology (Pasquinelli & Ruvkun, 2003), immunology, cardiovascular biology, cancer (Reddy, 2015) and neurobiology (Kawahara, 2008); investigation into the connection between genetic variation in miRNAs or miRNA-binding sites and phenotypic evolution is still in its early stages. To understand how evolutionary changes in miRNA sequence and abundance contribute to phenotypic evolution, comparative studies will be necessary, on miRNAs across a broad set of taxa within their natural environments. A first step towards this end, is to characterize the transcriptome of the respective study organism, including the miRNAome, i.e. the set of expressed miRNAs in a tissue or organism. The current metazoan species and tissue diversity, represented in the miRNA databases is limited. This chapter adds a

miRNAome of three tissues in the teleost fish *Poecilia reticulata* to complement existing miRNA sequences in miRbase and mirGeneDB.

The basic shared mechanism of the sRNA classes is that they directly bind effector proteins, and are directed by sequence specificity, targeting RNA transcripts via base-pairing interactions. For all three sRNA classes, the core effector proteins are members of the Argonaute superfamily. miRNAs and small interfering RNAs (siRNAs) are both formed from double-stranded precursors, are typically ~22 nt in length, and have a wider set of physiological roles, and broader phylogenetic conservation compared to Piwi-interacting RNAs (piRNAs). PiRNAs are derived from single-stranded RNA precursors (Hirakata & Siomi, 2016), have lengths in the range of 24-35 nt (Hirakata & Siomi, 2016; Kim, Han, & Siomi, 2009), and are found in animals (Carthew & Sontheimer, 2009). The known functions of piRNAs are most clearly present in the germline, in which piRNAs associate with a distinct subset of Argonaute effector proteins, i.e. members of the Piwi lineage; in contrast, miRNAs and siRNAs associate with the Ago lineage. Transposable elements (TEs) are important components of eukaryotic genomes, but their mobilization often has detrimental effects on the host genome. To protect against this, host cells have developed genetic and epigenetic mechanisms to silence TEs. One such mechanism involves the Piwi-piRNA complex, which suppresses TEs in animal gonads by either cleaving TE transcripts in the cytoplasm or inducing specific chromatin modifications at TE sites in the nucleus (Yamanaka, Siomi, & Siomi, 2014).

Large scale non-coding DNA studies have shown miRNA and piRNA genomic loci are not randomly distributed across chromosomes; instead, for both they are usually found as multiple (clustered) sequences (Zhang, Zhang, & Su, 2009); although, the size and complexity of the clusters differ between miRNAs and piRNAs. MiRNA clusters are defined as a set of two or

more miRNAs that are transcribed from tandem miRNA genes, in the same orientation, and not separated by a transcriptional unit (Lai & Vera, 2013). Typical miRNA clusters consist of two or three miRNAs, but larger clusters such as the human miR-17-92 with six miRNAs have also been discovered. Clustered miRNA genes are often functionally related, and share sequence similarity (Zhang, Zhang, & Su, 2009). In Drosophila about ~50% of miRNA genes in the genome are clustered (Lai, Tomancak, Williams, & Rubin, 2003). PiRNAs, which are derived from genomic regions called piRNA clusters, predominantly originate from transposable elements (TEs) present within these clusters. As a result, the sequence of piRNAs obtained from these clusters can target and suppress not only the TEs within the clusters but also related TEs located elsewhere in the genome, acting as guide molecules to regulate TEs in trans (Yamanaka et al., 2014).

In the present study, we characterized the miRNAome of the Trinidadian guppy in whole brain, ovary and testis tissues, using next generation sequencing. Guppies have become a common model system for evolutionary studies. We called predicted miRNA sequences, i.e. precursor (pre-miRNA) and 5' and 3' mature sequences, with the goal of characterizing sequence and abundance variation within and between these three tissues. We perform a clustering analysis to determine what percentage of our discovered miRNAs were found in clusters. Our current miRNA characterization in the brain complements a recent study (Eva K. Fischer et al., 2021) in guppies, and adds brain tissue expression patterns from RNA-seq data in individuals of the same lineage (drainage) as this study, such that future studies could link guppy miRNAs to their respective mRNA targets.

Methods

MiRNA and piRNA processing enzyme evolutionary trees

We constructed evolutionary trees to visualize the relationships of the Drosha, Dicer and Argonaute family of proteins. We combined the Argonaute subfamilies Ago and Piwi into one tree, and two separate trees were constructed for Drosha and Dicer. We collected protein sequences by gene or protein name searches in the NCBI protein and gene databases for human, mouse, zebrafish, fruit fly, and guppy. We selected these organisms as they are common model organisms with ample information on their small RNA processing enzymes, i.e. Drosha, Dicer and Argonaute family proteins. We used NCBI's ortholog database to find orthologous proteins after a single organism's gene was located. For proteins with multiple isoforms, we selected the first isoform as the representative sequence. We made multiple sequence alignments using the EMBL-EBI online application, Multiple Sequence Comparison by Log-Expectation (MUSCLE). Maximum likelihood trees were constructed using IQ-TREE (Kalyaanamoorthy et al. 2017, Nguyen et al. 2015, Hoang et al. 2018) webserver, http://www.iqtree.org, and viewed and edited in iTOL, https://itol.embl.de/.

Husbandry

Our lab collected Trinidadian guppies in 2014 from both high-predation and low-predation localities in the Quare river drainage (J. F. Gilliam, Fraser, & Alkins-Koo, 1993; D. Reznick et al., 2001). We established unique families (20-25 family lines) in laboratory aquaculture from wild-caught gravid females (E. K. Fischer et al., 2016). The second-generation lab reared fish were established by crossing first-generation lab reared fish with unrelated (different family) first-generation fish from the same source population; siblings from the second-generation

matings were split into rearing environments with predator cues or without cues (E. K. Fischer et al., 2016). All experimental methods were approved by the Colorado State University Animal Care and Use Committee (Protocol #12-3818A).

Tissue collection, processing and RNA extraction

We extracted whole brains from the Quare lineage males in 2015 within 10 minutes of lights-on (E. K. Fischer et al., 2016). Fish were anesthetized by immersion in ice water followed by rapid decapitation. Whole brains were removed, flash frozen in liquid nitrogen, and stored at -80°C until further processing (E. K. Fischer et al., 2016). The brains were homogenized, and total RNA was extracted from the brain tissue using Qiagen RNeasy Lipid Tissue Mini Kit (Qiagen, Germany) following manufacturer guidelines (E. K. Fischer et al., 2016).

Small RNA library construction and sequencing

We prepared separate sequencing libraries for each individual using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® - primer set 1 (New England Biolabs, Massachusetts, USA) following manufacturer instructions. We constructed and size selected 61 sRNA libraries (51 brain, 9 ovary and 1 testis sample). For the ovary libraries, we size selected twice, first on the RNA insert, and secondly on the final library molecule (adapters, primers and insert sequence). The size selection was done by running an agarose gel and cutting bands from the gel at the desired fragment length, ~150 nt, (adapters, primers, and insert sum). The gel slices were electrophoretically transferred to blots, washed and extracted for further processing (see supplemental file1 for complete ovary library preparation procedure).

We prepared the brain samples identically, except the size selection of cDNA libraries was performed using BluePippin (Sage Science, MA, USA) with 3% agarose Q3 internal marker

cassettes (Prod. no. BDQ3010). The other major difference was the Blue Pippin size selected libraries were only size selected once, i.e. final libraries molecules were size selected, but not the input RNA. We started first with the size selection window from the specifications provided in the NEB sRNA library prep kit, 105-155 nt. We used TapeStation (D1000 ScreenTape, Part no. 5067-5582) to check the distribution of library molecule lengths. The size selection window of 105-155 nt yielded an average library molecule length of 143 nt (adaptors and primers contribute 126 nt to the molecule length) and average insert length of 17 nt. To size select for both miRNA and piRNA small RNAs, we enlarged the window size (104-190nt) to include the larger piRNA lengths of 26-31nt (Dorner, Eulalio, Huntzinger, & Izaurralde, 2007; Edge, 2009; Seto, Kingston, & Lau, 2007). The larger window size yielded an average library molecule length of 152 nt (26 nt average insert length). Due to COVID-19 pandemic supply limitations, we had to substitute the 3% agarose cassettes for 2% agarose V1 internal marker cassettes (Prod no. BDF2010) for 51 samples. To account for the agarose concentration difference we narrowed and shifted the size selection range from 104-190 (range of 86 nt) to 115-180 nt (range of 65 nt), (see https://doi.org/10.5281/zenodo.8127693 for a list of specifications for each sample's preparation). All libraries were combined with unique barcodes into six pools of 12 libraries per sequencing lane. We balanced experimental groups across sequencing lanes, and blocks of individual libraries prepared in the same week were sequenced together. Libraries were sequenced on an Illumina NextSeq500 at the Colorado State University Genomics Core in November 2020 and January 2021.

Adapter trimming, quality and low complexity filtering

We utilized bbduk.sh (Decontamination Using Kmers) for adapter trimming and read quality filtering, setting a minimum average read quality cutoff of Phred score 30. The complete

command and options used for adapter trimming and quality filtering can be found at (https://doi.org/10.5281/zenodo.8127705). Additionally, we applied low complexity filtering to all reads prior to mapping, employing the low complexity filter tool from the PRINSEQ package with the lc_method set to 'DUST' and lc_threshold set to 32.

Reads mapping and miRNA discovery using mirdeep2

The miRDeep2 package (Friedländer, MacKowiak, Li, Chen, & Rajewsky, 2012) was used for discovery of novel and conserved miRNAs in our guppy sRNA sequencing data . We ran separately by tissue (51 brain, 9 ovary, and 1 testis) the (mapper.pl) script for reads preprocessing and mapping to the most recent guppy reference genome (GCA_904066995.1). Pre-processing collapsed reads into read groups and size-selected for sequences 18-34 nt, then miRDeep2 detected miRNAs in each tissue using Bayesian statistics to score the "fit" of sequenced small RNAs to the biological model of miRNA biogenesis (miRDeep2_core_algorithm.pl; Friedlander et al. 2012). The core algorithm assigns a log odds score to each sequence generated from a number of features, including: (1) the number of reads (mature, loop and star sequences) within the sequencing data that map to the miRNA precursor signature during the mirdeep2-core-algorithm; (2) the short 3' duplex overhangs typically generated from Drosha/Dicer cleavage ; and (3) the thermodynamic stability of the hairpin structures (Friedländer et al., 2008). Because each tissue had a unique mapping file used for running the miRDeep2 core algorithm, we determined the union and intersection for the provisional identifiers assigned uniquely to each miRNA signature within the three tissues.

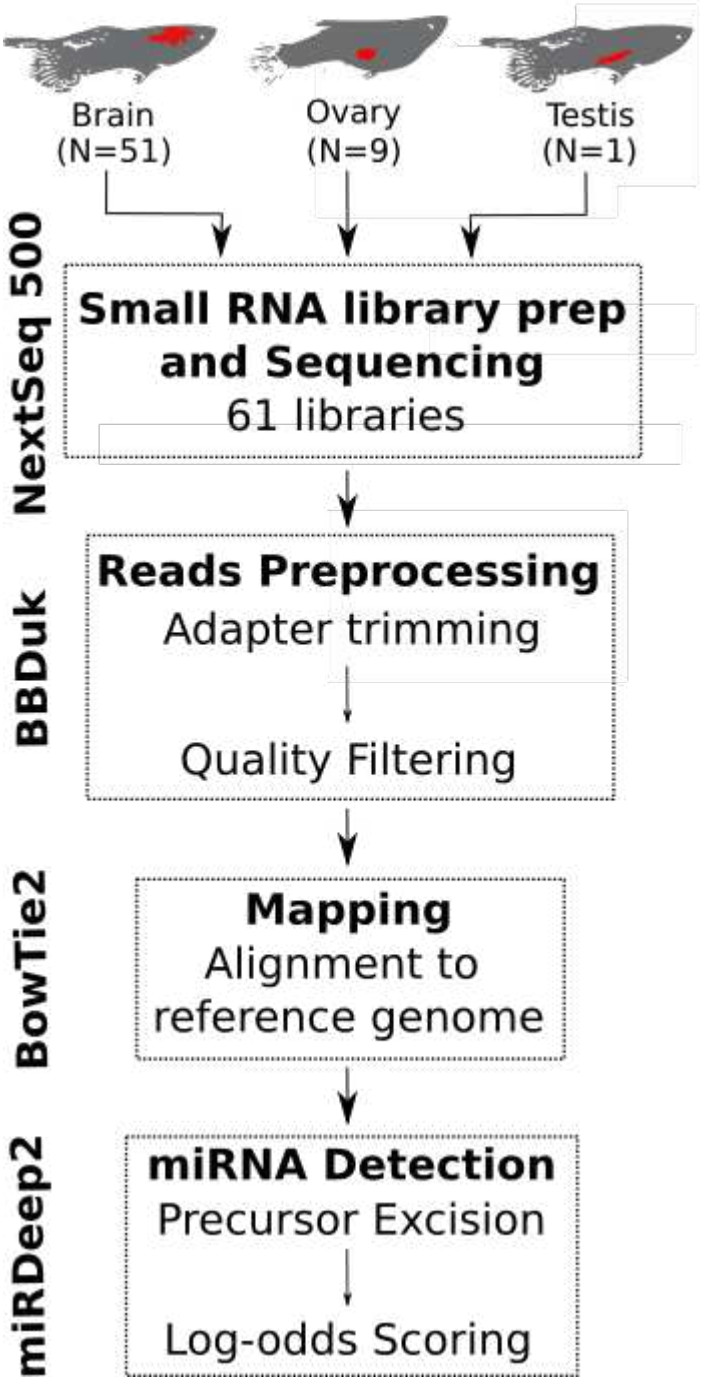

Figure 2.1: Overview of our miRNA prediction process by small RNA Next Generation sequencing (NGS) and miRDeep2 algorithms. We generated small RNA libraries from brain, ovary and testis samples. Libraries were sequenced using Illumina NGS. We used "Decontamination Using Kmers" (BBDuk) from the BBtools package for adapter trimming and quality filtering. Bowtie aligned reads to the male guppy reference genome, and miRDeep2 used the mapped reads files for excision of putative miRNA precursors, and log-odds scoring for assessment of the quality of detected miRNAs.

Putative miRNA precursors assessment

We assessed the putative miRNA precursors (pre-miRNA) for confidence of the prediction (https://doi.org/10.5281/zenodo.8127717 for a complete list of all the putative pre-miRNAs, N= 2883 all tissues). Upon inspection of pre-miRNA, mature, and star sequence alignments, it was evident that some of the lower read count signatures (e.g., 3 RPM) likely represented true miRNAs based on miRNA and miRNA* read alignment and exact matches in miRBase. To maximize miRNA discovery rate (as a first pass of the guppy miRNAome, with a relatively low estimated false negative rate), a more lenient cutoff of 3 mature reads mapped per million (RPM) was used. We excluded mirdeep2 score from post-filtering to avoid removal of putative miRNAs with strong alignment support. We also required independent calls of the mature sequence in at least 10 samples, and the vast majority of precursors from all three tissues were called independently from at least 5 individuals (data not shown). MiRNA count normalization was calculated as follows,

$$RPM\ of\ miRNA = \frac{Number\ of\ reads\ mapped\ \text{to pre-miRNA} * 10^6}{Total\ N\ of\ mapped\ reads\ \text{from}\ library}$$

Sequence based identification of miRNA primary transcript loci by miRBase searches

To determine primary transcript loci of the putative miRNA precursors (pre-miRNAs), we blasted the set of discovered pre-miRNAs against all guppy records in the standard NCBI collection. Ribosomal RNA (rRNA), transfer RNA (tRNA), and miRNA target sequences are not included within these records. We required a minimum cutoff of 95% sequence identity. We filtered 14,337 blast hits by selecting alignments with zero gaps and the top three longest alignments for each query, and then identified the alignment with the lowest E-value for each

putative pre-miRNA. The alignments were then categorized into three genome feature bins: unnanotated region of the guppy genome, messenger RNA (mRNA), or noncoding RNA (ncRNA). We used Batch Entrez to look up the complete records of the mRNA (N = 180) and ncRNA (N = 53) alignments. Additionally, we used Rfam batch search to align our set of pre-miRNAs against the Rfam database collections (the Rfam collections consist of families of structural RNAs including ncRNA genes as well as cis-regulatory elements).

Annotations of putative miRNAs

We applied minimum alignment length filtering to the blastn alignments to find pre-miRNA loci in intragenic and intergenic regions of the guppy genome. We filtered alignments based on minimum mean lengths of our pre-miRNA sequences, using the logic that pre-miRNAs contain 5' and 3' arms (18-24 nt) and a loop sequence (~10-15nt) and thus are usually in the 60-80 nt range (Ambros et al., 2003). We required a minimum alignment length of 60 nucleotides, and allowed no more than one base mismatch per alignment. We used Batch Entrez to lookup the full records from the blastn filtered alignment output see (https://doi.org/10.5281/zenodo.8131978) for the Batch Entrez uploaded file; we then separated the coding DNA sequences (CDS) features from the total mRNA (XM_ accession) records.

PiRNA read characterization

We used the small RNA pipeline (Tate A.J et al. 2023) tiny-count: a counting tool for hierarchical classification and quantification of small RNA-seq reads with single-nucleotide precision (github.com/MontgomeryLab/tinyRNA) to hierarchically assign and count mapped reads to five RNA categories: (1) rRNA/tRNA, (2) protein coding gene, (3) miRNA hairpin/miRNA/miRNA*, (4) interspersed repeats, and (5) unannotated region. Mapped reads were assigned based on their overlap with annotated genome features. The hierarchy of

assignments aimed to remove known read sequences, leaving predominantly piRNA sequences. The interspersed repeat category comprised DNA transposons and retrotransposon features, which are known loci for piRNA primary transcripts. Overlapping reads from germline samples with these features were expected to be enriched in piRNA sequences. The 'unassigned region' category, although lacking a specific biological basis for piRNA detection, was assigned last, with the expectation that the remaining sequences would be enriched in piRNAs. To create the 'unassigned region', a GFF file with 100kb features spanning the entire guppy genome was generated, with 20kb steps. The GFF file was provided to tinyRNA as a feature GFF.

miRNA clustering

To determine the percentage of our miRNA set found within clusters we used a sliding window algorithm (10kb with 2kb steps). If two or more pre-miRNA genomic loci are found within the same window, they are marked as part of a cluster.

Results

Characterization of orthologous proteins in miRNA biogenesis and post-transcriptional regulation

To confirm that guppies have the core components of the Drosha/Dicer-dependent miRNA biogenesis pathway, and the argonaute family of proteins (Fig. 2.2), we searched for DROSHA (ribonuclease III), DICER (endoribonuclease Dicer), AGO1-4, and PIWI proteins PIWI and PIWIL in 4 species, *Poecilia reticulata, Danio rerio, Homo sapiens,* and *Drosophila melanogaster*. We constructed dendrograms from AGO and PIWI proteins based on amino acid sequence similarity, and confirmed that DROSHA and DICER are both present in the annotated guppy genome (data not shown).

Guppies have two PIWI proteins (GILI and GILI2) (Fig. 2.2). *P. reticulata* had predicted protein sequences for four AGO proteins (gAGO1-4). For all four AGO proteins the teleost species clustered in a separate group from the *H. sapiens* (Fig 2.2). *D. melanogaster* sequences, PIWI and Aubergine were included as an invertebrate outgroup.

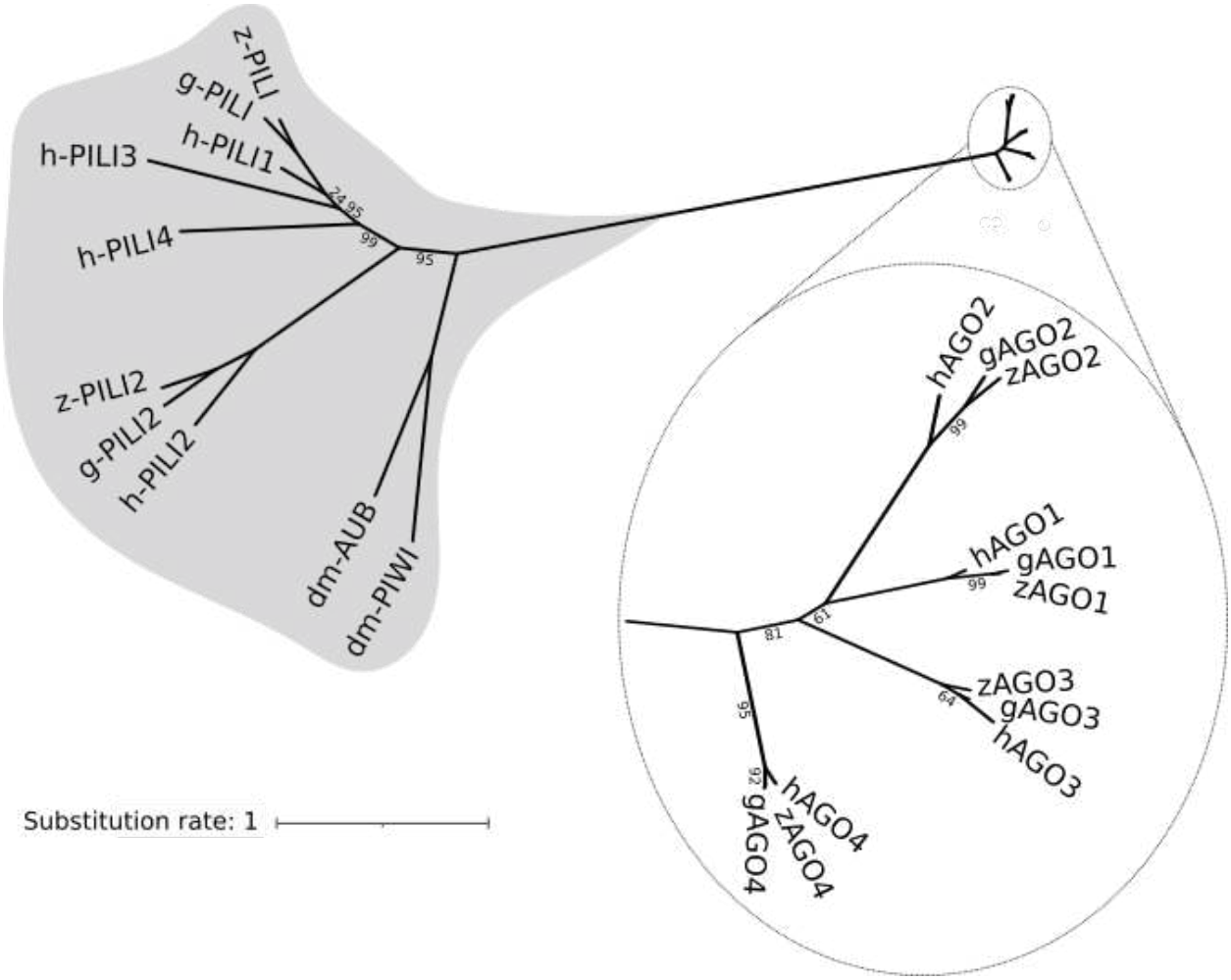

Figure 2.2: Argonaute family evolutionary tree derived from protein sequence alignments. Two distinct subclasses are present, the germline specific PIWI branch (grey) and somatic AGO branch (zoom out). Bootstrap values are indicated in the middle of branches for all values less than 100. Guppy PIWI proteins are g-PILI (accession XP_008415675.1) and g-PILI2 (accession XP_008415818.1). dm - *Drosophila melanogaster*. g - guppy, z - zebrafish, h - human.

sRNA read distributions in brain, ovary, and testis

We mapped to the guppy genome (GCA_904066995.1), 276,722,633 brain, 18,764,380 ovary, and 6,751,682 testis total reads as a starting point for mirdeep2-based miRNA discovery. (Fig. 2.3). Brain sRNA reads were largely distributed in the lower miRNA range (18-24 nt); ovary and testis distributions had more reads in the higher piRNA range (25-34 nt). The 5' nucleotide distributions of the mature sequences were similar across tissues, with strong bias toward uridine as the 5' nucleotide (Fig. 2.3), as expected for small RNAs.

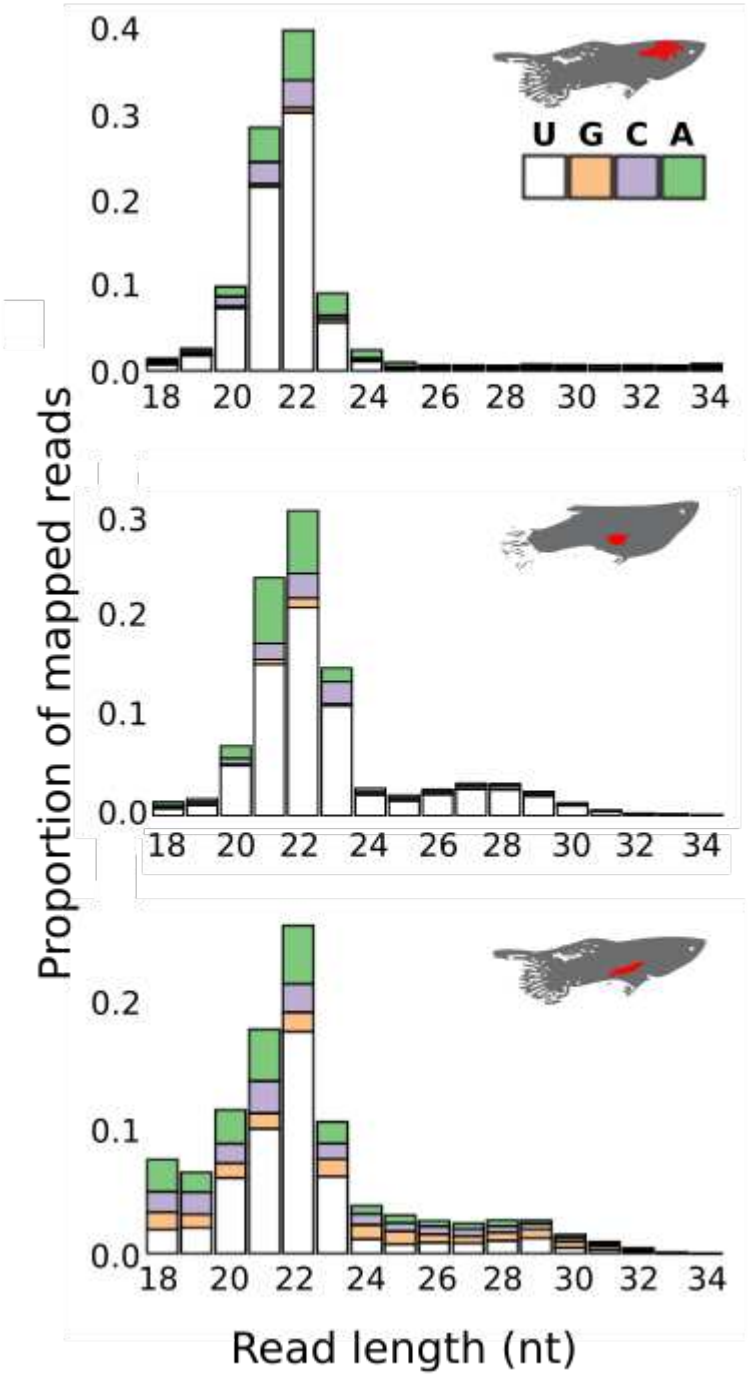

Figure 2.3: All samples exhibit strong 5' uridine bias, and read length distributions support presence of miRNAs and piRNAs. Read length distribution and 5' nucleotide bias of mapped reads for brain (N= 51), ovary (N= 9), and testis (N= 1). X-axis is read length in nucleotides. Y-axis is the proportion of total mapped reads, stacked color coded bar corresponds to read distribution of the 5' nucleotide.

Putative miRNA precursors assessment and filtering

We used the mirdeep2 core algorithm to constuct (1,487 brain, 558 ovary, and 350 testis), miRNA signatures, consisting of mature precursor miRNA and star sequences. After post-filtering for minimum library presence (10 libraries from brain, 3 for ovary, 1 for testis), and minimum reads (3 RPM), we retained 1120 miRNA signatures for further analyses see (https://doi.org/10.5281/zenodo.8132659) for read counts, mature sequences, and tissue expression. Most of the post-filtered miRNAs were detected in a single tissue. The tissues had a variable number of unique miRNA sequences that did not appear to correlate with the number of libraries analyzed (Fig. 2.4).

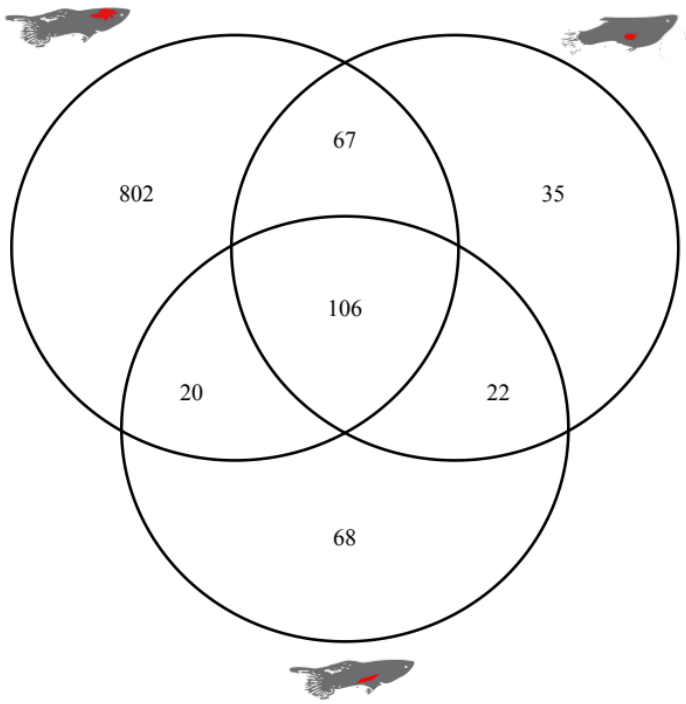

Figure 2.4: Venn diagram depicting the union and intersection of the miRNAs surviving post-filtering. Post-filtering required a minimum total mature read count of 3 reads per million

mapped (RPM), and present in at least 10 libraries. N = 51 brain samples, N = 9 ovary samples, and N = 1 testis sample.

Coordinate based identification of miRNA primary transcript loci

Our blastn results from NCBI nucleotide teleostei (taxid:32443) records showed most of the putative pre-miRNAs alignments are located within intergenic (unannotated) regions of the guppy genome (between genes). The gene annotations were determined in an automated fashion for both guppy and zebrafish, intergenic regions are assigned as a position that falls within the boundaries of the genes start and stop codon. Intragenic annotations consisted of brain miRNAs detected in 24 genes total, ovary 14 genes, and testis 18. Within the gene annotations, most miRNAs were detected in introns. Very few miRNAs were found in exons (brain = 1, and testis = 1), and the ovary samples had no miRNAs overlapping exons.

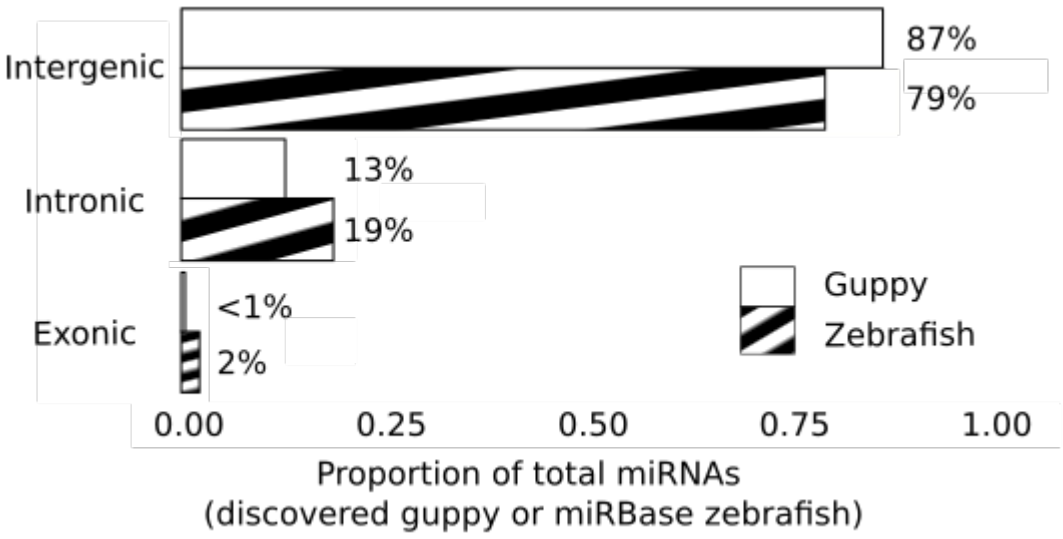

Figure 2.5: MiRNA genomic loci are consistent between Zebrafish and guppy. Pre-miRNA genomic loci were annotated, bar chart illustrates the proportions of pre-miRNA transcripts detected in intergenic (uncharacterized), or intragenic, and intronic or exonic, as compared to

zebrafish. Zebrafish pre-miRNAs were downloaded from miRBase. N= 1120 guppy pre-miRNAs, and N= 355 total zebrafish miRNAs.

miRNA quantification

We estimated miRNA abundances using the mirdeep2 quantifier module and used the normalized counts matrix output to cluster and visualize tissue specific expression patterns of 1120 putative miRNAs. Brain, ovary, and testis tissues expressed 61 miRNAs in common. Clustering by individual libraries produced three distinct groups, corresponding to the three tissues (Fig. 2.6). The ovary libraries consisted of both fertilized ova and embryos at different developmental stages. The unfertilized libraries were more similar to each other, and the distribution of reads varied across developmental groups.

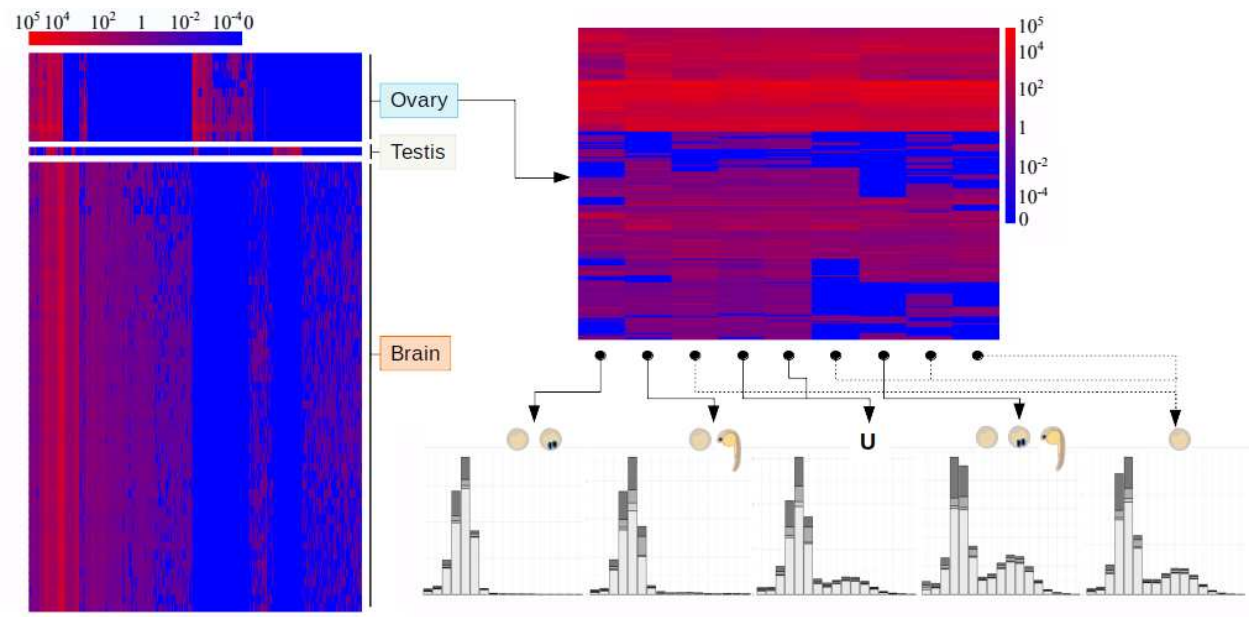

Figure 2.6: Tissues show distinct miRNA expression patterns. Heatmaps - blue indicates low to no expression and red indicates high expression (based on mature read abundance). Left heatmap – columns are individual miRNAs and rows are samples (brain = 51, ovary = 9, and testis = 1). Ovary zoom in (top right heatmap) shows samples as columns and rows as individual miRNAs.

The bottom read length distributions are connected to their respective columns of the heatmap. The symbols above the read length distribution plots indicate the developmental stages of the embryos. From left to right fertilized/eyed, fertilized/neurulation, unfertilized, fertilized/eyed/neurulation, fertilized.

MiRNA sequence orthology and clustering

We assessed miRNA sequence similarity to a custom blast database composed of all miRBase mature sequences. Of the 1120 miRNAs, 228 (83.5%) aligned with the miRBase custom database with an exact match or one mismatch. The 228 matched miRNAs were derived from 184 pre-miRNAs. The 228 matches were assigned to a total of 75 miRNA families based on the seed sequence, i.e. bases 2-7 from the 5' end of the miRNA, and none of the single nucleotide mismatches were within the seed sequence (Table 2.1). The remaining high confidence miRNAs that were not matched were considered to be novel guppy miRNAs (N= 45; Table 2.1).

We performed a clustering analysis (see methods) to determine which of our high confidence miRNAs are expressed from tandem miRNA genes. A large proportion (43.8%) of the 274 miRNAs were found in clusters. Most of the clusters (all except two), had two miRNAs within the cluster, and two clusters had three miRNAs.

Table 2.1: Table of miRNA families for discovered miRNAs, and summary table (bottom) including numbers of miRNA sequences with a given seed sequence, either with or without orthologous matches. Nucleotide mismatches were in non-seed sequence. High confidence sequences without orthologous matches are considered novel guppy sequences (grey background).

MiRNAs with orthologous matches

Family	Seed	Number of Sequences	Mismatches	Family	Seed	Number of Sequences	Mismatches
let-7	gaggua	23	0	miR-126	cguacc	2	0
miR-125	cccuga	10	0	miR-187	cguguc	2	0
miR-9	cuuugg	9	0	miR-18	cugccc	2	0
miR-30	guaaac	9	0	miR-146	gagaac	2	0
miR-181	acauuc	8	0	miR-143	gagaug	2	0
miR-199	ccagug	8	0	miR-103	gcagca	2	0
miR-92	auugca	7	0	miR-138	gcuggu	2	0
miR-26	ucaagu	7	0	miR-1	ggaaug	2	0
miR-27	ucacag	7	0	miR-724	uaaagg	2	0
miR-20	aaagug	6	0	miR-204	ucccuu	2	0
miR-21	agcuua	5	0	miR-734	aaaugc	1	0
miR-135	auggcu	5	0	miR-462	aacgga	1	0
miR-101	acagua	4	0	miR-365	aaugcc	1	0
miR-100	acccgu	4	0	miR-137	acgggu	1	0
miR-10	acccug	4	0	miR-449	agcuca	1	1
miR-22	agcugc	4	0	miR-10544	aggcgu	1	0
miR-128	cacagu	4	0	miR-456	aggcug	1	0
miR-152	cagugc	4	0	miR-2188	aggucc	1	0
miR-222	gcuaca	4	0	miR-731	augaca	1	0
miR-24	ggcuca	4	0	miR-124	caaggu	1	0
miR-203	ugaaau	4	0	miR-730	ccucau	1	0
miR-129	uuuuug	4	0	miR-139	cuacag	1	0
miR-16	agcagc	3	0	miR-29	cugauu	1	1
miR-23	ucacau	3	0	miR-8160	gaauaa	1	0
miR-153	ugcaua	3	0	miR-219	gaauug	1	1
miR-133	uggucc	3	0	miR-489	gacauc	1	0
miR-200	aacacu	2	0	miR-184	ggacgg	1	0
miR-132	aacagu	2	0	miR-34	ggcagu	1	0
miR-200	aauacu	2	0	miR-193	gggucu	1	0
miR-216	aaucuc	2	0	miR-22	guucuu	1	0
miR-338	acaaca	2	0	miR-7552	uacaau	1	0
miR-217	acugca	2	0	miR-458	uagcuc	1	0
miR-183	auggca	2	0	miR-725	ucaguc	1	0
miR-455	augugc	2	0	miR-145	uccagu	1	0
miR-214	cagcag	2	0	miR-210	ugugcg	1	0
miR-727	cagucu	2	0	miR-218	ugugcu	1	0
miR-205	ccuuca	2	0	miR-182	uuggca	1	0
miR-212	ccuugg	2	0	miR-375	uuguuc	1	0

Novel guppy miRNAs (no orthologous match)

Seed	Number of Sequences
gcggcu	9
aagugc	4
cagaua	3
agcucc	2
cccucc	2
cugucc	2
cagcug	2
cguccg	1
cuauca	1
caucgc	1
cgaucc	1
cugcuu	1
ucgcug	1
ugaucc	1
ugcucc	1
aagccc	1
ccgccc	1
ccugcu	1
cgaccc	1
uugucg	1
uuuuuu	1
cagucc	1
cccuau	1
ccuacc	1
cggaac	1
cucucu	1
ucaagu	1
uucccu	1

Total mature miRNAs	45
Unique mature miRNAs	29

Summary of orthologous and novel miRNA detection

	Orthologous matched	No orthologous match
Total mature miRNAs	228 (83.5%)	45 (16.4%)
Unique mature miRNAs	125	19
Unique pre-miRNAs	184	39
Unique star sequences	161	37

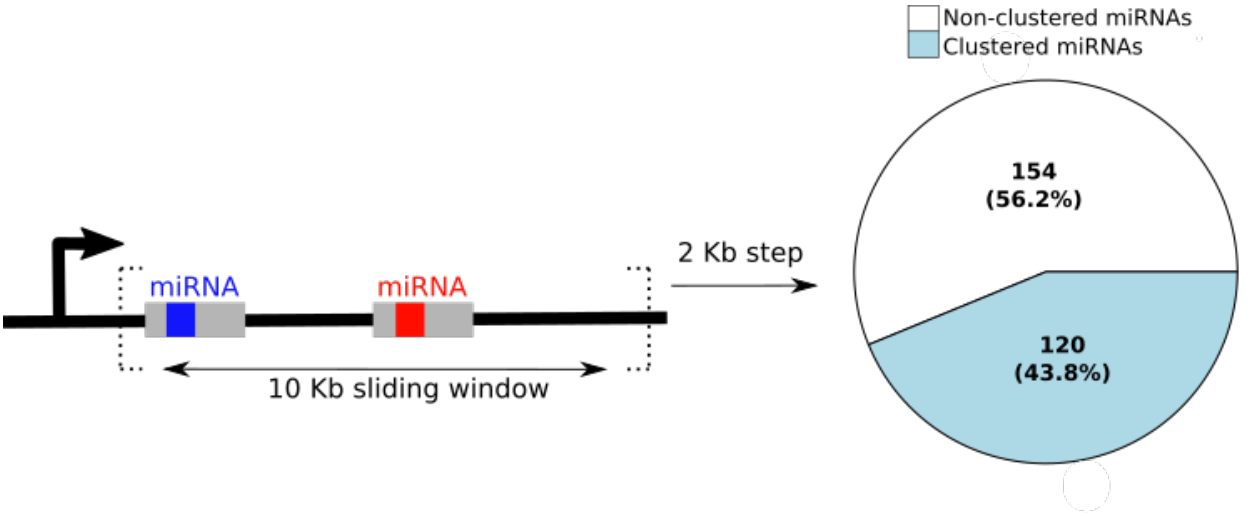

Figure 2.7: Nearly half of our discovered miRNAs were found in clusters (within ~10kb). Diagram of the miRNA cluster detection sliding window (10Kb with 2kb steps) algorithm, and pie chart of proportion of clustered vs. non-clustered miRNAs.

Highest abundance miRNAs

We identified the most abundant miRNAs in our dataset from the proportion of normalized counts generated by the mirdeep2 quantification module. Only the ovary and testis tissues had miRNAs in common in their top five highest abundance miRNAs, i.e. the second most abundant ovary and testis miRNA was the same miRNA. Brain has a miRNA detected without a miRBase orthologous match (no annotation), indicating the putative miRNA is novel (Fig. 2.8).

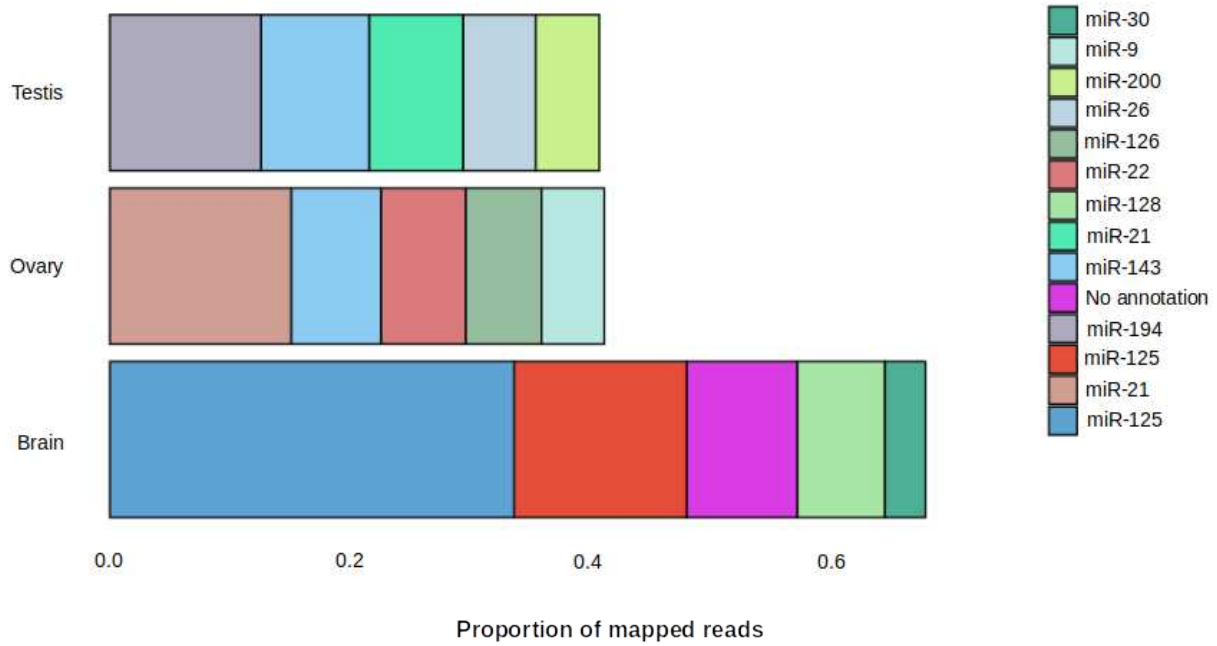

Figure 2.8: The top five most abundant miRNAs from our aligned dataset in brain, ovary and testis. Colors indicate families, established by matches to miRBase teleost miRNAs.

Small RNA classification of mapped reads

To analyze the distribution of miRNA and piRNA reads in our ovary samples, we classified the reads into different classes using tinyRNA, based on their overlap with annotated genomic features, and compared the results with the small RNA classification data from tinyRNA to identify piRNA reads.

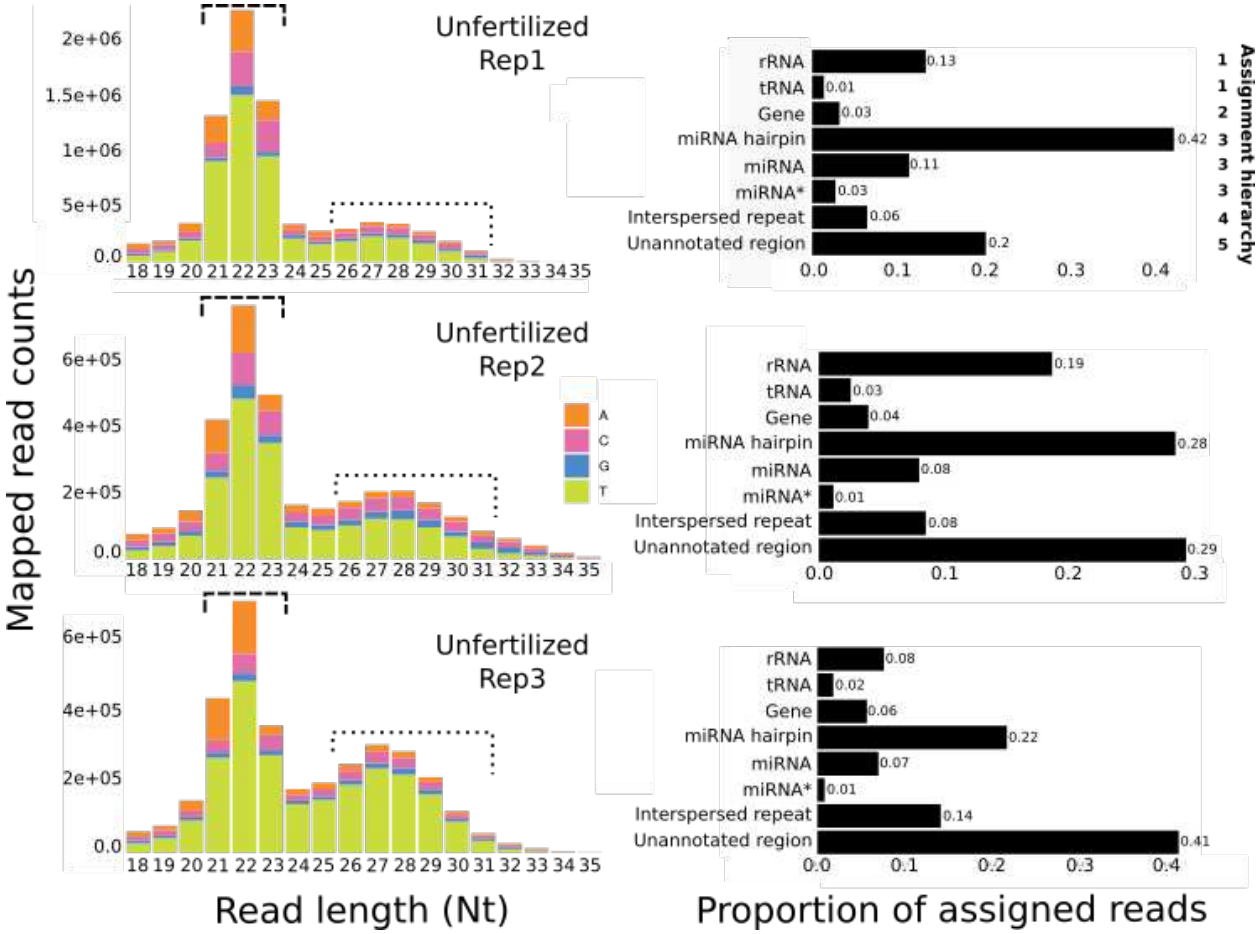

Figure 2.9: Assignment of miRNA and piRNA reads to genomic RNA features. Long dash brackets refer to miRNA read distribution (20-23nt), short dash brakcets refer to piRNA read distributions (24-32nt). Mapped read absolute counts (left bar plots) are assigned to genomic features (e.g. gene, rRNA, tRNA, etc., right, black bar plots) based on overlap of the read with the genomic feature (right bar plot). Bar chart data labels indicate proportion of total reads assigned for the RNA category.

Discussion

This study focused on the characterization of miRNAs and piRNAs in the Trinidadian guppy. Essential proteins in the Dicer-dependent miRNA pathway were confirmed, and our data supported the bias for 5'-U/A in miRNAs. We applied our own filtering criteria to reduce false discovery of miRNAs. After quantifying miRNAs, we found tissue-specific expression patterns.

The study also explored piRNA expression, with an inverse abundance trend compared to miRNAs. Overall, this comprehensive characterization of guppy miRNAs and piRNAs lays the groundwork for future investigations into their roles in gene regulation, evolutionary processes, and phenotypic plasticity.

We confirmed the essential proteins in the Dicer-dependent miRNA pathway are present in Trinidadian guppies, and determined the human homologous complement of AGO and PIWI proteins. We assessed filtering criterion of our putative miRNAs, we suggest post-filtering for the detection of miRNAs (using miRDeep2) from sRNA data, for confidence scoring, and downstream analyses. We chose to filter with stringent requirements for our miRNA characterization, but also included analyses with more relaxed filtering (Fig 2.6). The minimum cutoff post-filtering was put in place to reduce the initial false negatives that were moved forward for annotation and quantification. Initially we applied a basic cutoff of the 'mirdeep' log-odds ratio; however we found (by inspection of the precursor signature) many qualitatively good confidence miRNA signatures that were removed by the mirdeep score cutoff. For this reason we instead used minimum mature read depth, and presence throughout multiple libraries for removing low quality calls. Additionally, we did not account for hairpin secondary structure, although miRDeep2 does provide significance testing statistics for stability of the proposed hairpin structure. We found many non-significant structures to have otherwise strong attributes of miRNA precursor hairpins both passing our stringent post-filtering requirements as well as individual inspection (reads mapped to both strands of precursor sequence with 3' overhang) of the alignments. Lower probability of false positive miRNA predictions can come from increasing the minimum reads requirement, assessing the mirdeep2 log odds ratio, and miRBase or MirGeneDB matches that infer phylogenetic conservation; however, as the requirement for these

filters increases, the probability of false negatives will also go up. Finding a particular filter set and level largely depends on the type analyses that will be performed on the discovery set. Together the read length distributions and 5'-uridine bias indicate a sRNA enriched dataset. All three tissues had reads distributed around the miRNA expected average length, indicating a large fraction of the data is made up of mature miRNA 5p and 3p strands. The read length and abundance distributions showed a higher proportion of reads in the 25-34 nt range in gonads compared to brain tissue (Fig. 2.3). Animal gonads are the only known tissue type abundant in piRNAs (Hirakata & Siomi, 2016), consistent with the 25-34 nt reads in our ovary and testis samples indicating the presence of piRNAs.

Our data supports a 5'-U/A bias in guppies. One of the few sequence motifs identified in miRNAs is the frequent 5'-uridine (5'-U) (Seitz, Tushir, & Zamore, 2011), or 5'-adenosine (5'-A) (Meijer et al., 2014). In flies, the 5'-U directs sRNAs to AGO1, and 5'-C ends favor AGO2 loading (Czech et al., 2009; Ghildiyal et al., 2008; Ghildiyal, Xu, Seitz, Weng, & Zamore, 2010; Okamura, Liu, & Lai, 2009). The bias is related to how the duplex is oriented during Argonaute loading; the mature sequence is usually the strand with the less stable 5' end of the duplex (Dianne S. Schwarz & Tingting Du, Zuoshang Xu, Neil Aronin, 1996; Khvorova, Reynolds, & Jayasena, 2003; Seitz et al., 2011). The 5' nt is important for sorting small RNAs between effector complexes, and the 5' nt dictates which strand is bound by the Ago.

The post filtering removed many low abundance putative miRNAs, such that the original higher diversity of miRNAs detected in the brain (likely due to the larger number of individual brain libraries and deeper sequencing, compared to ovary, and testis) was diminished. Additionally, the brain is known to express more distinct and a larger number of miRNAs than any other tissue in vertebrates (Adlakha & Saini, 2014), which may also explain our larger pre-

filtered brain set. Brain has a wide variety of cell types, both neuronal and nonneuronal (Adlakha & Saini, 2014; Motti, Bixby, & Lemmon, 2012), and there is a direct correlation between the number of miRNAs and morphological complexity of organisms (Berezikov, 2011).

In addition to expression and biogenesis criterion established by the mirdeep2 core algorithm, phylogenetic conservation adds evidence for the existence of miRNAs, and is another common criteria for miRNA discovery. We searched the guppy miRNAs against all miRBase mature miRNA sequences, and found 85.3% of our query miRNAs had at least one species with an identical or up to one mismatched base-pairing miRNA, and most had numerous species matches (data not shown). Of the miRBase teleostei miRNA sequences (3,688 total), many (42.7%) are single entries, and the remaining (57.3%) have one or more duplicates, i.e. multiple entries with the same miRNA sequence within or across species. Although the duplicated sequences make up more than half of all miRBase teleostei entries, this duplicated set is made up of relatively few unique sequences. Therefore, most of the teleostei miRNA sequence diversity is found as unique single entries, i.e. one entry from a single species. If the miRBase teleostei miRNA sequence diversity is at all telling of individual species diversity within the teleostei clade, then most miRNAs, ~43% are unique to the species and roughly 57% have one or more conserved matches in miRBase. This suggests any homology criterion for designating a sequence as a miRNA should (at least) take into account a possibility that a novel miRNA has little or no phylogenetic conservation, and is not likely represented in any miRNA databases. Phylogenetic conservation improves confidence in the prediction, but the absence of conservation shouldn't be used as disqualifying, i.e. if a conserved miRNA doesn't exist in miRBase or MirGeneDB.

In guppies embryonic development occurs in the ovary (Wourms, 1981), therefore our data provides an aggregated snapshot of putative miRNA and piRNA expression for multiple

embryonic stages (Fig. 2.7). We were able to loosely designate developmental stages by observing characteristic features that were present in the individual when the ovariectomy was performed (see methods – tissue collection). Although the mixed developmental staging muddles the interpretation of miRNA profile association with a particular embryonic stage, we can see dynamic miRNA and piRNA expression profiles across the developmental timeline (Fig. 2.7). The two unfertilized samples have the highest developmental homogeneity we assessed, as both females were virgins, and were carrying only unfertilized eggs. The miRNA expression profiles of the two unfertilized samples were also most similar to each other (Fig. 2.7), possibly showing a correlation of the profile with developmental stage. The other samples had fertilized eggs, and embryos in various stages of development. It was more difficult to see a relationship for the samples that had developing embryos, likely due to the overlapping developmental stages of two or more broods, i.e. with the occurrence of superfetation. Future studies on isolated embryos could disentangle the miRNA and piRNA expression profiles of individual embryos.

In addition to miRDeep2 for the discovery of miRNAs, we also used tinyRNA for analysis of our sRNA-seq data. Guppy annotations are fairly limited, as compared to more popular model systems, however we put together a set of annotation files for the assignment of reads in our sRNA-seq data. There are a large number of unique piRNA sequences (Girard, Sachidanandam, Hannon, & Carmell, 2006), and less detail is understood about the biogenesis of piRNAs compared to miRNAs.

We used the small RNA pipeline tinyRNA to assess the presence of piRNA reads in our data set. Read length and 5' nucleotide distributions differed among the nine ovary samples. The unfertilized replicate1 had approximately 42% of assigned reads to miRNA hairpins, and 20% to unannotated genomic regions. The unfertilized replicate3 had the largest putative piRNA

signature, and accordingly also had the largest proportion (0.41) of assigned reads to the unannotated region, and the smallest proportion (0.22), of miRNA hairpin assignments (Fig. 2.9). These trends taken together indicate that tinyRNA is able to detect piRNA and miRNA reads in our samples, and could be used to quantitatively differentiate sRNA distributions.

This study represents the first characterization of the guppy miRNA and piRNA transcriptome. We include confidence assessments for a large set of computational miRNA predictions, orthologous DICER-dependent analysis of phylogentic conservation, and characterize miRNA and piRNA expression during guppy embryonic development. We characterized a high confidence set of miRNAs consisting of 1120 miRNAs. We also showed miRNA expression patterns across three tissues. This characterization of an extensive repertoire of guppy miRNAs provides a starting point for future studies aimed towards understanding the physiological roles of miRNAs, the evolution of gene regulation and complex phenotypes, and the influence of phenotypic plasticity on evolutionary trajectories.

CHAPTER 2 WORKS CITED

Adlakha, Y. K., & Saini, N. (2014). Brain microRNAs and insights into biological functions and therapeutic potential of brain enriched miRNA-128. *Molecular Cancer*, *13*(1), 1–18. https://doi.org/10.1186/1476-4598-13-33

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*(7006), 350–355. https://doi.org/10.1038/nature02871

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., … Tuschl, T. (2003). *A uniform system for microRNA annotation*. 277–279. https://doi.org/10.1261/rna.2183803.One

Anne E. Magurran and Benoni H. Seghers. (1991). *Variation in Schooling and Aggression Amongst Guppy (Poecilia reticulata) Populations in Trinidad*. *118*(3), 214–234.

Barson, N. J., Cable, J., & Van Oosterhout, C. (2009). Population genetic analysis of microsatellite variation of guppies (Poecilia reticulata) in Trinidad and Tobago: Evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *Journal of Evolutionary Biology*, *22*(3), 485–497. https://doi.org/10.1111/j.1420-9101.2008.01675.x

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function Review. *Biosensors and Bioelectronics*, *116*, 281–297. https://doi.org/10.1016/j.bios.2017.04.001

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, *136*(2), 215–233. https://doi.org/10.1016/j.cell.2009.01.002

Baskerville, S., & Bartel, D. P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna*, *11*(3), 241–247. https://doi.org/10.1261/rna.7240905

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics*, *12*(12), 846–860. https://doi.org/10.1038/nrg3079

Bohnsack, M. T., Czaplinski, K., & Görlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna*, *10*(2), 185–191. https://doi.org/10.1261/rna.5167604

Cai, Y., Yu, X., Hu, S., & Yu, J. (2009). A Brief Review on the Mechanisms of miRNA Regulation. *Genomics, Proteomics and Bioinformatics*, *7*(4), 147–154. https://doi.org/10.1016/S1672-0229(08)60044-3

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–655. https://doi.org/10.1016/j.cell.2009.01.035

Chen, K., & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, *8*(2), 93–103. https://doi.org/10.1038/nrg1990

Cho, W. C. S. (2007). OncomiRs: The discovery and progress of microRNAs in cancers. *Molecular Cancer*, *6*, 1–7. https://doi.org/10.1186/1476-4598-6-60

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., … Hannon, G. J. (2009). Hierarchical Rules for Argonaute Loading in Drosophila. *Molecular Cell*, *36*(3), 445–456. https://doi.org/10.1016/j.molcel.2009.09.028

Dianne S. Schwarz, G. rgy H., & Tingting Du, Zuoshang Xu, Neil Aronin, and P. D. Z. (1996). Asymmetry in the Assembly of the RNAi Enzyme Complex. *Ukrain'skyi Biokhimichnyi Zhurnal*, *68*(5), 24–25.

Dorner, S., Eulalio, A., Huntzinger, E., & Izaurralde, E. (2007). Delving into the diversity of silencing pathways. Symposium on MicroRNAs and siRNAs: Biological Functions and Mechanisms. *EMBO Reports*, *8*(8), 723–729. https://doi.org/10.1038/sj.embor.7401015

Drakaki, A., & Iliopoulos, D. (2009). MicroRNA Gene Networks in Oncogenesis. *Current Genomics*, *10*(1), 35–41. https://doi.org/10.2174/138920209787581299

Dzikowski, R., Hulata, G., Harpaz, S., & Karplus, I. (2004). Inducible reproductive plasticity of the guppy Poecilia reticulata in response to predation cues. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, *301*(9), 776–782. https://doi.org/10.1002/jez.a.61

Edge, L. (2009). Molecular Biology Select. *Cell*, *137*(2), 187–189. https://doi.org/10.1016/j.cell.2009.04.015

Elvidge, C. K., Ramnarine, I., & Brown, G. E. (2014). Compensatory foraging in Trinidadian guppies: Effects of acute and chronic predation threats. *Current Zoology*, *60*(3), 323–332. https://doi.org/10.1093/czoolo/60.3.323

Endler, J. A. (1995). Multiple-trait coevolution and environmental gradients in guppies. *Philos. Tmns. R. Sot. London SW B*, *13*(29), 107–122.

Fischer, E. K., Ghalambor, C. K., & Hoke, K. L. (2016). Plasticity and evolution in correlated suites of traits. *Journal of Evolutionary Biology*, *29*(5), 991–1002. https://doi.org/10.1111/jeb.12839

Fischer, Eva K., Harris, R. M., Hofmann, H. A., & Hoke, K. L. (2014). Predator exposure alters stress physiology in guppies across timescales. *Hormones and Behavior*, *65*(2), 165–172. https://doi.org/10.1016/j.yhbeh.2013.12.010

Fischer, Eva K., Soares, D., Archer, K. R., Ghalambor, C. K., & Hoke, K. L. (2013). Genetically and environmentally mediated divergence in lateral line morphology in the Trinidadian guppy (Poecilia reticulata). *Journal of Experimental Biology*, *216*(16), 3132–3142. https://doi.org/10.1242/jeb.081349

Fischer, Eva K., Song, Y., Hughes, K. A., Zhou, W., & Hoke, K. L. (2021). Nonparallel transcriptional divergence during parallel adaptation. *Molecular Ecology*, *30*(6), 1516–1530. https://doi.org/10.1111/mec.15823

Fraser, B. A., Whiting, J. R., Paris, J. R., Weadick, C. J., Parsons, P. J., Charlesworth, D., Weigel, D. (2020). Improved Reference Genome Uncovers Novel Sex-Linked Regions in the Guppy (Poecilia reticulata). *Genome Biology and Evolution*, *12*(10), 1789–1805. https://doi.org/10.1093/gbe/evaa187

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., & Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology*, *26*(4), 407–415. https://doi.org/10.1038/nbt1394

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). MiRD eep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, *40*(1), 37–52. https://doi.org/ 10.1093/nar/gkr688

Gasparini, C., Daymond, E., & Evans, J. P. (2018). Extreme fertilization bias towards freshly inseminated sperm in a species exhibiting prolonged female sperm storage. *The Royal Society Publishing*. https://doi.org/10.1098/rsos.172195

Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, *525*(7569), 372–375. https://doi.org/10.1038/nature15256

Ghalambor, C. K., Reznick, D. N., & Walker, J. A. (2004). Constraints on adaptive evolution: The functional trade-off between reproduction and fast-start swimming

performance in the Trinidadian guppy (Poecilia reticulata). *American Naturalist*, *164*(1), 38–50. https://doi.org/10.1086/421412

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., … others. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science*, *320*(5879), 1077–1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., & Zamore, P. D. (2010). Sorting of drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *Rna*, *16*(1), 43–56. https://doi.org/10.1261/rna.1972910

Gilliam, J. F., Fraser, D. F., & Alkins-Koo, M. (1993). Structure of a tropical stream fish community: a role for biotic interactions. *Ecology*, *74*(6), 1856–1870. https://doi.org/10.2307/1939943

Gilliam, James F, Fraser, D. F., & Alkins-Koo, M. (1993). Structure of a tropical stream fish community: a role for biotic interactions. *Ecology*, *74*(6), 1856–1870.

Girard, A., 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.

Gosline, A. K., & Rodd, F. H. (2008). Predator-induced plasticity in guppy (Poecilia reticulata) life history traits. *Aquatic Ecology*, *42*(4), 693–699. https://doi.org/10.1007/s10452-007-9138-7

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., & Eddy, S. R. (2003). Rfam: An RNA family database. *Nucleic Acids Research*, *31*(1), 439–441. https://doi.org/10.1093/nar/gkg006

HaigH. Kazazian Jr, Corinne Wong, & Hagop Youssoufian*, A. F. S. D. G. P. & S. E. A. (1988). Haemophilia A resulting from de novo insertion of Ll sequences represents a novel mechanism for mutation in man. *Letters to Nature*, *1*(7), 226–227.

Handelsman, C. A., Ruell, E. W., Torres-Dowdall, J., & Ghalambor, C. K. (2014). Phenotypic plasticity changes correlations of traits following experimental introductions of Trinidadian guppies (Poecilia reticulata). *Integrative and Comparative Biology*, *54*(5), 794–804. https://doi.org/10.1093/icb/icu112

Harris, S., Ramnarine, I. W., Smith, H. G., & Pettersson, L. B. (2010). Picking personalities apart: estimating the influence of predation, sex and body size on boldness in the guppy Poecilia reticulata. *Oikos*, *119*(11), 1711–1718.

Harrison, P. W., Wright, A. E., & Mank, J. E. (2012). The evolution of gene expression and the transcriptome-phenotype relationship. *Seminars in Cell and Developmental Biology*, *23*(2), 222–229. https://doi.org/10.1016/j.semcdb.2011.12.004

Hirakata, S., & Siomi, M. C. (2016). piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1859*(1), 82–92. https://doi.org/10.1016/j.bbagrm.2015.09.002

Hoang, Diep Thi, Olga Chernomor, Arndt Von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35 (2): 518–22.

Hui, J. H. L., Marco, A., Hunt, S., Melling, J., Griffiths-Jones, S., & Ronshaugen, M. (2013). Structure, evolution and function of the bi-directionally transcribed iab-4/iab-8 microRNA locus in arthropods. *Nucleic Acids Research*, *41*(5), 3352–3361. https://doi.org/10.1093/nar/gks1445

Huizinga, M., Ghalambor, C. K., & Reznick, D. N. (2009). The genetic and environmental basis of adaptive differences in shoaling behaviour among populations of Trinidadian guppies, Poecilia reticulata. *Journal of Evolutionary Biology*, *22*(9), 1860–1866. https://doi.org/10.1111/j.1420-9101.2009.01799.x

Ishizu, H., Siomi, H., & Siomi, M. C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes \& Development*, *26*(21),

Juliano, C., Wang, J., & Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annual Review of Genetics*, *45*, 447–469.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K F Wong, Arndt Von Haeseler, and Lars S Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14 (6): 587–89.

Kaufman, E. J., & Miska, E. A. (2010). The microRNAs of Caenorhabditis elegans. *Seminars in Cell and Developmental Biology*, *21*(7), 728–737. https://doi.org/10.1016/j.semcdb.2010.07.001

Kawahara, Y. (2008). [MicroRNAs in neurobiology]. *Brain and Nerve = Shinkei Kenkyu No Shinpo*, *60*(12), 1437–1444. Retrieved from https://europepmc.org/article/med/19110755

Khurana, J. S., & Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. *Journal of Cell Biology*, *191*(5), 905–913.

Khvorova, A., Reynolds, A., & Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, *115*(2), 209–216. https://doi.org/10.1016/S0092-8674(03)00801-8

Kim, V. N., Han, J., & Siomi, M. C. (2009). *Biogenesis of small RNAs in animals*. *10*(fEBRuARy). https://doi.org/10.1038/nrm2632

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). MiRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Krol, J., Loedige, I., & Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics*, *11*(9), 597–610. https://doi.org/10.1038/nrg2843

Lai, E. C., Tomancak, P., Williams, R. W., & Rubin, G. M. (2003). Computational identification of Drosophila microRNA genes. *Genome Biology*, *4*(7), 1–20. https://doi.org/10.1186/gb-2003-4-7-r42

Lai, X., & Vera, J. (2013). MicroRNA clusters. *Encyclopedia of Systems Biology. New York: Springer*

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Tuschl, T.

(2007). A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell*, *129*(7), 1401–1414. https://doi.org/10.1016/j.cell.2007.04.040

Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., & Bartel, D. P. (2003). *Vertebrate MicroRNA Genes*. *299*(March), 1.

Liu, J. (2008). Control of protein synthesis and mRNA degradation by microRNAs. *Current Opinion in Cell Biology*, *20*(2), 214–221.

Magurran, A. E. (2005). *Evolutionary ecology: the Trinidadian guppy*. Oxford University Press on Demand.

Magurran, A. E., & Seghers, B. H. (1990). Population differences in predator recognition and attack cone avoidance in the guppy Poecilia reticulata. *Animal Behaviour*, *40*(3), 443–452. https://doi.org/https://doi.org/10.1016/S0003-3472(05)80524-X

Malone, C. D., & Hannon, G. J. (2009). Molecular evolution of piRNA and transposon control pathways in Drosophila. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*, 225–234. https://doi.org/10.1101/sqb.2009.74.052

Meijer, H. A., Smith, E. M., & Bushell, M. (2014). Regulation of miRNA strand selection: Follow the leader? *Biochemical Society Transactions*, *42*(4), 1135–1140. https://doi.org/10.1042/BST20140142

Motti, D., Bixby, J. L., & Lemmon, V. P. (2012). MicroRNAs and neuronal development. *Seminars in Fetal and Neonatal Medicine*, *17*(6), 347–352. https://doi.org/10.1016/j.siny.2012.07.008

Nguyen, Lam-Tung, Heiko A Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies."*Molecular Biology and Evolution* 32 (1): 268–74.

Nordell, S. E. (1998). The response of female guppies, Poecilia reticulata, to chemical stimuli from injured conspecifics. *Environmental Biology of Fishes*, *51*(3), 331–338. https://doi.org/10.1023/A:1007464731444

Okamura, K., Liu, N., & Lai, E. C. (2009). Distinct Mechanisms for MicroRNA Strand Selection by Drosophila Argonautes. *Molecular Cell*, *36*(3), 431–444. https://doi.org/10.1016/j.molcel.2009.09.027

Pasquinelli, A. E., & Ruvkun, G. (2003). Control of Developmental Timing by MicroRNAs and Their Targets. *Https://Doi.Org/10.1146/Annurev.Cellbio.18.012502.105832*, *18*, 495–513. https://doi.org/10.1146/ANNUREV.CELLBIO.18.012502.105832

Penrith, M.-L. (1970). *The distribution of the fishes of the family Clinidae in southern Africa*. South African Museum.

Phillips, T. (2008). Regulation of Transcription and Gene Expression in Eukaryotes. *Nature Education*, *1*(1), 199.

Pillai, R. S., & Chuma, S. (2012). piRNAs and their involvement in male germline development in mice. *Development Growth and Differentiation*, *54*(1), 78–92. https://doi.org/10.1111/j.1440-169X.2011.01320.x

Reddy, K. B. (2015). MicroRNA (miRNA) in cancer. *Cancer Cell International*, *15*(1), 1–6. https://doi.org/10.1186/S12935-015-0185-1/FIGURES/1

Renn, S. C. P., & Schumer, M. E. (2013). SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION Genetic accommodation and behavioural evolution : insights from genomic studies SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION. *Animal Behaviour*, *85*(5), 1012–1022. https://doi.org/10.1016/j.anbehav.2013.02.012

Reznick, D. A., Bryga, H., & Endler, J. A. (1990). Experimentally induced life-history evolution in a natural population: Discovery Service for Concordia University Library. *Nature Letters*, *346*(July), 357–359.

Reznick, D., Butler IV, M. J., & Rodd, H. (2001). Life-history evolution in guppies. VII. The comparative ecology of high- and low-predation environments. *American Naturalist*, *157*(2), 126–140. https://doi.org/10.1086/318627

Reznick, D. N., & Bryga, H. (1987). Life-History Evolution in Guppies (Poecilia reticulata): 1. Phenotypic and Genetic Changes in an Introduction Experiment. *Evolution*, *41*(6), 1370. https://doi.org/10.2307/2409101

Reznick, D. N., Shaw, F. H., Rodd, F. H., & Shaw, R. G. (1997). Evaluation of the rate of evolution in natural populations of guppies (Poecilia reticulata). *Science*, *275*(5308), 1934–1937. https://doi.org/10.1126/science.275.5308.1934

Roberts, J. T., & Borchert, G. M. (2017). Computational prediction of microRNA target genes, target prediction databases, and web resources. *Bioinformatics in MicroRNA Research*, 109–122.

Ruell, E. W., Handelsman, C. A., Hawkins, C. L., Sofaer, H. R., Ghalambor, C. K., & Angeloni, L. (2013). Fear, food and sexual ornamentation: Plasticity of colour development in Trinidadian guppies. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1758). https://doi.org/10.1098/rspb.2012.2019

Scaria, V., Hariharan, M., Maiti, S., Pillai, B., & Brahmachari, S. K. (2006). Host-virus interaction: A new role for microRNAs. *Retrovirology*, *3*, 1–9. https://doi.org/10.1186/1742-4690-3-68

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864.

Scrimshaw, N. (1944). *Superfetation in Poeciliid Fishes*. *1944*(3), 180–183.

Seitz, H., Tushir, J. S., & Zamore, P. D. (2011). A 5′-uridine amplifies miRNA/miRNA* asymmetry in Drosophila by promoting RNA-induced silencing complex formation. *Silence*, *2*(1), 1–10.

Senti, K. A., & Brennecke, J. (2010). The piRNA pathway: A fly's perspective on the guardian of the genome. *Trends in Genetics*, *26*(12), 499–509.

Seto, A. G., Kingston, R. E., & Lau, N. C. (2007). The Coming of Age for Piwi Proteins. *Molecular Cell*, *26*(5), 603–609. https://doi.org/10.1016/j.molcel.2007.05.021

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, *8*(4), 272–285.

Starega-Roslan, J., Koscianska, E., Kozlowski, P., & Krzyzosiak, W. J. (2011). The role of the precursor structure in the biogenesis of microRNA. *Cellular and Molecular Life Sciences*, *68*(17), 2859–2871. https://doi.org/10.1007/s00018-011-0726-2

Tarver, J. E., Sperling, E. A., Nailor, A., Heimberg, A. M., Robinson, J. M., King, B. L., … Peterson, K. J. (2013). Mirnas: Small genes with big potential in metazoan phylogenetics. *Molecular Biology and Evolution*, *30*(11), 2369–2382. https://doi.org/10.1093/molbev/mst133

Torres-Dowdal, J., Handelsman, C. a, Reznick, D. N., & Ghalambor, C. K. (2012). Local adaptation and the evolution of of phenotypic plasticity in Trinidadian guppies (Poecilia reticulata). *Evolution*, *66*(11), 3432–3443. https://doi.org/10.5061/dryad.84gf5

Torres Dowdall, J., Handelsman, C. A., Ruell, E. W., Auer, S. K., Reznick, D. N., & Ghalambor, C. K. (2012). Fine-scale local adaptation in life histories along a continuous environmental gradient in Trinidadian guppies. *Functional Ecology*, *26*(3), 616–627.

Tsuchiya, S., Okuno, Y., & Tsujimoto, G. (2006). MicroRNA: Biogenetic and functional mechanisms and involvements in cell differentiation and cancer. *Journal of Pharmacological Sciences*, *101*(4), 267–270. https://doi.org/10.1254/jphs.CPJ06013X

Tzur, G., Israel, A., Levy, A., Benjamin, H., Meiri, E., Shufaro, Y., … Galun, E. (2009). Comprehensive gene and microRNA expression profiling reveals a role for microRNAs in human liver development. *PLoS ONE*, *4*(10). https://doi.org/10.1371/journal.pone.0007511

Veith, W. J. (1979). Reproduction in the live-bearing teleost Clinus Superciliosus.

South African Journal of Zoology, *14*(4), 208–211.

Venkatesh, B., Tan, C. H., & Lam, T. J. (1990). Steroid hormone profile during gestation and parturition of the guppy (Poecilia reticulata). *General and Comparative Endocrinology*, *77*(3), 476–483. https://doi.org/10.1016/0016-6480(90)90239-I

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America*, *102 Suppl*(2), 6543–6549. https://doi.org/10.1073/pnas.0501844102

Willing, E. M., Bentzen, P., Van Oosterhout, C., Hoffmann, M., Cable, J., Breden, F., … Dreyer, C. (2010). Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology*, *19*(5), 968–984. https://doi.org/10.1111/j.1365-294X.2010.04528.x

Winter, J., Jung, S., Keller, S., Gregory, R. I., & Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology*, *11*(3), 228–234.

Wourms, J. P. (1981). Viviparity: The maternal-fetal relationship in fishes. *Integrative and Comparative Biology*, *21*(2), 473–515. https://doi.org/10.1093/icb/21.2.473

Wu, J., Cai, H., Xiang, Y. B., Matthews, C. E., Ye, F., Zheng, W., … Shu, X. O. (2018). Intra-individual variation of miRNA expression levels in human plasma samples. *Biomarkers*, *23*(4), 339–346. https://doi.org/10.1080/1354750X.2018.1427794.

Yamanaka, S., Siomi, M. C., & Siomi, H. (2014). PiRNA clusters and open chromatin structure. *Mobile DNA*, *5*(1), 1–12. https://doi.org/10.1186/1759-8753-5-22/FIGURES/7

Zhang, Y. F., Zhang, R., & Su, B. (2009). Diversity and evolution of MicroRNA gene clusters. *Science in China, Series C: Life Sciences*, *52*(3), 261–266. https://doi.org/10.1007/s11427-009-0032-5

CHAPTER 3: EVOLUTION OF MIRNA EXPRESSION PLASTICITY

Introduction

Phenotypic plasticity is the capacity of a single genome to produce multiple distinct phenotypes in response to environmental inputs. Behavior provides many striking examples of animal's flexibility when confronted with diverse environmental conditions (Ghalambor et al., 2010). Underlying plastic phenotypic variation is differential regulation of gene expression, either for single genes or networks of genes. Phenotypic plasticity at the organismal level, ie. plasticity in morphology, physiology, and behavior, have been the primary historical focus of phenotypic plasticity studies, but examining the molecular mechanisms by which gene expression plasticity evolves undoubtedly will provide a deeper understanding of plasticity at the organismal level (Renn and Schumer 2013). The advancement of genomic and transcriptomic sequencing through Next Generation Sequencing has allowed for the sampling of virtually all genes, and now also in organisms that historically had few genomic and genetic tools available (Renn and Schumer 2013). These genomic advances have begun to provide compelling data sets for comparative studies of the evolutionary processes that are dependent on changes in phenotypic plasticity, e.g. (Aubin-Horth and Renn 2009; Beldade, Mateus, and Keller 2011). A deeper understanding of the evolution of gene expression plasticity will further our insight into the forces limiting and driving the evolution of complex traits and further our understanding of how organisms respond to dynamic environments (Fischer et al. 2021).

All organisms encounter environmental variation, and so the ubiquitous presence of phenotypic plasticity in biology is expected, i.e. because plasticity can enhance fitness by reducing mismatch between environment and organism's phenotype (C. K. Ghalambor et al. 2007). West-

Eberhard (2003) formally defines genetic accommodation as a mechanism of evolution that refines novel phenotypes (generated in response to mutations or environmental perturbations) into adaptive phenotypes through serial quantitative genetic changes. The series of quantitative genetic changes lead to a change in plasticity that largely reflects the demands of the particular environment. Generally, high phenotypic plasticity in a trait is selected when the organism is confronted by environmental variation, e.g. when no fixed trait is appropriate for all environmental conditions (DeWitt and Scheiner 2004; Ehrenreich and Pfennig 2016), when reliable environmental cues signal local change (Levin et al. 2012), and when the costs of plasticity are outweighed by its fitness benefits (Whitman and Agrawal, 2009).

One form of gene expression plasticity that has received less attention are microRNAs (miRNAs), a class of small non-coding RNAs (sRNAs), that typically average ~21nt in length and function in transcriptional and translational silencing (Ambros 2004; Bartel 2004; Chen and Rajewsky 2007). MiRNAs play a significant role in the modulation of a multitude of physiological (and pathopysiological) processes (Bhaskaran and Mohan 2014). Animals that fail to express miRNAs do not survive or have abnormal reproduction (Ketting et al. 2001; Wienholds et al. 2003; Kloosterman and Plasterk 2006). Most mammalian mRNAs are believed to be conserved targets of miRNA regulation, and more than 60% of human genes are estimated to be regulated by miRNAs at the translational level (Friedman et al. 2009). Although miRNA mediated regulation of gene expression has become a major focal point for phenotypic plasticity research, especially in biomedical research, lacking from this depth of knowledge on miRNAs is their *evolutionary* role in the formation of complex phenotypes.

This chapter examines miRNA plasticity in two populations of guppies to address how plasticity diverges. We ask whether plasticity in miRNAs largely was lost or gained when

guppies colonized low predation habitats, or whether plasticity tended to change extent or direction. The prevalence of miRNA expression patterns can provide insights into the relationship between the evolution of expression plasticity and organismal traits. Fig. 3.1 describes general patterns possible when plasticity evolves. If most miRNAs exhibit an accommodation pattern, it suggests that changes in gene expression plasticity are associated with underlying changes in organismal traits. On the other hand, a predominance of assimilation would indicate a loss of miRNA expression plasticity, which may be necessary for the variation in organismal traits between ancestral and derived populations. By regulating mRNA targets, changes in miRNA expression plasticity can directly modify organismal phenotypes, i.e. through alterations in the mRNA network.

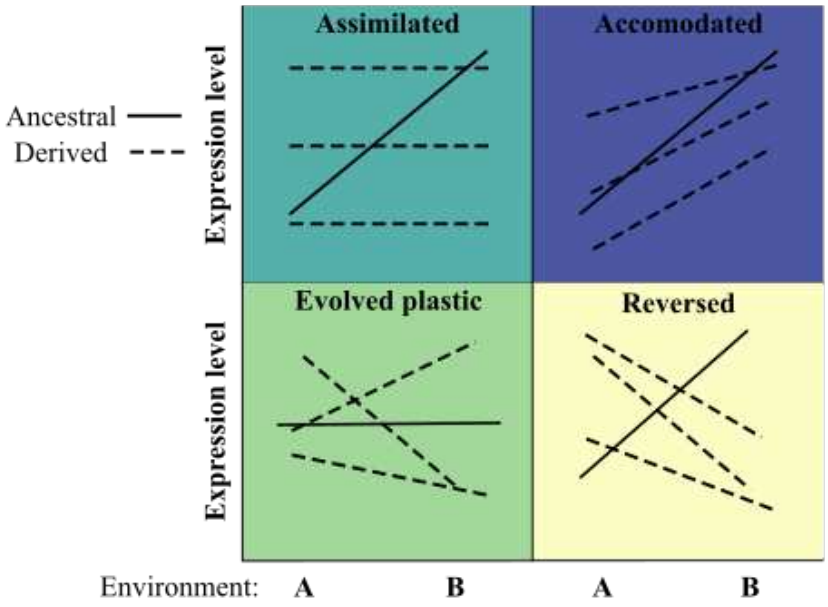

Figure 3.1: Four patterns of evolved plasticity proposed for categorizing evolved miRNA expression plasticity. The solid line represents the degree of plasticity of the ancestral population when exposed to two different environments. The dashed line represents the derived population. *Assimilated* is an ancestral to derived change from plasticity to no plasticity. *Evolved plastic* is the opposite pattern, a change from no plasticity to plasticity. *Accommodated* is a change in the magnitude of plasticity. In the *reversed* pattern both populations exhibit plasticity, and the direction of plasticity is opposite each other. Adapted from (Renn & Schumer, 2013).

Given the functional relationship of miRNAs targeting mRNAs leading to the suppression of translation or increased mRNA decay rates (with some exceptional miRNAs actually increasing the abundance of their targets) (Xu et al. 2022), we decided an analysis of this major transcriptional and translational regulator was a logical next step. The current study uses a small RNA-seq (sRNA-seq) approach that complements both the (Fischer et al. 2021) and (Ghalambor et al. 2015) studies to determine if patterns of evolved plasticity in *miRNAs* are consistent or deviate from previously established patterns of evolved *mRNA* expression plasticity.

Trinidadian guppies (*Poecilia reticulata*) have become a valuable model system for studying ecology and evolution due to their repeated adaptation to low and high predation environments (Fig 3.2; Endler 1995; D. Reznick, Butler IV, and Rodd 2001; Fischer et al. 2021). These adaptations have occurred independently in paired populations across multiple river drainages, representing separate evolutionary lineages (Barson, Cable, and Van Oosterhout 2009b; Gilliam, Fraser, and Alkins-Koo 1993; Willing et al. 2010; Fischer et al. 2021). The guppies exhibit parallel adaptations in various morphological, and behavioral traits (Endler 1995; Magurran 2005; D. N. Reznick 1997; D. A. Reznick, Bryga, and Endler 1990; D. Reznick, Butler IV, and Rodd 2001). While previous research has focused on whole organism phenotypes, little is known about the molecular mechanisms underlying these adaptive differences (Fischer et al. 2021). Two studies have looked at brain gene expression in the guppy high and low predation context. Ghalambor et al. 2015 examined gene expression in guppy brains and found an inverse relationship between the direction of plasticity and gene expression evolution. Fischer et al, 2021 investigated mRNA plasticity and divergence during parallel adaptation in guppies, revealing nonparallel transcriptional divergence and the evolution of plasticity in mRNA expression.

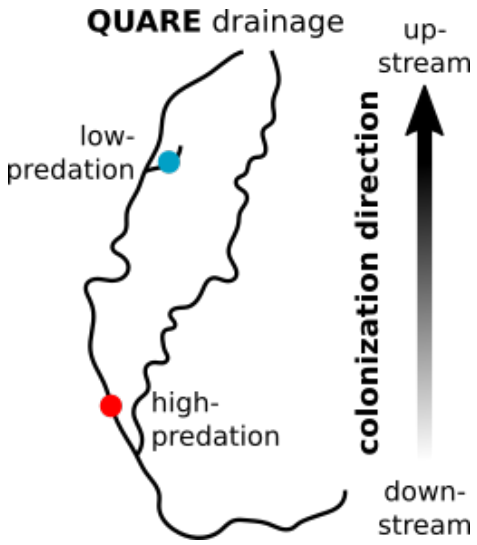

Figure 3.2: The ancestral high (QH) and derived low (CM) predation populations in the Quare drainage. Low predation guppy populations were established from guppies swimming up waterfalls and colonizing upstream habitats that are absent of their cichlid fish predators. Adapted from (Fischer et al. 2021)

This chapter will explore the population differences in miRNA expression, as well as the effect of predation on miRNA expression, leveraging two natural populations the Quare (QH) high predation and Campo (CM) low predation populations that are native to Trinidad (Fig. 3.2). Most notably we analyzed the interactions of population and rearing condition, i.e. high or low predation evolutionary history and reared in the presence or absence of predator cues. We characterized these interaction effects using the reaction norm framework, as established by Renn and Schumer (2013) and used in the Fischer et al. (2021) experiments. We are asking two primary questions; does expression plasticity itself evolve? If so, we can also ask, what qualitative patterns are observed in the evolution of expression plasticity?

Methods

Husbandry

These are the same animals as described in CH2 methods *husbandry*. Briefly, guppies were collected from high (Quare) and low predation (Campo) sites in Trinidad. We established unique families and reared guppies with or without predator cues. All experimental methods were approved by the Colorado State University Animal Care and Use Committee (Protocol #12-3818A) (E. K. Fischer et al., 2016).

Table 3.3: 2x2 factorial study design. Quare high-predation (QH) guppies were split into two rearing conditions either with (QH+) or without (QH-) predator cues. Campo low-predation (CM) guppies followed the same regiment producing two additional experimental groups CM+ and CM-.

	Predator Cue Present (+)	Predator Cue Absent (-)
Quare high-predation **(QH)**	**QH+** (N = 12)	**QH-** (N = 11)
Campo low-predation **(CM)**	**CM+** (N = 15)	**CM-** (N = 13)

Tissue collection, RNA extraction, and sRNA library construction

These are the same male whole brain samples as described in methods, CH2. Briefly, fish were anesthetized and brains were removed. Whole brains were flash frozen in liquid nitrogen and homogenized. We extracted total RNA from brain tissue using Qiagen RNeasy Lipid Tissue Mini Kit (Qiagen, Germany) following manufacturer guidelines (E. K. Fischer et al., 2016). We prepared sRNA libraries using the NEBNext® Multiplex Small RNA Library

Prep Set for Illumina® - primer set 1 (New England Biolabs, Massachusetts, USA) following manufacturer instructions. 51 brain libraries were prepared and size selected using BluePippin (Sage Science, MA, USA). We used TapeStation (D1000 ScreenTape, Part no. 5067-5582) to check the distribution of library molecule lengths. We pooled samples with unique barcodes and balanced experimental groups across sequencing lanes. Libraries were sequenced on an Illumina NextSeq500 at Colorado State University.

Differential expression analysis

As a first step, we performed a differential expression (DE) analysis on our miRNA data set. We wrote a custom R script to for statistical modeling of differences in expression across population, rearing condition, and the interaction of the two. To test for miRNA differential expression, we used a generalized linear mixed model (GLMM), and family was included as a random variable. The R function glmer.nb() from the package lme4 Bates D, Mächler M, Bolker B, Walker S (2015) was used as the base function in the script, which uses a negative binomial probability function. Model estimates and p-values for population, rearing, and interaction effects (as well as simple contrasts) were calculated from the test() and lsmeans() functions with formula, ~Pop:Rearing. We calculated FDR adjusted p-values for population, rearing, and interaction effects using the R package 'fdrtool' and the fdrtool() function with the default cutoff method set to default 'fndr'. We considered any effects with FDR ≤ 0.05 to be significant.

Categorization of evolved patterns of plasticity

As per Renn and Schumer (2013) the 'patterns of evolved plasticity' in gene expression can be characterized using four categories of reaction norms, *assimilated*, *accommodated*, *evolved plastic*, and *reversed* (see Fig. 3.1). We assessed patterns of plasticity based on evaluating posthoc simple effects only after finding a significant interaction effect. Interaction effects were consid-

ered significant if they had a false discovery rate less than 0.05. Specifically we used the following technical definitions, *accommodation* refers to an increase or fine adjustment to plasticity between ancestral (QH) and derived (CM) populations (that doesn't otherwise fit the assimilation, evolved plastic or reversed categories). The *assimilation* pattern was represented by QH plasticity that is lost in the CM population. E*volved plastic* indicated the QH population exhibited no plasticity, and CM did exhibit plasticity. Lastly, *reversed* indicated plasticity was observed in both populations, but the direction of plasticity was opposite of each other.

Results

Differential expression analysis

Our results showed a total of 181 (18.2% of the 995 total miRNAs analyzed) differentially expressed miRNAs including main effects of population, rearing condition as well as population by rearing interactions (Fig. 3.4).

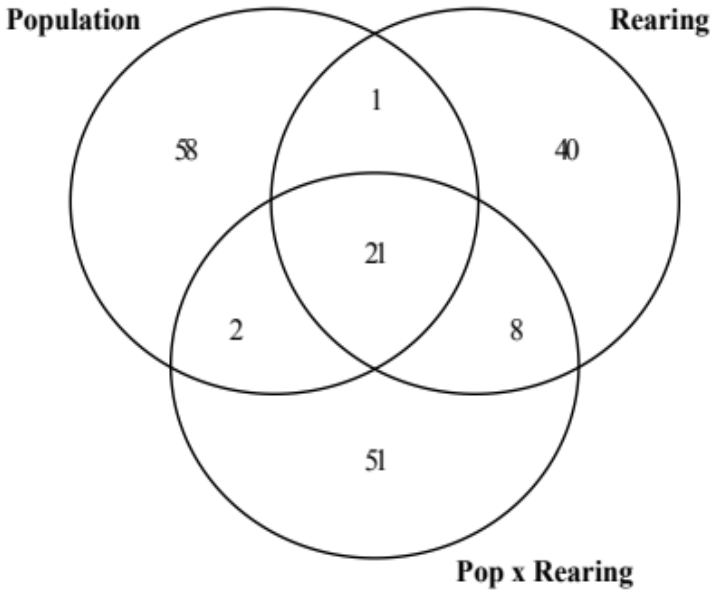

Figure 3.4: Overview of miRNA differential expression results. Interaction effects were the most common (N=82, 45.3%), followed by population main effects (N=58, 32.0% of DE miRNAs), and rearing effects (N=51, 28.2%).

Population main effects

Log2 fold changes of the population significant DE miRNAs ranged from -6.03 to 6.02. The QH population had 31 DE miRNAs with higher expression compared to CM, and 28 had higher levels in CM as compared to QH. Population DE miRNAs were mainly distributed in the mid-and lower expression level, i.e. log10 mean expression levels ranging from -1.14, 1.67 (Fig. 3.2).

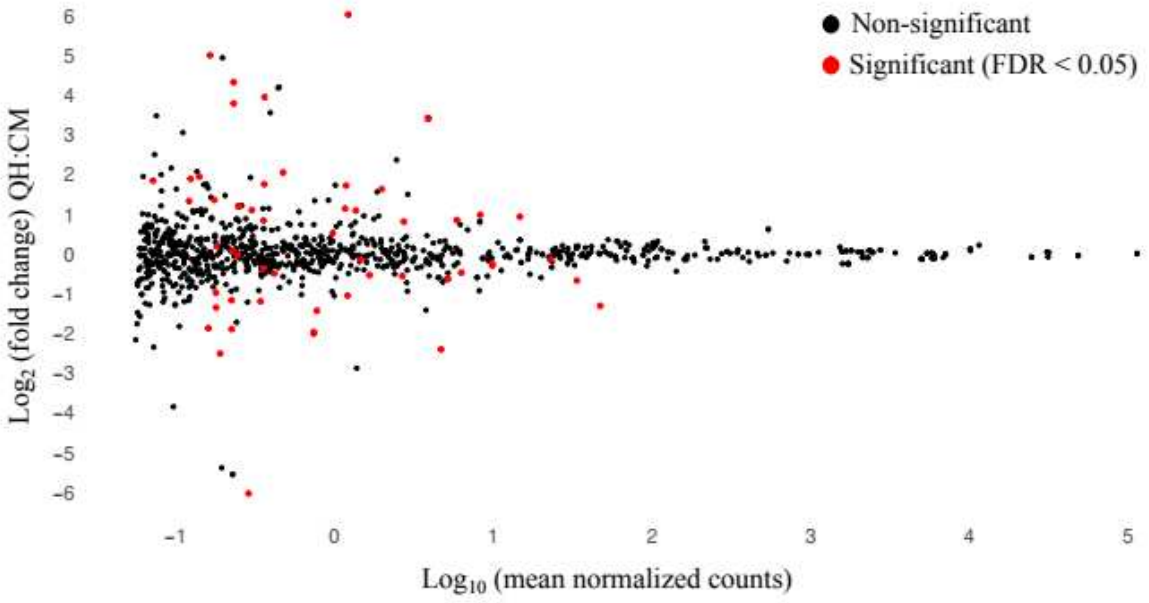

Figure 3.5: Differentially expressed genes were approximately equally likely to be up- and down-regulated in miRNAs with low to moderate abundances. The Y-axis is the difference in expression as Log_2(fold change) QH : CM. The X-axis is the Log_{10} of the mean normalized counts. N = 59 DE miRNAs, FDR < 0.05.

Rearing condition main effects

Rearing condition main effects were the least common of the significant DE miRNAs (N= 41, after intersecting interaction effects were removed) (FDR < 0.05) (Fig. 3.1 and Fig 3.3).

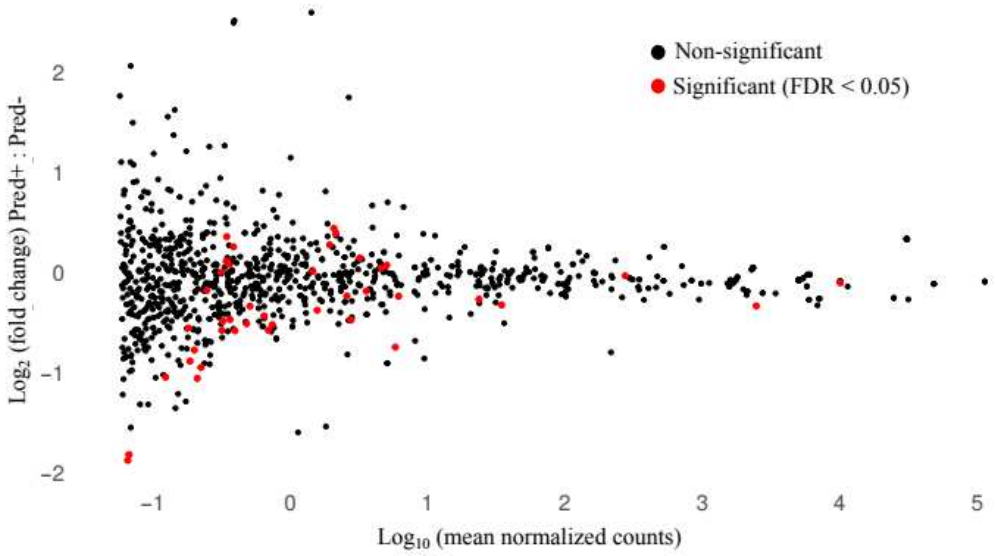

Figure 3.6: Differentially expressed genes were biased towards up-regulation in predation absent condition. MiRNAs with differential expression were observed from low to high abundances. The Y-axis is the difference in expression as Log2(fold change) predator cue present : predator cue absent. The X-axis is the Log10 of the mean normalized counts. N = 41 DE miRNAs (FDR < 0.05).

Evolution of miRNA expression plasticity

We detected 82 miRNAs with evolved expression plasticity, i.e. miRNAs with significant population-by-rearing interaction effects (FDR < 0.05). For miRNAs with significant interaction effects, we observed a similar number of genes expressing plasticity within each population. We found 37 miRNAs exhibited plasticity in both populations. To characterize the patterns of evolved plasticity we used the previously described reaction norm framework (Fig 3.1), we classified all but 1 interaction effects into one of four categories (Fig 3.7). All four categories were represented, and fairly evenly distributed. *Reversed* (plasticity in both populations, in opposite directions) and *evolved plastic* (no plasticity in ancestral, plasticity in derived) were the largest categories (N=24, 29%), and (N= 23, 28%), followed by *assimilated*

(ancestral plasticity, no derived plasticity) (N=21, 26%), and *accommodated* (change in magnitude of plasticity between ancestral and derived) (N= 13, 14%) (Fig. 3.7).

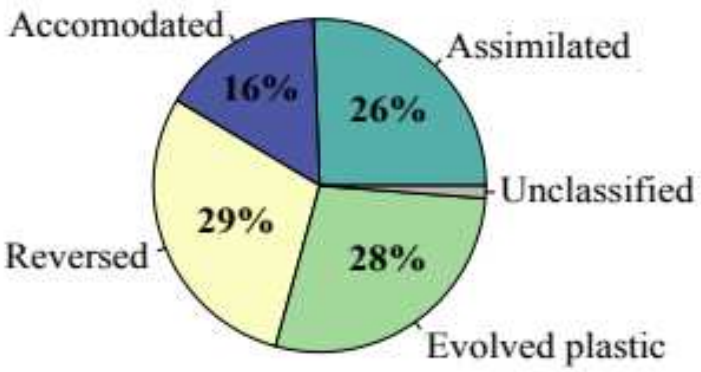

Figure 3.7: Pie chart of percentages of each DE population-by-rearing miRNAs (N= 43) within the four proposed categories of evolved plasticity. These categories are based on simple effects with a FDR less than 0.05.

Discussion

We found miRNAs that differed in expression levels between populations, rearing condition, and their interaction (Fig 3.4). Interestingly, we discovered a similar percentage (~18%) of the known guppy brain miRNAome (N=995 brain miRNAs) was differentially expressed as was found in Fischer et al. (2021) mRNA analysis (~22% or 4,387 total Quare drainage differentially expressed mRNAs from a total set of 19,902).

Differential miRNA expression between populations has genetic causes, and we interpret population main effects as originating from genetic background differences between the two populations (divergence), leading to differential regulation (through either direct or indirect mechanisms) of miRNA expression.

The responsiveness of miRNAs to environment differs based on genomic background, as we found multiple significant interaction effects. We observed many population by rearing effects indicating both plasticity in miRNA expression as well the evolution of plasticity itself. Since we did observe interaction effects, we were then able to perform the second analysis to determine what categories of evolved expression plasticity are represented within out data set. We found equal numbers of gains of plasticity, losses of plasticity, changes in extent of plasticity, and changes in direction of plasticity as high-predation fish colonized low-predation environments (Fig. 3.7). Fischer et al. (2021) showed that all four patterns of evolved expression, i.e. ancestral plasticity gain, ancestral plasticity lost, change in plasticity, or reversed plasticity were present. Due to the function of miRNAs in targeting mRNAs, the similar distribution of evolved plasticity in miRNAs and mRNAs is consistent with our expectations.

In this study we assessed effects of population, predation-based rearing condition, and the interaction of the two, on miRNA expression from whole brain small RNAs. We found numerous miRNAs exhibiting differential expression, indicating both influence from genetic background, as well as miRNAs that specifically respond to predator based environmental variation. Lastly, we characterized the patterns of evolved plasticity in a similar manner to that described by (Renn & Schumer, 2013), and can be compared to the mRNA characterization of evolved plasticity (Fischer et al. 2021), as these were the same animals used in both studies. Future work may aim towards linking gene expression patterns to organismal traits.

CHAPTER 3 WORKS CITED

Alexander, Heather J., John S. Taylor, Sampson Sze-Tsun Wu, and Felix Breden. 2006. "PARALLEL EVOLUTION AND VICARIANCE IN THE GUPPY (POECILIA RETICULATA) OVER MULTIPLE SPATIAL AND TEMPORAL SCALES." *Evolution* 60 (11): 2352–69. https://doi.org/10.1111/J.0014-3820.2006.TB01870.X.

Ambros, Victor. 2004. "The Functions of Animal MicroRNAs." *Nature* 431 (7006): 350–55. https://doi.org/10.1038/nature02871.

Aubin-Horth, Nadia, and Susan C.P. Renn. 2009. "Genomic Reaction Norms: Using Integrative Biology to Understand Molecular Mechanisms of Phenotypic Plasticity." *Molecular Ecology* 18 (18): 3763–80. https://doi.org/10.1111/J.1365-294X.2009.04313.X.

Barson, N. J., J. Cable, and C. Van Oosterhout. 2009a. "Population Genetic Analysis of Microsatellite Variation of Guppies (Poecilia Reticulata) in Trinidad and Tobago: Evidence for a Dynamic Source-Sink Metapopulation Structure, Founder Events and Population Bottlenecks." *Journal of Evolutionary Biology* 22 (3): 485–97. https://doi.org/10.1111/j.1420-9101.2008.01675.x.

Bartel, David P. 2004. "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function Review." *Biosensors and Bioelectronics* 116: 281–97. https://doi.org/10.1016/j.bios.2017.04.001.

Beldade, Patrãcia, Ana Rita A. Mateus, and Roberto A. Keller. 2011. "Evolution and Molecular Mechanisms of Adaptive Developmental Plasticity." *Molecular Ecology* 20 (7): 1347–63. https://doi.org/10.1111/j.1365-294X.2011.05016.x.

Bhaskaran, M., and M. Mohan. 2014. "MicroRNAs: History, Biogenesis, and Their Evolving Role in Animal Development and Disease." *Veterinary Pathology* 51 (4): 759–74. https://doi.org/10.1177/0300985813502820.

Carvalho, G., P. W. Shaw, A. E. Magurran, and B. H. Segers. 1991. "Marked Genetic Divergence Revealed by Allozymes among Populations of the Guppy Poecilia Reticulata (Poeciliidae), in Trinidad." *Biological Journal of the Linnean Society* 42 (3): 389–405. https://doi.org/10.1111/J.1095-8312.1991.TB00571.X.

Chen, Kevin, and Nikolaus Rajewsky. 2007. "The Evolution of Gene Regulation by Transcription Factors and MicroRNAs." *Nature Reviews Genetics* 8 (2): 93–103. https://doi.org/10.1038/nrg1990.

DeWitt, Thomas J., and Samuel M. Scheiner. 2004. "Phenotypic Plasticity : Functional and Conceptual Approaches," 247.

Dukas, R. (2013). SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION Effects of learning on evolution : robustness , innovation and speciation SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION. *Animal Behaviour*, *85*(5), 1023–1030. http://dx.doi.org/10.1016/j.anbehav.2012.12.030

Ehrenreich, Ian M., and David W. Pfennig. 2016. "Genetic Assimilation: A Review of Its Potential Proximate Causes and Evolutionary Consequences." Annals of Botany. 2016. https://doi.org/10.1093/aob/mcv130.

Endler, John A. 1995. "Multiple-Trait Coevolution and Environmental Gradients in Guppies." *Philos. Tmns. R. Sot. London SW B* 13 (29): 107–22.

Fischer, Eva K., Youngseok Song, Kimberly A. Hughes, Wen Zhou, and Kim L. Hoke. 2021. "Nonparallel Transcriptional Divergence during Parallel Adaptation." *Molecular Ecology* 30 (6): 1516–30. https://doi.org/10.1111/mec.15823.

Friedman, Robin C., Kyle Kai How Farh, Christopher B. Burge, and David P. Bartel. 2009. "Most Mammalian MRNAs Are Conserved Targets of MicroRNAs." *Genome Research* 19 (1): 92. https://doi.org/10.1101/GR.082701.108.

Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. "Adaptive versus Non-Adaptive Phenotypic Plasticity and the Potential for Contemporary Adaptation in New Environments." *Functional Ecology* 21 (3): 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x.

Ghalambor, C., Angeloni, L., & Carroll, S. (2010). Behaviour as phenotypic plasticity. *Evolutionary Behavioural Ecology*, *December*, 90–107.

Ghalambor, Cameron K., Kim L. Hoke, Emily W. Ruell, Eva K. Fischer, David N. Reznick, and Kimberly A. Hughes. 2015. "Non-Adaptive Plasticity Potentiates Rapid Adaptive Evolution of Gene Expression in Nature." *Nature* 525 (7569): 372–75. https://doi.org/10.1038/nature15256.

Gilliam, James F., Douglas F. Fraser, and Mary Alkins-Koo. 1993. "Structure of a Tropical Stream Fish Community: A Role for Biotic Interactions." *Ecology* 74 (6): 1856–70. https://doi.org/10.2307/1939943.

Huey, Raymond B., Paul E. Hertz, and B. Sinervo. 2003. "Behavioral Drive versus Behavioral Inertia in Evolution: A Null Model Approach." *The American Naturalist* 161 (3): 357–66. https://doi.org/10.1086/346135.

Itzkowitz, M. S. N. (2001). Parental division of labour and the shift from minimal to maximal role specializations: An examination using a biparental fish. https://doi.org/10.1006/anbe.2000.1724

Ketting, R. F., S. E.J. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, and R. H.A. Plasterk. 2001. "Dicer Functions in RNA Interference and in Synthesis of Small RNA Involved in Develop-

mental Timing in C. Elegans." *Genes & Development* 15 (20): 2654–59. https://doi.org/10.1101/GAD.927801.

Kloosterman, Wigard P., and Ronald H.A. Plasterk. 2006. "The Diverse Functions of MicroR

NAs in Animal Development and Disease." *Developmental Cell* 11 (4): 441–50. https://doi.org/10.1016/J.DEVCEL.2006.09.009.

Künstner, Axel, Margarete Hoffmann, Bonnie A. Fraser, Verena A. Kottler, Eshita Sharma, Detlef Weigel, and Christine Dreyer. 2016. "The Genome of the Trinidadian Guppy, Poecilia Reticulata, and Variation in the Guanapo Population." *PLoS ONE* 11 (12): 1–25. https://doi.org/10.1371/journal.pone.0169087.

Lee, Rosalind C, Rhonda L Feinbaum, and Victor Ambros. 1993. "The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to Lin-14 Ambros and Horvitz, 1987). Animals Carrying a Lin-4 Loss-of-Function (Lf) Mutation, Lin-4(E912), Display Reiterations of Early Fates at Inappropriately Late Developmental Stages; Cell Lineage Patterns Normally Specific for the L1 Are Reiter." *Cell* 75: 843–54.

Levin, Simon, Stephen R. Carpenter, H. Charles J. Godfray, Ann P. Kinzig, Michel Loreau, Jonathan B. Losos, Brian Walker, and David S. Wilcove. 2012. "The Princeton Guide to Ecology," 3–6.

Magurran, Anne E. 2005. *Evolutionary Ecology: The Trinidadian Guppy*. Oxford University Press on Demand.

Pfennig, David W., Matthew A. Wund, Emilie C. Snell-Rood, Tami Cruickshank, Carl D. Schlichting, and Armin P. Moczek. 2010. "Phenotypic Plasticity's Impacts on Diversification and Speciation." *Trends in Ecology and Evolution* 25 (8): 459–67. https://doi.org/10.1016/j.tree.2010.05.006.

Price, T. D., A. Qvarnstrom, and D. E. Irwin. 2003. "The Role of Phenotypic Plasticity in Driving Genetic Evolution." *Proceedings of the Royal Society B: Biological Sciences* 270 (1523): 1433–40. https://doi.org/10.1098/rspb.2003.2372.

Rachagani, Satyanarayana, Ye Cheng, and James M. Reecy. 2010. "Myostatin Genotype Regulates Muscle-Specific MiRNA Expression in Mouse Pectoralis Muscle." *BMC Research Notes* 3 (1): 1–5. https://doi.org/10.1186/1756-0500-3-297/FIGURES/2.

Renn, Suzy C P, and Molly E Schumer. 2013. "SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION Genetic Accommodation and Behavioural Evolution : Insights from Genomic Studies SPECIAL ISSUE : BEHAVIOURAL PLASTICITY AND EVOLUTION." *Animal Behaviour* 85 (5): 1012–22. https://doi.org/10.1016/j.anbehav.2013.02.012.

Reznick, D., M. J. Butler IV, and H. Rodd. 2001. “Life-History Evolution in Guppies. VII. The Comparative Ecology of High- and Low-Predation Environments.” *American Naturalist* 157 (2): 126–40. https://doi.org/10.1086/318627.

Reznick, David A., Heather Bryga, and John A. Endler. 1990. “Experimentally Induced Life-History Evolution in a Natural Population: Discovery Service for Concordia University Library.” *Nature Letters* 346 (July): 357–59.

Reznick, David N., and Heather Bryga. 1987. “Life-History Evolution in Guppies (Poecilia Reticulata): 1. Phenotypic and Genetic Changes in an Introduction Experiment.” *Evolution* 41 (6): 1370. https://doi.org/10.2307/2409101.

Reznick, David N. 1997. “Life History Evolution in Guppies (Poecilia Reticulata): Guppies as a Model for Studying the Evolutionary Biology of Aging.” *Article in Experimental Gerontology*. https://doi.org/10.1016/S0531-5565(96)00129-5.

Rodd, F. 1997. “PHENOTYPIC PLASTICITY IN THE LIFE HISTORY TRAITS OF GUPPIES: RESPONSES TO SOCIAL ENVIRONMENT.” Ecology. 1997.

Schlichting, C. D., and M. Pigliucci. 1998. “Phenotypic Evolution: A Reaction Norm Perspective.” *Phenotypic Evolution: A Reaction Norm Perspective.*

Schlichting, Carl D. 2008. “Hidden Reaction Norms, Cryptic Genetic Variation, and Evolvability.” *Annals of the New York Academy of Sciences* 1133 (1942): 187–203. https://doi.org/10.1196/annals.1438.010.

Shaw, P. W., G. R. Carvalho, A. E. Magurran, and B. H. Seghers. 1991. “Population Differentiation in Trinidadian Guppies (Poecilia Reticulata): Patterns and Problems.” *Journal of Fish Biology* 39: 203–9. https://doi.org/10.1111/J.1095-8649.1991.TB05084.X.

Steen, Shyril O, Alistair J Cullum, and Albert F Bennett. 2014. “Rapid Evolution of Escape Ability in Trinidadian Guppies (Poecilia Reticulata) RAPID EVOLUTION OF ESCAPE ABILITY IN TRINIDADIAN GUPPIES (POECILIA RETICULATA)” 56 (4): 776–84.

Waddington, C. H. 1952. “Selection of the Genetic Basis for an Acquired Character.” *Nature* 169 (4294): 278. https://doi.org/10.1038/169278A0.

West-Eberhard, Mary-Jane. 2003. *Developmental Plasticity and Evolution*. New York: Oxford University Press.

Whitman, Douglas W, and Anurag A Agrawal. n.d. “What Is Phenotypic Plasticity and Why Is It Important?”

Wienholds, Erno, Marco J. Koudijs, Freek J.M. Van Eeden, Edwin Cuppen, and Ronald H.A. Plasterk. 2003. “The MicroRNA-Producing Enzyme Dicer1 Is Essential for Zebrafish Development.” *Nature Genetics 2003 35:3* 35 (3): 217–18. https://doi.org/10.1038/ng1251.

Willing, Eva Maria, Paul Bentzen, Cock Van Oosterhout, Margarete Hoffmann, Joanne Cable, Felix Breden, Detlef Weigel, and Christine Dreyer. 2010. “Genome-Wide Single Nucleotide Polymorphisms Reveal Population History and Adaptive Divergence in Wild Guppies.” *Molecular Ecology* 19 (5): 968–84. https://doi.org/10.1111/j.1365-294X.2010.04528.x.

Xu, Peng, Xianbin Li, Yijun Liang, Zhenshen Bao, Fengyue Zhang, Lili Gu, Saeed Kosari, and Wenbin Liu. 2022. “PmiRtarbase: A Positive MiRNA-Target Regulations Database.” *Computational Biology and Chemistry* 98: 107690.

CHAPTER 4: PREDATOR EXPOSURE AND GENETIC BACKGROUND DO NOT ALTER A-TO-I EDITING

Introduction

RNA editing generates post-transcriptional sequence modifications (Farajollahi & Maas, 2010; Maydanovych & Beal, 2006), primarily RNA nucleotide changes, including adenosine-to-inosine (A-to-I, recognized as guanine in translation and sequenced as A-to-G) (Ramaswami, Lin, Piskol, Tan, & Davis, 2012) and cytosine-to-uracil (C-to-U, sequenced as C-to-T) (Farajollahi & Maas, 2010; Picardi et al., 2015). At the center of the metazoan A-to-I RNA editing system is a class of adenosine deaminases called ADARs that introduce the A-to-I substitutions (Nishikura, 2010). During the process of translation, inosine is interpreted as guanosine. Consequently, A-to-I editing can lead to modifications in amino acid sequences. Aberrant editing of some sites has been associated with deleterious, even lethal phenotypic outcomes (Anne-Laure, 2010; Boris Zinshteyn, 2010). In most human tissues, A-to-I editing accounts for over 95% of all editing events (Bahn et al., 2012; Ramaswami et al., 2012; Q. Zhang & Xiao, 2015). A-to-I editing is found universally in metazoans (Grice & Degnan, 2015), and C-to-U editing is found both in animals and plants (Blanc & Davidson, 2003). Although RNA editing sometimes generates protein products distinct form those specified by the genomic DNA, RNA editing is also known to alter other functional outcomes including changes to alternative splicing (Rueter, Dawson, & Emeson, 1999), miRNA binding (Borchert et al., 2009), and numerous other biological processes (Blanc & Davidson, 2003; Zipeto, Jiang, Melese, & Jamieson, 2015). Further, RNA editing is thought to be a substantial contributor to transcriptome diversity (Fumagalli et al., 2015; Han et al., 2015; Paz-Yaacov et al., 2015). (Wahlstedt, Daniel, & Hman, 2009) showed an important role for RNA editing in brain development, and RNA

editing has been linked to multiple human diseases (Slotkin & Nishikura, 2013; Zipeto et al., 2015). Several adaptive explanations have been proposed for A-to-I editing, such as mutational buffering, gene regulation, proteome diversification, and optimization of genomic GC content. However, alternative non-adaptive mechanisms for the spread of editing sites have also been documented (Covello & Gray, 1993; Sloan, 2017). Furthermore, the influence of genetic background on A-to-I editing variation has not been well established. A major question remains, does A-to-I editing substantially contribute to natural phenotypic variation (attributable to either genetics or developmental plasticity), or are individual edit sites with aberrant editing that have been linked to major deleterious phenotypes (and have been largely studied outside of the context of natural populations) rare occurrences?

To determine whether A-to-I editing is has biologically meaningful impacts (in a natural context), we will examine whether they change in consistent ways based on genetic background, and a major ecological variable, predation pressure. The Trinidadian guppy system (Magurran, 2005) provides a natural context to test the phenomenon of A-to-I editing. Guppies exhibit local adaptions to high predation or low predation environments (Endler, 1995; Haskins, C. Haskins, E., McLaughlin, J., 1961) measured by life history, behavioral and morphological metrics, and developmental plasticity in response to predator cues alters many of these traits in ways that depend on the population of origin (Torres-Dowdal et al., 2012). We can leverage the Trinidad guppy system, using the natural phenotypic variation to determine if differential A-to-I editing depends on population of origin or developmental experience with predators.

In this study we model A-to-I editing in a transcriptome-wide fashion to aid in determining if A-to-I editing is associated with phenotypic variation, or if A-to-I editing is largely a factor of the stochastic nature of ADAR meditated editing. We investigated whether

intragenic A-to-I editing is associated with developmental plasticity or population divergence, or both, in whole brain (tissues known for high diversity of A-to-I editing). Our analysis reveals that A-to-I editing was not associated with population, rearing condition, or population by rearing interaction. Furthermore our analyses includes two separate lineages for the investigation of parallel evolution in A-to-I editing. Because we did not detect differential editing in either drainage, we conclude A-to-I editing likely does not play a major role in expression of developmental plasticity or population divergence in guppies.

Methods

Husbandry and rearing

To determine the influence of evolutionary history with predators and developmental experience with predators on whole brain A-to-I RNA editing patterns, we used a lab based breeding design (Fischer, Song, Hughes, Zhou, & Hoke, 2021) of which the same samples were used as (Fischer, et al. 2021). Our lab collected Trinidadian guppies from high-predation and low-predation population pairs in the Aripo (2012) and Quare (2014) river drainages. We established unique families (20-30 family lines) for each population in laboratory aquaculture from wild-caught gravid females (E. K. Fischer et al., 2016). The second-generation lab reared fish were established by crossing first-generation lab reared fish with unrelated first-generation fish from the same source population. Siblings from the second-generation matings were split into rearing environments with predator cues or without cues (E. K. Fischer et al., 2016). All experimental methods were approved by the Colorado State University Animal Care and Use Committee (Protocol #12-3818A).

RNA-seq library preparation and sequencing

For library preparation we used the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer's protocol to generate individual sample sequencing libraries. Libraries were sequenced on an Illumina HiSeq 2000 in May 2014 (Aripo data set) and January 2016 (Quare data set). The Quare data set had 58 samples (N = 13-15 per group), here group is defined as the particular level for population and rearing condition, e.g. high predation population and presence of predator cues rearing condition. The Aripo data set contained 36 samples (N = 8-10 per group). We balanced all experimental groups across sequencing lanes (Fischer et al., 2021).

Bioinformatics

Reads pre-processing and mapping

We trimmed Illumina adapter and primer sequences using Trim Galore! (Babraham Bioinformatics) following developer recommendations we used a quality score cutoff of 33, a stringency of five, and minimum read length of 36bp. As part of the SNP-free RNA editing IdeNtification Toolkit (SPRINT) (F. Zhang, Lu, Yan, Xing, & Tian, 2017) pipeline for the identification of RNA edit sites (RESs), we mapped the trimmed and quality filtered paired-end reads to the most recent guppy reference genome (assembly GCA_904066995.1) using BWA (v. 0.7.12) (Li & Durbin, 2009) with command options 'bwa aln fastqfile'. The sprint 'main' script was used for mapping and computational RES prediction. We used recommended SPRINT developer default options for the detection of regular RESs.

A-to-I RES prediction

We chose the SPRINT package because it accurately detects edits without a complete SNP database (F. Zhang et al., 2017). Briefly, we used the pipeline to first sort the mapping files

(BAM) using samtools (Li et al., 2009). The sorted BAM files were PCR deduplicated using picard-tools (version 1.119, http://broadinstitute.github.io/pic ard/) with the command option 'MarkDuplicates.jar REMOVE_DUPLICATES=true'. We considered reads with mapping quality ≥ 20 as mapped reads. We identified single nucleotide variants (SNVs) as mismatches between the guppy reference genome and mapped RNA-seq reads, and only SNVs with a base quality score of ≥ 25 were retained. There is a higher occurrence of mismatches near intron/exon junctions, primarily due to increased mapping errors near splice sites (F. Zhang et al., 2017). To avoid calling SNVs that are near splice sites, (Zhang et al. 2017) introduced 'fragment-loc', the distance of a mismatch to the nearest end of the mapped fragment it belongs to. We kept mismatches as SNVs with a 'fragement-loc' of 5 nucleotides or greater (see Zhang et al., 2017) for discussion on 'frag-loc'.

Ramaswami et al. (2012) observed that A-to-I RNA editing sites (RESs) tend to cluster in the genome, while the density of single nucleotide polymorphisms (SNPs) compared to RESs is low. Zhang et al. (2017) found that RES-based SNV duplets (pairs of SNVs with the same type of variation) are often located within a shorter distance of each other (within 400nt), whereas SNP-based SNV duplets are less frequently found in close proximity (within 1600nt). We implemented the suggested window size of 200nt to scan the guppy genome from the start to end position of each chromosome. SNV duplets sharing a common SNV are merged to form clusters, and the merging process stops when a different type of SNV duplet is encountered. The resulting clusters of SNV duplets are considered RESs if they meet a minimum size criterion. When the minimum cluster size and window size were set to 2 and 200nt, respectively, they found the false positive rate (1 – percentage of identified RESs that are actually SNPs; Q. Zhang & Xiao, 2015), as approximating 95%, and percentage of the true RESs identified as 80.3%.

These were the best parameters for both precision and recall; we used these same parameters for the detection of A-to-I editing in our RNA-seq data set. The SPRINT output files contain the depth count or total number of reads covering the RES, and supporting reads, a read count of reads exhibiting the A-to-G (or T-to-C, the reverse complement present for non-stranded RNA-seq data) mismatch between reference sequence and mapped RNA-seq read that also passed the mapping quality, baseseq quality, and fragment-loc filters. Depth is the count of reads covering the RES (RNA Editing Site) and supporting reads, which are the count of reads showing the A-to-G (or T-to-C reverse complement for our non-stranded RNA-seq data) mismatch between the reference sequence and mapped RNA-seq reads that passed the filters.

Analysis

RES genomic loci and pre-filtering

To determine the genomic loci of our RESs, and to remove non-RESs from our non-stranded RNA-seq data, we found the overlap between our RES set and gene annotations from the most recent guppy assembly annotations (Fraser et al., 2020). These annotations do not include untranslated regions (UTRs). We wrote a custom R script that checks each RES against the GFF file. We first selected all RESs that were within gene boundaries. See github repository for R script. We next filtered the set of intragenic RESs by keeping only A-to-G (+) strand and T-to-C (-) strand RESs. Because we have non-stranded RNA seq data (after second strand cDNA synthesis it becomes ambiguous as to which strand is the original transcript, hence the information indicating which DNA strand the transcript originated from), we also keep the reverse complement of A-to-G, T-to-C. We omitted ~41% of the RESs overlapping gene annotations with the types A-to-G (-) and T-to-C (+), indicating for SPRINT non-stranded RNA-

seq users, a rather large overestimation, if the A-to-G rate is simply estimated by A-to-G (+/-) and T-to-C (+/-).

From the concatenated sprint output we calculated the number of fish from each drainage editing the RES, as well as the sum of their supporting reads (A-to-G mismatch passing filtering requirements) and depth (total reads covering RES for all fish editing). We calculated "edit level" as the quotient of supporting reads (SR) to total reads (Depth). We pre-filtered the RES set before the quantitative analysis to those with a depth of at least 10 reads and at least 4 fish editing the RES (among both drainages), i.e. 4 fish with at least one supporting read, coming from either drainage. We considered these to be good confidence RESs predictions and were kept for further analyses, see github repository xx for R script..

Depth matrix

The sprint output does not include depth (total coverage) counts for non-editing fish, i.e. only fish with at least one supporting read have depth counts. To count depth for the non-editing fish for quantitative comparisons, we wrote a custom R script (github.com/mileswhedbee/AI_glmm) that checked for overlap of the RES genomic location with the mapped reads and counted all reads that mapped to each RES site for all fish.

Differential editing analysis

To analyze differences in editing (DE) across our high confidence RES set based on population and rearing conditions (with or without predator cues), we modeled each site using a GLMM in R. See github.com/mileswhedbee/AI_glmm for R script and data. The script uses the glmmTMB() function (Brooks et al. 2017) with the model design: $SR \sim Pop + Rearing + Pop{:}Rearing + Family + offset(DP)$, where *SR* is supporting reads, *Pop* is population, *Rearing* is

rearing condition, and *family* is a random variable included to control for genetic relatedness (siblings) among samples. We decided not to model "edit level" as our response variable to preserve information about total counts of both SR and depth. Instead we used SR as a count response variable while accounting for depth in the model as an offset. We initially ran the full pre-filtered set (N = 1027 sites) and evaluated p-value histograms (Pop, Rearing and Pop:Rearing terms) from the model output. All three histograms showed similar distributions with a single peak at ≥ 0.95. Upon inspection of the RESs with p-values in the ≥ 0.95 range, we increased the sum of supporting reads to ≥ 12, and also increased the number of fish required to be editing a site to at least 8. These filter parameters were empirically derived, so our DE set had stricter requirements than our characterization set, i.e. a second smaller DE set was generated from stricter filtering requirements that provided enough counts for statistical power to compare edit levels. Our model was designed to support zero-inflated SR count values by using the negative binomial probability model; however for some of our hypothesis testing, i.e. some of the RESs, the models did not converge and therefore were not able to be tested. The results of our fitted models were corrected for multiple hypotheses testing by false discovery rate (FDR). We used the R package "fdrtool" (Klaus and Strimmer, 2021) to calculate FDR values, and considered tests with FDRs less than 0.05 to be significant.

Results

Characterization of A-to-I editing in Aripo and Quare drainages

In the Aripo drainage (N = 36 individuals), 90.9% of the edits called by Sprint were consistent with A-to-I editing (genomic sequence of A and transcriptomic sequence of G, or genomic sequence of T and transcriptomic sequence of C; 351,140 out of a total of 386,092 calls)

(Fig. 4.1). The C-to-U rate was 6.4%, approximately double that of the non A-to-I or C-to-U rate (2.6%; Table 4.1). We observed in the Quare and Aripo datasets, 85-91% of the edits were A-to-I and 6.5%-7.6% of the edits were C-to-U (Table 4.1).

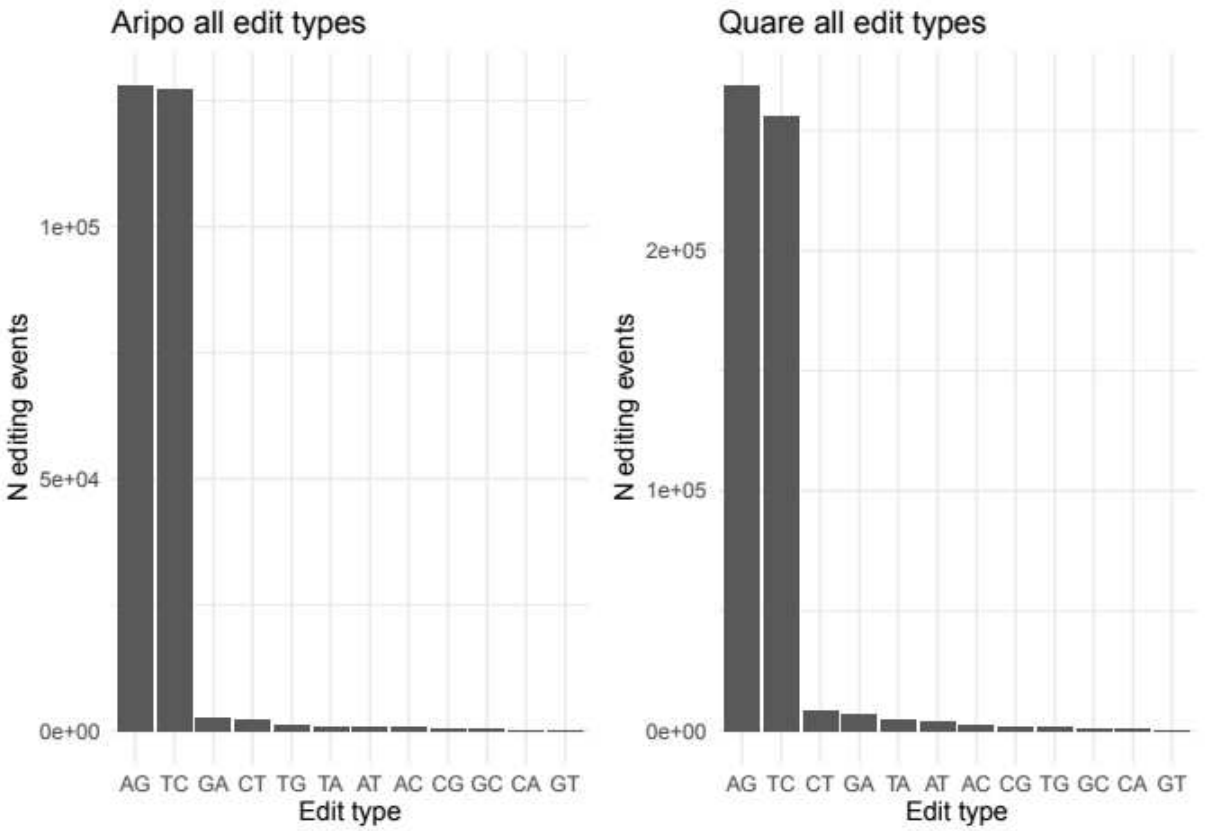

Figure 4.1. Evidence of A-to-I editing in guppy transcriptomes from two drainages. 'Edit type' refers to the difference between genomic and transcriptomic nucleotides, e.g. AG represents a mismatch between genomic 'A' and transcriptomic 'G'. The high abundance edit types are 'AG' and (reverse complement) 'TC', the expected biological edit types (A-to-I) in metazoans. The CT and GA, correspond to C-to-U editing are also elevated relative to the non-biological types.

Table 4.1: Summary of RNA edit type calls, and their respective rates for Aripo and Quare drainages.

Aripo (N = 36 fish)

	N	Rate
Total Editing events	386,092	
A-to-I (AG/TC +/-)	351,150	90.9%
C-to-U (CT/GA +/-)	24,882	6.4%
Non A-to-I or C-to-U	10,060	2.6%

Quare (N = 58 fish)

	N	Rate
Total Editing events	795,582	
A-to-I (AG/TC +/-)	681,704	85.7%
C-to-U (CT/GA +/-)	53,736	6.8%
Non A-to-I or C-to-U	60,142	7.6%

Table 4.2: Summary of intragenic and intergenic RES genomic loci. The majority of RESs were found in intergenic loci.

Item	N	Percentage total
Total RES	12,986	
Intragenic RESs	1,027	7.91%
Intergenic RESs	11,959	92.09%
Edited genes	110	
Mean RESs per gene	9.3	

RES Genomic loci

To determine the genomic loci of our RESs, and to obtain a more accurate A-to-I rate from non-stranded RNA-seq data, we determined the overlap of all called RESs and the most recent guppy gene annotations. We found 7.9% (N = 1,027) of the total RESs (combined from both drainages) were intragenic and 92.1% (N = 11,959) were intergenic (Table 4.2). The

intergenic regions also includes 5' and 3' UTRs. The intragenic RESs were distributed among 110 genes, with a mean of ~9 RESs per gene.

RES Drainage differences

We compared the Quare and Aripo RESs sets to determine the union and intersection of our characterization RES post-filtered set. Of the 1,027 sites found within genes, 30.6% (N = 314) were shared between drainages, 54.5% (N = 560) of the total set were unique to Quare (N = 58 samples), and 14.9% were unique to Aripo (N = 36 samples).

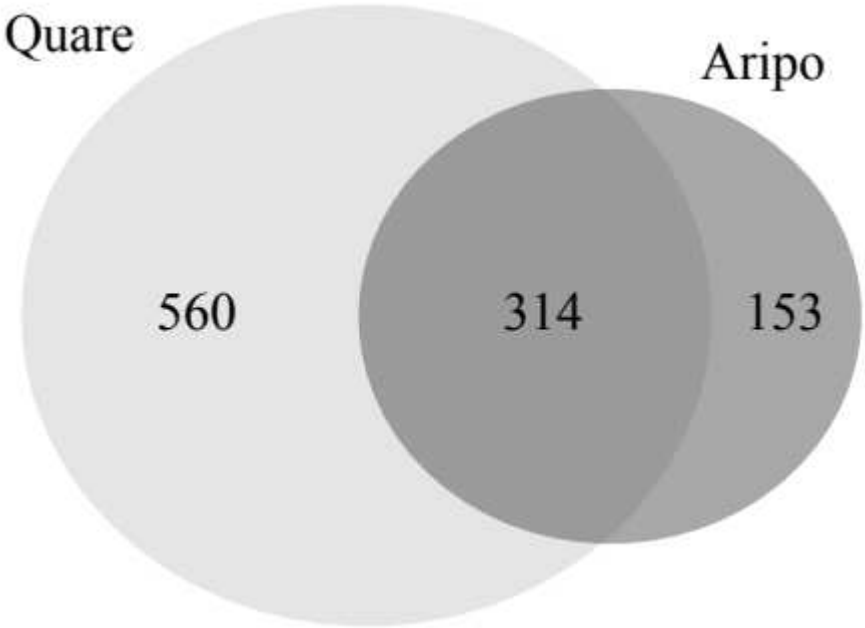

Figure 4.2: Total number of discovered RESs and their overlap between the Aripo and Quare drainages after post-filtering. 3.7x (560 to 153) more unique sites were discovered in Quare as compared to Aripo. The two drainages shared 314 sites in common.

No major difference between group or drainage edit level distributions

To characterize the distribution of edit levels in our filtered data set (at least 4 fish editing with a minimum supporting reads of at least 10, and a minimum edit level of 5%), and to look for major pattern differences in distribution of edit levels based on population of origin and rearing condition, by drainage, we looked at distributions of edit levels (Fig 4.3). The distributions were all similar with a total edit level mean (all groups both drainages) of 0.35 (median=0.25) and standard deviation = 0.27. Over half of our analyzed RESs had an edit level of 25% or less, indicating A-to-I editing targets are commonly edited at low levels.

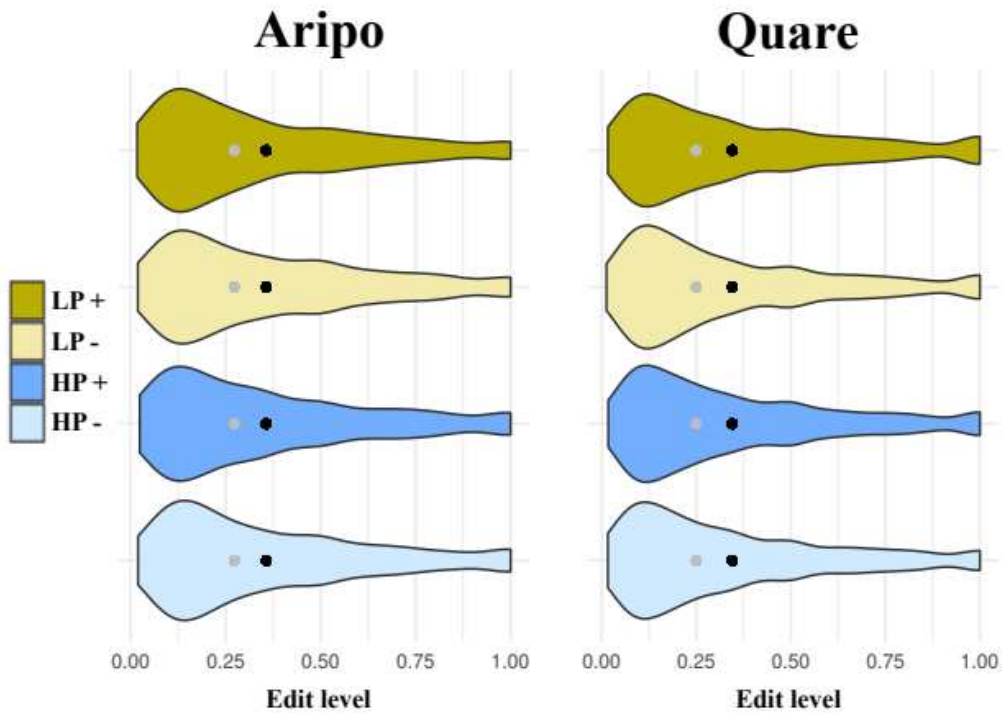

Figure: 4.3: Violin plots showing similar distributions of edit levels for RESs statistically modeled from Aripo and Quare drainages in all populations and rearing conditions. Plotted are RESs identified in at least 4 fish. Grey dots indicate the mean edit level, black dots indicate median edit level. LP+: low predation population, readred with predators; LP-: low predation population, reared without predators; HP+: high predation population, reared with predators; HP-: high predation population, reared without predators.

Inconclusive differential editing results

The results of our statistical modeling suggest no A-to-I editing differences of population, rearing environment or population by rearing. Although the Quare analyzed data set produced 4 significant ($FDR \leq 0.05$) population by rearing effects; however plots of predicted vs. observed supporting read counts indicated poor model fit. See discussion for more details on our preliminary statistical modeling results.

Discussion

A major question about the functions of A-to-I editing is whether or not the effects exerted on the transcriptome largely result in minor or no (organismal) phenotypic variation, i.e. if editing only rarely results in major phenotypic consequences, or if editing is causally implicated in most trait variation. Towards this end we produced and characterized the first guppy A-to-I 'editome' and found numerous unique and shared RESs across Aripo and Quare drainages (Fig. 4.2).

We expected the highest enrichment for A-to-I editing, but our A-to-I call rate (the fraction of calls that were A-to-I) was lower than reported by Ramaswami et al. (2012), Zhang et al. (2017), and Zhang & Xiao (2015) by approximately 7-10%. However, these rates were derived from human B-cell derived cell lines (GM12878). Additionally, our rate was calculated (as per Zhang et al. 2017) after applying minimum post-filtering criteria (required a sum of at least 2 supporting reads) to the SPRINT data. Differences in species, tissue, and post-filtering requirements may explain the discrepancies in A-to-I call rates. A study characterizing the A-to-I editome of Zebrafish showed a lower A-to-I rate of ~80% across 762 genes (Yan et al., 2017).

Our mean call rate was slightly higher than Yan et al. (2017), but our calculation included the total (intra- and intergenic) A-to-I RES set, whereas their calculation was based only on intragenic RESs. Our A-to-I rate suggests good specificity of A-to-I (and possibly C-to-U) editing.

We observed many sites in common between drainages, but also sites unique to each drainage, indicating for at least some sites there is conservation of across guppy lineages (Fig 4.2). We looked broadly across our RES characterization set to determine if fish from different drainages, populations, or rearing conditions had major transcriptome wide differences in edit levels, such as hypo or hyper editing. We did not see strong indication of global differences in editing based on population, drainage, or rearing condition. Our next analysis focused on specific RES quantitative differences.

We statistically modeled each site (glmm), and for Aripo did not find any significant effects of population, rearing or their interaction. At an FDR of 0.05, we did detect 4 significant interaction effects in Quare. Our models weakly suggest that populations and rearing conditions do not differ in edit level, but further statistical modeling is necessary to confirm these conclusions. Additionally, approximately 20% (Quare) and 30% (Aripo) of all modeled RESs did not converge during model fitting. This may be due to the chosen model architecture, as well as attempting to account for the highly zero-inflated count variable. The high non-convergence rate could also be attributed to the challenge of modeling different types of distributions that do not fit the negative binomial distribution. To improve convergence, alternative optimization algorithms could be explored instead of relying solely on the default algorithm. Taken together we did not detect strong differences in editing by population, rearing condition, or population by rearing interaction, but further work is necessary to confirm this.

It's clear some RESs are associated with strong phenotypic outcomes, such as the aberrant RNA editing of the serotonin 2C receptor (HTR2C), which has been linked to numerous major mental disorders in humans (Iwamoto & Kato, 2003; Iwamoto, Nakatani, Bundo, Yoshikawa, & Kato, 2005; Weissmann et al., 2016). Although altered editing of a single locus has been connected to multiple mental health related phenotypes, A-to-I editing has also been implicated in a more global way, for example widespread RNA hypoediting was linked to schizophrenia via two recoding sites in the mitochondrial gene, mitofusin 1 (MFN1) (Choudhury et al., 2023). Much attention has been focused on the deleterious effects of A-to-I editing; however there is evidence of adaptive benefits of editing as well. Although transcriptome-wide mapping studies have shown that A-to-I editing is widespread in all metazoans (Ramaswami & Li, 2014, 2016). It is important to note (in all vertebrates) that the majority of A→I editing sites are located outside of mRNA coding regions and do not directly impact protein products (Rosenthal & Eisenberg, 2023).

Numerous editing sites have been found in humans, and significant numbers have even been observed in eumetazoans, such as corals (Porath et al., 2017). In an analysis of approximately 10,000 human samples, around 1,000 recoding sites were identified, with approximately 200 of them conserved in non-primate mammals (Gabay et al., 2022). In contrast, the number of noncoding sites identified was in the millions. Similarly, in zebrafish (Buchumenski et al., 2021), ants (Li et al., 2014), and bees (Porath et al., 2019), the number of recoding sites identified ranged from ~100-200.

Cephalopods show striking differences; the transcriptome of the squid nervous system first revealed a significant number of recoding sites (57,108), impacting the majority of the encoded proteins (6,991 of 12,039 open reading frames identified) (Alon et al., 2015). Further

studies have shown widespread recoding in coleoid cephalopods, including *Octopus bimaculoides* (Albertin et al., 2015; Liscovitch-Brauer et al., 2017), *Octopus vulgaris*, cuttlefish (*Sepia officinalis*) (Liscovitch-Brauer et al., 2017), Hawaiian bobtail squid (*Euprymna scolopes*), and striped pajama squid (*Sepioloidea lineolata*) (Shoshan, Liscovitch-Brauer, Rosenthal, & Eisenberg, 2021). In cephalopods, recoding levels exhibit significant variation among tissues, with neural tissues showing the highest levels, recoding constitutes 11-13% of the overall RNA editing activity observed (Albertin et al., 2015; Liscovitch-Brauer et al., 2017), in contrast to less than 1% in mammals (Bazak et al., 2014). It's clear from the cephalopod, vertebrate (including primate specific focuses), and plant (primarily C-to-U editing) research that RNA editing functions in different ways according to the particular phylogenetic system.

Our characterization set shows A-to-I editing is fairly widespread within protein-coding regions of the guppy genome. Although widespread, our quantitative analyses showed no differences in A-to-I editing across population, rearing condition, or their interaction levels, despite known major phenotypic consequences of both rearing environment and genetic background. Our study suggests A-to-I editing is commonly found throughout the guppy transcriptome, but we did not find evidence that editing likely contributes to the phenotypic variation observed in our samples. However, our statistical modeling will require further refinements to draw conclusive inferences. Transcriptome-wide changes in editing such as hypo or hyper editing may have minimum functional impact on most RESs targets (at least in vertebrates), and incur minimal phenotypic consequences, but occasionally hit rare targets that result in major phenotypic consequences, often deleterious, but highly dependent on the particular phylogeny in question.

CHAPTER 4 WORKS CITED

Anne-Laure. (2010). *RNA Biology Plant RNA editing*. https://doi.org/10.4161/rna.7.2.11343

Bahn, J. H., Lee, J.-H., Li, G., Greer, C., Peng, G., & Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research*. https://doi.org/10.1101/gr.124107.111

Blanc, V., & Davidson, N. O. (2003). C-to-U RNA editing: Mechanisms leading to genetic diversity. *Journal of Biological Chemistry*, *278*(3), 1395–1398. https://doi.org/10.1074/jbc.R200024200

Borchert, G. M., Gilmore, B. I., Spengler, R. M., Xing, Y., Lanier, W., Bhattacharya, D., & Davidson, B. L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. *Human Molecular Genetics*, *18*(24), 4801–4807. https://doi.org/10.1093/HMG/DDP443

Boris Zinshteyn, K. N. (2010). NIH Public Access. *Wiley Interdiscip Rev Syst Biol Med*, *1*(2), 202–209. https://doi.org/10.1002/wsbm.10.Adenosine-to-inosine

Choudhury, M., Fu, T., Amoah, K., Jun, H.-I., Chan, T. W., Park, S., … Xiao, X. (2023). Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function. *Science Advances*, *9*(14). https://doi.org/10.1126/SCIADV.ADE9997

Covello, P., & Gray, M. (1993). *On the evolution of RNA editing*.

Endler, J. A. (1995). Multiple-trait coevolution and environmental gradients in guppies. *Philos. Tmns. R. Sot. London SW B*, *13*(29), 107–122.

Farajollahi, S., & Maas, S. (2010). Molecular diversity through RNA editing: a balancing act. *Trends in Genetics*, *26*(5), 221–230. https://doi.org/10.1016/J.TIG.2010.02.001

Fischer, E. K., Song, Y., Hughes, K. A., Zhou, W., & Hoke, K. L. (2021). Nonparallel transcriptional divergence during parallel adaptation. *Molecular Ecology*, *30*(6), 1516–1530. https://doi.org/10.1111/mec.15823

Fraser, B. A., Whiting, J. R., Paris, J. R., Weadick, C. J., Parsons, P. J., Charlesworth, D., … Weigel, D. (2020). Improved Reference Genome Uncovers Novel Sex-Linked Regions in the Guppy (Poecilia reticulata). *Genome Biology and Evolution*, *12*(10), 1789–1805. https://doi.org/10.1093/gbe/evaa187

Fumagalli, D., Gacquer, D., Rothé, F., Lefort, A., Libert, F., Brown, D., … Detours, V. (2015). Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. *Cell Reports*, *13*(2), 277–289. https://doi.org/10.1016/J.CELREP.2015.09.032

Gaisler-Salomon, I., Kravitz, E., Feiler, Y., Safran, M., Biegon, A., Amariglio, N., & Rechavi, G. (2014). Hippocampus-specific deficiency in RNA editing of GluA2 in Alzheimer's disease. *Neurobiology of Aging*, *35*(8), 1785–1791. https://doi.org/10.1016/J.NEUROBIOLAGING.2014.02.018

Grice, L. F., & Degnan, B. M. (2015). The origin of the ADAR gene family and animal RNA editing. *BMC Evolutionary Biology*, *15*(1), 1–7. https://doi.org/10.1186/s12862-015-0279-3

Han, L., Diao, L., Yu, S., Xu, X., Li, J., Zhang, R., … Liang, H. (2015). The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell*, *28*(4), 515–528. https://doi.org/10.1016/J.CCELL.2015.08.013

Haskins, C. Haskins, E., McLaughlin, J., & H. (1961). *Polymorphism and population structure in Lebistes...* Retrieved from https://scholar.google.com/scholar?oi=gsb95&q=Polymorphism and population structure in Lebistes reticulatus%2C an ecological study&lookup=0&hl=en

Hideyama, T., Yamashita, T., Aizawa, H., Tsuji, S., Kakita, A., Takahashi, H., & Kwak, S. (2012). Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiology of Disease*, *45*(3), 1121–1128. https://doi.org/10.1016/J.NBD.2011.12.033

Iwamoto, K., & Kato, T. (2003). *RNA editing of serotonin 2C receptor in human postmortem brains of major mental disorders*. https://doi.org/10.1016/S0304-3940(03)00608-6

Iwamoto, K., Nakatani, N., Bundo, M., Yoshikawa, T., & Kato, T. (2005). *Altered RNA editing of serotonin 2C receptor in a rat model of depression*. https://doi.org/10.1016/j.neures.2005.06.001

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Magurran, A. E. (2005). *Evolutionary ecology: the Trinidadian guppy*. Oxford University Press on Demand.

Maydanovych, O., & Beal, P. A. (2006). Breaking the central dogma by RNA editing. *Chemical Reviews*, *106*(8), 3397–3411. https://doi.org/10.1021/CR050314A/ASSET/IMAGES/LARGE/CR050314AF00011.JPEG

Nishikura, K. (2010). Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry*, *79*, 321. https://doi.org/10.1146/ANNUREV-BIOCHEM-060208-105251

Paz-Yaacov, N., Bazak, L., Buchumenski, I., Porath, H. T., Danan-Gotthold, M., Knisbacher, B. A., … Levanon, E. Y. (2015). Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Reports*, *13*(2), 267–276. https://doi.org/10.1016/J.CELREP.2015.08.080

Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D'Erchia, A. M., & Pesole, G. (2015). Profiling RNA editing in human tissues: towards the inosinome Atlas. *Scientific Reports 2015 5:1*, *5*(1), 1–17. https://doi.org/10.1038/srep14941

Porath, H. T., Carmi, S., & Levanon, E. Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nature Communications*, *5*, 1–10. https://doi.org/10.1038/ncomms5726

Ramaswami, G., Lin, W., Piskol, R., Tan, M. H., & Davis, C. (2012). Accurate identification of human Aluand non-AluRNA editing sites. *Nature Protocols*, *9*(6), 579–581. https://doi.org/10.1038/nmeth.1982.Accurate

Rueter, S. M., Dawson, T. R., & Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. *Nature*, *399*(May), 75–80. https://doi.org/10.1038/19992

Sloan, D. B. (2017). Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. *Biology Letters*, *13*(8), 20170314. https://doi.org/10.1098/rsbl.2017.0314

Slotkin, W., & Nishikura, K. (2013). Adenosine-to-inosine RNA editing and human disease. *Genome Medicine*, *5*(11), 1–13. https://doi.org/10.1186/GM508/FIGURES/4

Torres-Dowdal, J., Handelsman, C. A., Reznick, D. N., Ghalambor, C. K., Torres-Dowdall, J., Handelsman, C. A., … Ghalambor, C. K. (2012). Local adaptation and the evolution of of phenotypic plasticity in Trinidadian guppies (Poecilia reticulata). *Evolution*, *66*(11), 3432–3443. https://doi.org/10.5061/dryad.84gf5

Wahlstedt, H., Daniel, C., & Hman, M. O. ¨. (2009). Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome Research*. https://doi.org/10.1101/gr.089409.108

Weissmann, D., van der Laan, S., Underwood, M. D., Salvetat, N., Cavarec, L., Vincent, L., … Pujol, J. F. (2016). Region-specific alterations of A-to-I RNA editing of serotonin 2c receptor in the cortex of suicides with major depression. *Translational Psychiatry 2016 6:8*, *6*(8), e878–e878. https://doi.org/10.1038/tp.2016.121

Yan, S., Lu, Y., He, L., Zhao, X., Wu, L., Zhu, H., … Xing, Q. (2017). Dynamic editome of zebrafish under aminoglycosides treatment and its potential involvement in ototoxicity. *Frontiers in Pharmacology*, *8*(NOV), 1–13. https://doi.org/10.3389/fphar.2017.00854

Zhang, F., Lu, Y., Yan, S., Xing, Q., & Tian, W. (2017). SPRINT: An SNP-free toolkit for identifying RNA editing sites. *Bioinformatics*, *33*(22), 3538–3548. https://doi.org/10.1093/bioinformatics/btx473

Zhang, Q., & Xiao, X. (2015). Genome sequence–independent identification of RNA editing sites. *Nature Methods 2015 12:4*, *12*(4), 347–350. https://doi.org/10.1038/nmeth.3314

Zipeto, M. A., Jiang, Q., Melese, E., & Jamieson, C. H. M. (2015). RNA rewriting, recoding, and rewiring in human disease. *Trends in Molecular Medicine*, *21*(9), 549–559. https://doi.org/10.1016/J.MOLMED.2015.07.001

CHAPTER 5: CONCLUSION

The 2nd chapter of this dissertation focused on detection and characterization of miRNAs and piRNAs in the Trinidadian guppy. Essential proteins in the Dicer-dependent miRNA pathway were confirmed, and our data supported the 5'-Uridine bias in miRNAs. We implemented strict filtering criteria to reduce false discovery of miRNAs. Tissue-specific expression patterns of miRNAs were observed, and the expression profiles of miRNAs and piRNAs showed dynamic patterns across the developmental timeline. We used a new program, tinyRNA for piRNA analysis and detected piRNA reads in our germline samples. This first characterization of guppy miRNAs and piRNAs lays the foundation for further investigations into their roles in gene regulation, evolutionary processes, and phenotypic plasticity.

In chapter 3, we identified differentially expressed (DE) miRNAs between populations, rearing conditions, and their interaction. Interestingly, approximately 18% of the known guppy brain miRNAome showed differential expression, which was comparable to the DE mRNA analysis (22%) conducted by Fischer et al. (2021) in the same guppy populations. We interpreted population main effects as arising from genetic background differences between populations, leading to differential miRNA regulation. We observed both plasticity in miRNA expression and the evolution of plasticity itself. The responsiveness of miRNAs to the environment varied based on genomic background, as evidenced by significant interaction effects. Analyzing different categories of evolved expression plasticity in our dataset, we found equal numbers of gains, losses, changes in extent, and changes in direction of plasticity.

Similar to the qualitative patterns found in the mRNAs by (Fischer et al. 2021), we also found (for miRNAs) four patterns of evolved plasticity with no single predominant category. This first step towards delineating the patterns of evolved plasticity in gene expression can be

used for comparisons to other levels of biological organization, i.e. to determine potential associations between molecular patterns of evolved plasticity, and organismal level patterns of evolved plasticity. Future research could also aim to link these miRNA expression patterns to specific organismal traits.

In chapter 4 we produced and characterized the first guppy A-to-I 'editome' and found numerous unique and shared RNA editing sites (RESs) across different populations and drainages. The A-to-I call rate in guppies was slightly lower than reported in human B-cell derived cell lines (F. Zhang, Lu, Yan, Xing, & Tian, 2017; Q. Zhang & Xiao, 2015). Differences in species, tissue, and post-filtering requirements may explain the difference. While some RESs were conserved across guppy lineages, others were unique to specific drainages.

Although some RESs have been linked to significant phenotypic outcomes, such as mental disorders (Iwamoto & Kato, 2003), A-to-I editing has also been implicated in a more global way. For example, widespread RNA hypoediting was linked to schizophrenia in humans (Choudhury et al., 2023). While much attention has been focused on the deleterious effects of A-to-I editing, there is evidence of adaptive benefits as well, especially in the coleoid cephalopods (Liscovitch-Brauer et al., 2017).

In guppies, the transcriptome-wide mapping of A-to-I editing showed it to be fairly widespread within protein-coding regions. However, quantitative analyses did not reveal significant differences in A-to-I editing levels across populations or rearing conditions, despite known major phenotypic consequences of both environmental conditions and genetic background. Our study suggests that while A-to-I editing is commonly found throughout the guppy transcriptome, it may not be a major contributor to the observed phenotypic variation. Transcriptome-wide changes in editing may have minimal functional impact on most targets but

occasionally hit rare targets that result in major phenotypic consequences, often deleterious, depending on the specific clade involved.

Comparative studies of miRNAs are enhanced when conducted across a diverse range of organisms with annotated miRNAs. This dissertation contributed to this endeavor by establishing the first Trinidadian guppy miRNAome, enriching the growing miRNA database. Our findings demonstrated that miRNA expression patterns show similar evolved plasticity as observed in mRNAs. Additionally, our results suggest that A-to-I editing, specifically in terms of edit levels, is not strongly associated with phenotypic variation among populations, rearing conditions, or their interactions. As we continue to explore different forms of transcriptome plasticity in various organisms and natural settings, we gain valuable insights into the molecular mechanisms underlying evolutionary processes and adaptation.

CHAPTER 5 WORKS CITED

Choudhury, M., Fu, T., Amoah, K., Jun, H.-I., Chan, T. W., Park, S., … Xiao, X. (2023). Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function. Science Advances, 9(14). https://doi.org/10.1126/SCIADV.ADE9997

Fischer, E. K., Song, Y., Hughes, K. A., Zhou, W., & Hoke, K. L. (2021). Nonparallel transcriptional divergence during parallel adaptation. Molecular Ecology, 30(6), 1516–1530. https://doi.org/10.1111/mec.15823

Iwamoto, K., & Kato, T. (2003). RNA editing of serotonin 2C receptor in human postmortem brains of major mental disorders. https://doi.org/10.1016/S0304-3940(03)00608-6

Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., Ziv, T., … Eisenberg, E. (2017). Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. Cell, 169(2), 191-202.e11. https://doi.org/10.1016/j.cell.2017.03.025

Zhang, F., Lu, Y., Yan, S., Xing, Q., & Tian, W. (2017). SPRINT: An SNP-free toolkit for identifying RNA editing sites. Bioinformatics, 33(22), 3538–3548. https://doi.org/10.1093/bioinformatics/btx473

Zhang, Q., & Xiao, X. (2015). Genome sequence–independent identification of RNA editing sites. Nature Methods 2015 12:4, 12(4), 347–350. https://doi.org/10.1038/nmeth.3314